1 A PATH Analytic Approach to Assess the Effects of Antipsychotic Treatment on Depressive Symptoms in Schizoaffective Disorder

Ibrahim Turkoz¹, Dong-Jing Fu², Cynthia A. Bossie², Larry Alphs²

¹Johnson & Johnson Pharmaceutical Research and Development, LLC, Titusville, NJ, USA; ²Ortho-McNeil Janssen Scientific Affairs, LLC, Titusville, NJ, USA

Introduction: Antipsychotics are often used to manage the affective (manic and depressive) and psychotic symptoms of patients with schizoaffective disorder. However, it is difficult to determine whether observed improvements in depressive symptoms are achieved through effects on depressive symptoms or effects on psychotic symptoms. This analysis evaluated the effect of antipsychotic treatment on depressive symptoms after accounting for effects on positive and negative psychotic symptoms and extrapyramidal symptoms in subjects with schizoaffective disorder.

Methods: A database (N=614) comprising two 6-week, randomized, placebo-controlled studies of antipsychotic treatment vs. placebo in subjects with schizoaffective disorder was used. Analysis included subjects with baseline depressive symptoms, defined as a HAM-D-17 score ≥16. Structural equation models (PATH analysis) separated total effects of antipsychotic treatment into effects on depressive symptoms and effects on positive, negative, and extrapyramidal symptoms. The change in HAM-D-17 score at week 6 endpoint was the dependent variable; the change in PANSS positive and negative factors and the change in SAS (for extrapyramidal symptoms) scores at week 6 endpoint were independent variables. These variables were thought to be mediators of change in depressive score. In each regression equation, a factor for treatment was also included so that comparisons between antipsychotic treatment and placebo could be made.

Results: 332 subjects were identified. Mean (SD) age was 37.7 (9.8) years; 54.5% were male. Mean (SD) baseline HAM-D-17 score was 22.9 (4.9). 51.8% had the depressive subtype of schizoaffective disorder, and 37.0% attempted suicide. Significant improvement from baseline in HAM-D-17 score at week 6 endpoint was observed with antipsychotic treatment vs. placebo (LS mean [SE] difference: –3.0 [0.9]; P=0.002). PATH analysis determined that 26.3% of treatment effect was attributed to improvement in depressive symptoms, whereas 46.1% and 28.2% of treatment effect were attributed to positive and negative symptoms, respectively. Treatment effect was not attributed to changes in extrapyramidal symptoms. Based on the PATH analysis, the total treatment effect corresponded to a 3-point greater improvement in the HAM-D-17 score compared with placebo group.

Conclusions: From the PATH analysis, up to 26% of treatment effect observed with an antipsychotic versus placebo in subjects with schizoaffective disorder was attributed to improvement in depressive symptoms, whereas 74% of treatment effect was attributed to improvement in positive and negative symptoms associated with psychosis. Supported by Ortho-McNeil Janssen Scientific Affairs, LLC Disclosures IT is an employee of Johnson & Johnson Pharmaceutical Research and Development, LLC, and a Johnson & Johnson stockholder. DJF, CB, and LA are employees of Ortho-McNeil Janssen Scientific Affairs, LLC, and Johnson & Johnson stockholders.

2 Relationship between the Clinical Global Impression of Severity for Schizoaffective Disorder and Manic and Depressive Mood Scales

Dong-Jing Fu¹, Cynthia A. Bossie¹, Ibrahim Turkoz¹, John Sheehan¹, Larry Alphs¹

¹Ortho-McNeil Janssen Scientific Affairs, LLC, Titusville, NJ, USA; ²Johnson & Johnson Pharmaceutical Research and Development, LLC, Titusville, NJ, USA

Background: The Hamilton Rating Scale–Depression 17- and 21-item versions (HAM-D-17, HAM-D-21) and Young Mania Rating Scale (YMRS) are widely used to assess mood symptoms in clinical trials. Data are often evaluated as change scores or in relation to a predefined threshold score. However, a major limitation of these approaches is that there is no accepted definition of a clinically meaningful change for these scales. The Clinical Global Impression–Severity (CGI-S) scale is widely used to measure overall clinical status. This analysis explored the relationship between...
ratings on the HAM-D-17 or YMRS and those on the depressive and manic subscales of the CGI-S for schizoaffective disorder (CGI-S-SCA) scale to determine what change score on the mood scales corresponds to a 1-point change on the CGI-S-SCA scale.

**Methods:** This post hoc analysis used a database (N=614) composed of two 6-week, randomized, placebo-controlled studies of antipsychotic versus placebo in symptomatic subjects with schizoaffective disorder assessed using HAM-D-17, YMRS, and CGI-S-SCA scales. Simple and multiple (with explanatory variables for treatment, protocol, concomitant medication stratum, and baseline score) regression models explored relationships between ratings on the YMRS and HAM-D-17 and the depressive and manic subscales of the CGI-S-SCA scale from baseline to the week-6 endpoint. Clinically meaningful improvement was defined as a 1-point change in the CGI-S-SCA scale. Preliminary data are presented.

**Results:** A simple regression model suggested that a 1-point change in the CGI-S-SCA scale of depression corresponded to a 4.4-point (SE=0.2) change in the HAM-D-17, whereas a multiple regression model suggested correspondence to a 3.6-point (SE=0.2) change in the HAM-D-17. A simple regression model suggested that a 1-point change in the CGI-S-SCA scale of mania corresponded to a 6.6-point (SE=0.2) change in the YMRS, and the multiple regression model suggested correspondence to a 5.8-point (SE=0.2) change in the YMRS.

**Conclusions:** The clinically meaningful improvement in manic and depressive domains of the CGI-S-SCA scale corresponded to approximately 4- and 6-point changes on the HAM-D-17 and YMRS, respectively, in a group of symptomatic subjects with schizoaffective disorder. Future studies are needed to confirm these findings in subjects with broader ranges of change and for worsening as well as improving.

**Disclosures:** Supported by Ortho-McNeil Janssen Scientific Affairs, LLC. IT is an employee of Johnson & Johnson Pharmaceutical Research and Development, LLC, and a Johnson & Johnson stockholder. DJF, CAB, and LA are employees of Ortho-McNeil Janssen Scientific Affairs, LLC, and Johnson & Johnson stockholders. JS is an employee of Ortho-McNeil Janssen Scientific Affairs, LLC.

### Impact of Ethnicity on Efficacy and Safety of Ziprasidone in Patients with Schizophrenia or Bipolar Disorder

Charles L. Bowden¹, Onur N. Karaya², Sheela Kolluri², Doug G. Vanderburg²

¹ University of Texas Health Science Center, San Antonio, TX, USA; ² Pfizer Inc. NY, USA

**Background:** Previous analyses have demonstrated similar efficacy and safety profiles for atypical antipsychotics within black and white patients with schizophrenia (1, 2). We investigated the impact of ethnicity on the efficacy and safety of ziprasidone in patients with schizophrenia or bipolar disorder.

**Methods:** Efficacy and safety data from 4 schizophrenia randomized controlled trials (RCTs) and 3 bipolar mania RCTs were analyzed by ethnicity (white, black, Asian, other). Efficacy variables comprised changes from baseline to primary endpoint (Week 4 in schizophrenia studies and Week 3 in bipolar studies) in Brief Psychiatric Rating Scale (BPRS) and Clinical Global Improvement-Severity (CGI-S) for schizophrenia and Mania Rating Scale (MRS) and CGI-S for bipolar disorder. Safety measures included incidence of ≥7% weight gain, extrapyramidal symptoms, somnolence, akathisia, and all-cause discontinuation.

**Results:** A total of 975 subjects were analyzed in the schizophrenia data set and 681 subjects in the bipolar data set. Efficacy outcomes were similar across ethnic groups in both the schizophrenia data set and the bipolar data set (Table 1). Small drug-placebo differences in efficacy outcomes for schizophrenia were likely due to pooling of dosage groups. All safety outcomes were also highly similar across ethnic groups (Incidence of ≥7% weight gain shown in Table, other data not shown).

#### Table 1. Efficacy and Safety Outcomes (continued on next page)
Conclusions: These analyses suggest that there are no substantial differences in the efficacy and safety/tolerability of ziprasidone amongst different ethnic groups with schizophrenia or bipolar disorder.

References:
1) Stauffer, V.L., Sniadecki, J.L., Piezer, K.W., Gatz, J., Kollack-Walker, S., Poole Hoffmann, V., Conley, R., Durell, T: Impact of race on efficacy and safety during treatment with olanzapine in schizophrenia, schizophreniform or schizoaffective disorder. BMC Psychiatry 2010; 10:89

4 A Posthoc Analysis of Eszopiclone Effects on Cardiometabolic Indices of Hyperarousal in the Treatment of Primary Insomnia

Randall Marshall¹, Todd Grinnell¹, William Spalding¹, Jacqueline Zummo¹, Edward Schweizer², Andrew Krystal³

¹Sunovion Pharmaceuticals Inc, Marlborough, MA, USA; ²Paladin Consulting Group, Hoboken, NJ, USA; ³Duke University, Durham, NC, USA

Introduction: Despite considerable evidence that chronic insomnia may be associated with underlying hyperarousal, little is known about the potential impact of pharmacologic insomnia therapy on measures of hyperarousal. The current study examined the effects of eszopiclone treatment on cardiometabolic indices of hyperarousal in insomnia patients.

Methods: Data were analyzed from two double-blind, placebo-controlled eszopiclone trials in patients diagnosed with primary insomnia: (1) a 2-week study in the elderly; and (2) a 6-week study in adults. Mean change±SD from baseline to Week 2 (elderly) or 6 (adults) in evening (adult only) and morning heart rate (HR), systolic and diastolic blood pressure (SBP; DBP) were assessed in a subset of these insomnia patients (elderly: n=78; adults: n=68) potentially enriched for hyperarousal via stratification by a 25th percentile split on HR, SBP, and DBP for each variable separately. Stratification was necessary because this population excluded patients with significant medical comorbidity.

Results: In the adult study, the eszopiclone group demonstrated a trend in mean reduction of morning HR and evening SBP vs. placebo (HR: -10.02±1.44 vs. -5.90±1.68; p=0.053; evening SBP: -11.4±2.41 vs. -4.68±3.19; p=0.08). In the elderly study, the eszopiclone group demonstrated a significant reduction in morning HR (-9.79±1.24 vs. -4.12±1.32; p=0.002) and a trend towards reduction in morning SBP (-11.52±2.46 vs. -4.19±3.12; p=0.06). There were no significant differences between treatment groups for the bottom 75th percentile subgroup in either study.

Conclusions: In this exploratory analysis, treatment with eszopiclone was associated with a tendency to decrease morning HR which was statistically significant in older insomnia patients. Trends for eszopiclone to decrease SBP but not DBP were also noted. These results suggest that eszopiclone treatment may lead to a decrease in sympathetic tone in insomnia patients with potential markers of hyperarousal. Future research should prospectively study insomnia patients selected for evidence of hyperarousal.

Disclosure: Funded by Sunovion Pharmaceuticals Inc.

5 Evidence for the Pharmacogenomic Effect of SULT4A1 on Response to Atypical Antipsychotics

Timothy L. Ramsey¹, Herbert Y. Meltzer³, Guy N. Brock³, Bharat Mehrotra¹, Karu Jayathilake², William Bobo², Mark D. Brennan¹

¹SureGene, LLC, Louisville, KY, USA; ²Psychiatric Hospital at Vanderbilt University, Nashville, TN, USA; ³Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, KY, USA

Introduction: A specific haplotype of the SULT4A1-1 gene, SULT4A1-1, impacts psychopathology and antipsychotic drug response in Caucasian subjects from Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) and a
Methods: Relationship to baseline psychopathology was evaluated using linear regression of Positive and Negative Syndrome Scale (PANSS) Total score. Drug response was evaluated using mixed model repeat measures (MMRM) for change in PANSS. Results: For the CATIE sample, SULT4A1-1(+) patients displayed higher baseline PANSS (p=0.03) and, when treated with olanzapine, demonstrated a significant haplotype-drug-time interaction (p=0.009) in the MMRM. Additionally, SULT4A1-1(+) patients treated with olanzapine displayed greater improvement compared to SULT4A1-1(-) patients treated with olanzapine (p=0.008) or to SULT4A1-1(+) patients treated with risperidone (p=0.006). In the replication sample, SULT4A1-1(+) patients had higher baseline PANSS, with the difference being significant for the BP sample (p=0.006). SULT4A1-1(+) SZ patients treated with olanzapine showed greater improvement than SULT4A1-1(-) patients treated with olanzapine (p = 0.05) or than SULT4A1-1(+) patients treated with risperidone (p=0.05). SULT4A1-1(+) BP patients treated with olanzapine showed greater improvement than SULT4A1-1(-) patients treated with risperidone (p=0.07).

Conclusions: SULT4A1-1 status may provide a predictor of differential response in subjects treated with olanzapine, and SULT4A1-1(+) subjects may respond better to olanzapine than risperidone.

6 Adjunctive Armodafinil for Negative Symptoms in Adults with Schizophrenia: A Double-Blind, Placebo-Controlled Study

John M. Kane¹, James M. Youakim², R. Yang², Jane Tiller²

¹Department of Psychiatry, The Zucker Hillside Hospital, Glen Oaks, NY, USA; ²Cephalon, Inc., Frazer, PA, USA

Introduction: A previous study suggested that adjunctive armodafinil may decrease negative symptoms in schizophrenia. The purpose of this study evaluated the efficacy and tolerability of adjunctive armodafinil for negative symptoms in adult patients with schizophrenia treated with antipsychotic medication.

Methods: Adult patients with schizophrenia being treated with olanzapine, risperidone, or paliperidone who had been clinically stable for at least 8 weeks were eligible for this 24-week, double-blind, placebo-controlled, parallel-group study. Patients were randomized 1:1:1:1 to once daily armodafinil 150 mg, 200 mg, or 250 mg, or placebo. The primary efficacy outcome was the negative symptoms subscale score of the Positive and Negative Syndrome Scale (PANSS). Patients were stratified according to PANSS negative symptoms subscale score (<17 vs. ≥17). The original inclusion criteria did not require a minimum PANSS negative symptoms subscale score; however, approximately half-way through the study, the protocol was amended to require a PANSS negative symptoms subscale score of ≥15 at screening and baseline visits. An interim analysis was done half-way through the study, thus generally including all patients enrolled before the protocol amendment. Tolerability was assessed throughout the study.

Results: The mean change from baseline (SD) of the PANSS negative scale score was -1.9 (3.8) for the armodafinil 150-mg group (n=70), -2.3 (3.6) for the 200-mg group (n=69), -2.0 (3.3) for the 250-mg group (n=71), and -2.2 (4.1) for the placebo group (n=70; all not significant, armodafinil vs. placebo). At baseline, patients included in the interim analysis (n=136) had a mean PANSS negative symptoms subscale score of 18.2 (4.64) and PANSS total score of 61.8 (10.60), respectively, for patients not in the interim analysis (n=144). At final visit, mean scores before and after the criteria change were similar resulting in greater mean decrease at final visit in patients enrolled after the criteria change. There were no significant changes in PANSS total or PANSS positive symptoms subscale scores. The frequency of adverse events was generally comparable across treatment groups.

Conclusions: This study found no advantage of adjunctive armodafinil as treatment of negative symptoms in adult patients with schizophrenia treated with antipsychotic medication. Armodafinil did not worsen the symptoms of schizophrenia. Future studies of negative symptoms in schizophrenia might benefit from accepting all comers rather than using a minimum threshold in an effort to avoid rater inflation and thus dilution of any real drug effect.

Disclosure: Sponsored by Cephalon, Inc.

7 Using the GAF to Assess Functional Impairment and Change in Shift Work Disorder (SWD)

Brian Rothman¹, Ryan Dammerman², Sofija Jovic¹, Christian Yavorsky¹, Mark Opler¹,³

¹ProPhase, LLC, New York, NY, USA; ²Cephalon, Inc., USA; ³New York University, NY, USA

Background: The Global Assessment of Functioning (GAF) addresses social, psychological, and occupational functioning and has been used as an outcome measure. Although some have questioned the GAF’s utility due to observed low inter-rater reliability estimates (e.g., Bates et al., 2002), these problems have been mitigated in research by training and monitoring (Hilsenroth et al., 2000; Vatnaland et al., 2007). We sought to determine whether the GAF can be used as a reliable and valid way to detect functional change in a group of sleep-disordered patients. Our secondary
aim was to contribute to existing research on the GAF by assessing whether training method (e.g., video vs. written vignette) impacts reliability and rater accuracy.

**Methods:** Raters received training and assigned GAF scores after reviewing videotaped interviews and vignettes of sleep-disordered patients. T-tests examined whether GAF ratings were sensitive to functional changes at different stages of illness, and were also used to determine concordance with gold standard scores.

**Results:** Mean comparisons in GAF ratings between the videos were significant ($t(42) = -10.69, p<.001$) and concordance with expert raters was achieved. Ratings also differed significantly between both vignette 1 and vignette 2 ($t(96) = -15.02, p<.001$), and vignette 2 and vignette 3 ($t(57) = 17.47, p<.001$). However, there was only moderate concordance with expert raters for the vignettes.

**Conclusions:** Results suggest that the GAF can assess functional changes in SWD patients, and that this population exhibits social, occupational, and psychological problems that are observable when using broad-based psychiatric scales. Consistent with prior research (e.g., Hilsenroth, 2000), results also indicate that raters are able to score the GAF accurately following training that addresses rating strategy and consistent conceptualization of GAF anchors. Video training materials were superior to vignettes for reaching concordance with expert ratings. Additionally, raters showed higher agreement with gold-standard scores when ratings changed in the expected direction (improvement in the second visit) rather than an unexpected one (worsening after baseline). Therefore, GAF training should address expectancy issues in addition to rater error and other inter-rater reliability problems. This approach will help to improve the GAF’s construct validity when applied to ratings of broader populations, as well as to this SWD cohort.

**Disclosure:** Cephalon sponsored the IM referenced in this abstract and provided a review of this analysis.

8 Inter-rater Reliability of the Conners' Adult ADHD Diagnostic Interview for DSM-IV (CAADID) in an Adult ADHD Trial

Elizabeth P. Sparrow¹, Brian Rothman², Drew Erhardt¹, Christian Yavorsky², E. Daly⁴, Gahan Pandina⁴, Mark Opler²,⁵

¹Sparrow Neuropsychology, P.A., Raleigh, NC, USA; ²ProPhase, LLC, New York, NY; ³Pepperdine University, Malibu, CA; ⁴Johnson & Johnson Pharmaceutical Research and Development, Titusville, NJ; ⁵New York University, New York, NY

**Introduction/Aims:** The CAADID is a semi-structured interview used by clinicians to gather information about DSM-IV criteria for ADHD in childhood and adulthood, including symptoms, age of onset, pervasiveness, and level of impairment. This study examined the reliability of clinician ratings with the CAADID and evaluated the appropriateness of using the CAADID in adult ADHD research.

**Methods:** Seventy-six licensed clinicians independently rated a video administration of the CAADID after receiving didactic training. All participants reported having at least 5 years of experience evaluating/treating ADHD in adults. Each participant watched the video then submitted a rating for each CAADID item. Gold standard ratings for the video were developed by an expert panel; each member of the panel independently rated the video, after which consensus discussions were held for each item.

**Results:** The initial inter-class correlation coefficient (ICC) for the entire data set, including dichotomous and scale data, was high at .991 ($p <.0001$, n=76), with all raters endorsing a diagnosis of ADHD in childhood as well as adulthood and 75 of 76 raters assigning “Combined Type” subtype. DSM-IV Part A (Symptom Criteria) was assessed by 18 yes/no items (9 Inattention symptoms, 9 Hyperactivity/Impulsivity symptoms). ICCs were good for Inattention symptoms in adulthood (.678, $p <.001$), but lower in childhood (.595, $p <.001$). ICCs for Hyperactivity/Impulsivity symptoms were high in adulthood (.917, $p <.001$) and childhood (.980, $p <.001$). DSM-IV Part B (Age of Onset) had perfect agreement (all raters indicated onset before age 7). DSM-IV Part C (Pervasiveness) had high ICCs for Inattention symptoms in adulthood (.969, $p <.001$) and childhood (.892, $p <.001$), and for Hyperactivity/Impulsivity symptoms in adulthood (.973, $p <.001$) and childhood (.730, $p <.001$). DSM-IV Part D (Impairment) had significant variability, with raters endorsing from one to four aspects of impairment. ICC for the Behavior Observation items was relatively low (.340, $p <.001$).

**Conclusions:** Raters showed very high levels of agreement on key items such as diagnosis and subtype. Ratings were more consistent for symptoms of Hyperactivity/Impulsivity than Inattention. Age of onset and pervasiveness of symptoms also had high levels of inter-rater agreement; in contrast, ratings of level of impairment across domains and behavior observations were less consistent. Use of the CAADID may provide some assurance of consistency in diagnostic practice within a clinical study of ADHD in adults. Further research is necessary to determine the impact of skills training for improving ratings of impairment and behavior observations.

**Disclosure:** The authors report no conflicts of interest for this work.
The Virtual Reality Functional Capacity Assessment Tool (VRFCAT): A New Co-Primary Cognitive Measure for Schizophrenia Drug Trials

Kolleen H. Fox¹, Stacy Ruse¹, V. Davis¹, Ricardo Pietrobon², Philip D. Harvey³, Ranga K. Krishnan², Richard S.E. Keefe¹,²

¹NeuroCog Trials, Inc., Durham, NC, USA; ²Duke University Medical Center, Durham, NC, USA; ³University of Miami Miller School of Medicine, Miami, FL, USA

Background: Representatives from the FDA have repeatedly stated that improvement on traditional measures of cognition will not be sufficient for approval of a new drug for cognitive enhancement; patients must also demonstrate improvement on a measure of “functional capacity” meant to approximate competence in activities of daily living. Currently, no co-primary measure has been universally endorsed and the method to achieve this end lacks definition. The Virtual Reality Functional Capacity Assessment Tool (VRFCAT) is designed to create a realistic, simulated environment using state-of-the-art gaming software to assess patients’ proficiency for navigating everyday life.

Methods: The authors worked closely with Applied Research Associates to create a realistic environment approximating daily life. The environment consists of three environments: a kitchen; a bus stop; and a supermarket. The patient is tasked with checking the kitchen for ingredients needed to fulfill a recipe; using the correct change and taking the correct bus to the grocery store based on a bus schedule; purchasing the correct quantity of groceries; counting exact change to pay the grocery bill; and taking the bus back to the apartment. Six versions of the task have been created with modifications to the details that minimize practice effects and increase utilization for clinical trials. A shorter tutorial version is used to minimize any ancillary effects due to patient differences in computer experience. Sixty-eight healthy controls were recruited through random telephone sampling and classified advertisements. Each participant was tested and then retested 7 to 14 days later with a different version. Results: The 68 participants, ranging in age from 19 to 68, were comprised of 22 males and 46 females. Forty-one were Caucasian, 26 were African-American, and 1 was Asian. Descriptive statistics for each version revealed that Version 4 contained significantly outlying data, thus the Factor Analysis conducted to determine the VRFCAT factor structure excluded data from Version 4. The total sample size for the FA was n=46, since any participant having received Version 4 at either visit was not included in this analysis. The FA explained 91% of variance and revealed three factors, which the authors have labeled Reasoning and Problem Solving; Speed of Processing; and Working Memory. The domains were combined into a composite, which yielded an ICC of .61 between testing visits.

Discussion: The pilot study for this new co-primary measure reveals a composite ICC consistent with or better than those for currently used co-primary measures. Although further testing is required, the VRFCAT shows promise as a new measure of functional capacity suited to be a co-primary measure in schizophrenia cognition drug trials. Further testing with a refined measure will be conducted shortly with patients with schizophrenia.

References:

The Profile of Attentional Deficits in Schizophrenia: Implications for Pharmacotherapy

Howard Hassman¹, Lawrence Brownstein¹, David Kretetz¹, Keith Wesnes², Helen Brooker², Chris Edgar²

¹CRI Worldwide, Mount Laurel, NJ, USA; ²United BioSource Corporation, Goring on Thames, UK

Introduction: There is a large body of work showing a range of impairments to cognitive function in schizophrenia. The MATRICS initiative has considered this research, and proposed a number of domains of cognitive function which are potential targets for treatment.

Methods: Attentional Deficits in a population of schizophrenic patients on stable medication will be closely evaluated, and will be related to the clinical severity of the condition using The Cognitive Drug Research computerised assessment system (CDR System). Further, the deficit profile of first time diagnosed previously unmedicated patients will be contrasted to that of patients on stable medication.

Results: The CDR System identified a wide range of deficits in both first time diagnosed as well as patients on stable medication. The most marked deficits were to the ability to focus attention and the time taken to retrieve information held in both working and episodic memory (effect sizes >3). Deficits to focused attention, cognitive processing time, and speed of retrieving information from memory are considerably greater in patients on stable medication, whereas the deficits to sustained attention and the time taken to retrieve information held in both working and episodic memory are
Conclusions: One or more of the major deficits identified in this work could be considered as the initial therapeutic targets for pharmaceuticals. However, even for those compounds which may be expected to have broad benefits, it may be sensible to restrict the primary cognitive outcomes to one or two major domains.

11 Using Translational Biomarkers and Clinical Pharmacodynamic Data to Inform Study Design for Phase II Proof of Efficacy

Kimberly E. Vanover, Robert E. Davis, Sharon Mates

Intra-Cellular Therapies, Inc., New York, NY, USA

Introduction: A novel investigational new drug is being developed for the treatment of schizophrenia, sleep disorders, depression, and other neuropsychiatric and neurological disorders. This new agent is a potent serotonin 5-HT2A receptor antagonist, a dopamine receptor phosphoprotein modulator (DPPM), and serotonin transporter (SERT) inhibitor. As a mesolimbic/mesocortical selective DPPM, it exhibits pre-synaptic partial agonism and post-synaptic antagonism at dopamine D2 receptors in vivo. This new agent also increases phosphorylation of NMDA NR2B receptors downstream of dopamine D1 receptor activation with mesolimbic selectivity, consistent with enhanced glutamatergic neurotransmission. Together, this unique pharmacological profile predicts low dose enhancement of sleep with antipsychotic and antidepressant efficacy at higher doses. Indeed, behavioral data are consistent with antipsychotic and antidepressant efficacy in animal models without adverse motor side effects. Moreover, low doses have demonstrated efficacy to improve sleep in animal models and in a Phase II clinical study in patients with insomnia. Higher doses were shown to be safe and well-tolerated in patients with stable schizophrenia. Pharmacodynamic data from stable patients together with brain occupancy data from healthy volunteers was used to select doses for a Phase II proof of efficacy trial.

Methods: Using positron emission tomography (PET) in healthy volunteers, target receptor and transporter occupancy was determined. [11C]-MDL100907 was used to image cortical 5-HT2A receptors, [11C]-raclopride was used to image striatal D2 receptors and [11C]-DASB was used to image serotonin transporter (SERT) occupancy. Exploratory efficacy measures were included in a Phase I/II safety study in patients with stable schizophrenia to provide pharmacodynamic signals: Positive and Negative Syndrome Scale (PANSS) and Calgary Depression Scale for Schizophrenia (CDSS).

Results: PET data provided proof of concept for engaging key molecular targets and evidence for rapid brain penetration and long lasting brain occupancy in healthy volunteers. Exploratory measures in patients with stable schizophrenia allowed for detection of an early signal of antipsychotic and antidepressant efficacy. Projecting doses from the PET study suggests that a full range of striatal D2 and SERT occupancy can be achieved at doses that are safe and well-tolerated in patients with schizophrenia while low doses already fully saturate 5-HT2A receptors.

Conclusions: The use of these biomarkers provided confirmation of dose-related target engagement in brain consistent with its preclinical profile and predictive of efficacy. Doses for Phase II evaluation of efficacy in acutely exacerbated schizophrenia patients will be based on these data.

12 Neurocognitive Performance and Self-Reported Sedation Following Administration of Lorazepam

Jane Yen-Revollo¹, Alan Boyd², Dexter Morgan², David Loring³, Kimford Meador³

¹University of North Carolina, School of Pharmacy, Chapel Hill, NC, USA; ²CNS Vital Signs, LLC, Morrisville, NC, USA; ³Emory University, Department of Neurology, Atlanta, GA, USA

Background: The goal of this study was to examine and compare the change in neurocognitive performance as well as measure self reported sedation on a 10 point visual analogue scale.

Methods: Using a randomized cross-over design, 28 subjects received placebo or 2 mg of Lorazepam. A series of tests were given including a computerized 10 point sedation scale and computerized neurocognitive tests measuring multiple cognitive domains including memory, psychomotor speed, reaction time, executive function, working memory and sustained attention. A summary age-matched standard score Neurocognitive Index (NCI) was calculated.

Results: The mean change in sedation score between the lorazepam and the placebo group was 2 points on a 10 point scale (p=0.005). The mean change in NCI between the lorazepam and placebo group was 10 points (p=0.005). NCI and Sedation correlated at -0.71 (p<0.001) for the placebo group. NCI and sedation correlated at -0.30 (p=0.11) for the drug group.

Conclusions: Using neurocognitive performance as a measure of drug effect, these data suggest subjects on lorazepam self report their level of sedation less accurately than patients on placebo.

Disclosure: Study was sponsored by GSK.
Variation in MADRS Reliability over the Course of Randomized Controlled Trials: Comparison of Computer vs. Site-Based Rater Administration.

Gary S. Sachs, Suzanne Edman, Dan DeBonis

United BioSource Corporation, Chevy Chase, MD, USA

Background: Determining the reliability of clinical rating scales across the range of circumstances of its proposed use is key to establishing the acceptability of any measure proposed as a drug development tool. Since most reports establishing scale and rater reliability derive their data from an exercise conducted apart from an actual clinical trial, there is a need to establish the performance characteristics of scales within the context of their intended use. Well established scales such as the Montgomery Asberg Depression Rating Scale (MADRS) may provide a useful benchmark against which to judge the merit of other proposed measures.

Methods: Three multicenter placebo controlled double blind studies were identified in which the MADRS was administered independently by site based raters (MADRS SBR) and by a computer (MADRS COMP) as part of a quality management program. Internal scale consistency was assessed using Cronbach’s alpha calculated for the MADRS SBR and MADRS COMP at baseline, the first post randomization visit, study endpoint, other visits, and all visits. All ratings were made in the subject’s native language. Comparisons of the MADRS SBR and MADRS COMP were made at overall and at four study time points Baseline, First Post-randomization visit, Study Endpoint and all other visits. Correlations between the pairs were calculated to examine measurement reliability overall and individually for each MADRS item.

Results: The sample included 7544 pairs of MADRS SBR and MADRS COMP ratings. The datasets had no subject identifiers other than the subject’s study identification number.

<table>
<thead>
<tr>
<th>Study indication</th>
<th>Baseline</th>
<th>First post randomization visit</th>
<th>Study end point (last visit)</th>
<th>Other Visits</th>
<th>All Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alpha</td>
<td>Alpha</td>
<td>Alpha</td>
<td>Alpha</td>
<td>Alpha</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Mean (S.D)</td>
<td>Mean (S.D)</td>
<td>Mean (S.D)</td>
<td>Mean (S.D)</td>
<td>Mean (S.D)</td>
</tr>
<tr>
<td>Overall MADRS SBR</td>
<td>.674</td>
<td>.809</td>
<td>.891</td>
<td>.879</td>
<td>.889</td>
</tr>
<tr>
<td></td>
<td>1020</td>
<td>936</td>
<td>686</td>
<td>4732</td>
<td>7544</td>
</tr>
<tr>
<td></td>
<td>30.3(5.8)</td>
<td>24.4 (8.2)</td>
<td>14.3(10.4)</td>
<td>17.8(10.2)</td>
<td>20.1(10.7)</td>
</tr>
<tr>
<td>Overall MADRS COMP</td>
<td>.688</td>
<td>.776</td>
<td>.854</td>
<td>.841</td>
<td>.847</td>
</tr>
<tr>
<td></td>
<td>1020</td>
<td>936</td>
<td>686</td>
<td>4732</td>
<td>7544</td>
</tr>
<tr>
<td></td>
<td>30.4(8.6)</td>
<td>23.4 (10.2)</td>
<td>15.0 (11.6)</td>
<td>18.1(11.5)</td>
<td>20.3(11.9)</td>
</tr>
<tr>
<td>1. BP Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS SBR</td>
<td>.741</td>
<td>.829</td>
<td>.882</td>
<td>.862</td>
<td>.877</td>
</tr>
<tr>
<td></td>
<td>296</td>
<td>257</td>
<td>186</td>
<td>871</td>
<td>1680</td>
</tr>
<tr>
<td></td>
<td>28.7(6.5)</td>
<td>22.1(8.8)</td>
<td>15.1(10.5)</td>
<td>16.9(9.6)</td>
<td>19.9(10.3)</td>
</tr>
<tr>
<td>MADRS COMP</td>
<td>.759</td>
<td>.807</td>
<td>.855</td>
<td>.849</td>
<td>.852</td>
</tr>
<tr>
<td></td>
<td>296</td>
<td>257</td>
<td>186</td>
<td>871</td>
<td>1680</td>
</tr>
<tr>
<td></td>
<td>29.1(9.9)</td>
<td>22.0(10.8)</td>
<td>16.0(12.0)</td>
<td>18.2(11.8)</td>
<td>20.8(12.2)</td>
</tr>
<tr>
<td>2. BP Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS SBR</td>
<td>.538</td>
<td>.803</td>
<td>.905</td>
<td>.878</td>
<td>.892</td>
</tr>
<tr>
<td></td>
<td>231</td>
<td>218</td>
<td>170</td>
<td>1158</td>
<td>1801</td>
</tr>
<tr>
<td></td>
<td>30.5(5.0)</td>
<td>24.6(8.1)</td>
<td>13.7(11.1)</td>
<td>16.3(10.5)</td>
<td>19.0(11.1)</td>
</tr>
<tr>
<td>MADRS COMP</td>
<td>.649</td>
<td>.787</td>
<td>.877</td>
<td>.849</td>
<td>.858</td>
</tr>
<tr>
<td></td>
<td>231</td>
<td>218</td>
<td>170</td>
<td>1158</td>
<td>1801</td>
</tr>
<tr>
<td></td>
<td>30.3(8.2)</td>
<td>22.7(10.5)</td>
<td>14.1(12.4)</td>
<td>16.3(11.7)</td>
<td>18.8(12.4)</td>
</tr>
<tr>
<td>3. UP Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS SBR</td>
<td>.655</td>
<td>.786</td>
<td>.888</td>
<td>.833</td>
<td>.891</td>
</tr>
<tr>
<td></td>
<td>493</td>
<td>461</td>
<td>330</td>
<td>2703</td>
<td>4063</td>
</tr>
<tr>
<td></td>
<td>31.1(5.6)</td>
<td>25.7(7.7)</td>
<td>14.1(10.1)</td>
<td>18.7(10.2)</td>
<td>20.7(10.6)</td>
</tr>
<tr>
<td>MADRS COMP</td>
<td>.642</td>
<td>.746</td>
<td>.839</td>
<td>.833</td>
<td>.840</td>
</tr>
<tr>
<td></td>
<td>493</td>
<td>461</td>
<td>330</td>
<td>2703</td>
<td>4063</td>
</tr>
<tr>
<td></td>
<td>31.2(7.9)</td>
<td>23.6(9.5)</td>
<td>15.0(11.0)</td>
<td>18.9(11.2)</td>
<td>20.8(11.6)</td>
</tr>
</tbody>
</table>

Conclusions: Guidance on the development of drug development tools makes clear the need for scale performance data. This analysis shows good reliability for the MADRS at all study visits after baseline. Close agreement between MADRS COMP and MADRS SBR across study visits and therapeutic indications during actual clinical trials supports the validity and reliability of computer-administered MADRS in studies of Unipolar and Bipolar Depression. The MADRS SBR and MADRS COMP may be useful benchmarks against which to judge the performance of other scales proposed for use in global multisite studies. The similar drop in performance observed for both the MADRS SBR and MADRS COMP at study baseline visits suggests that factors other than rater behavior may play an important role.
undercutting the reliability of baseline scores. Strengths and weaknesses of MADRS\textsubscript{SBR} and MADRS\textsubscript{COMP} will be considered.

References:


14 Adaptive Group Sequential Design for CNS Clinical Trials and Early Stopping Due to Slow Recruitment Rate and High Dropout Rate
Radhi F. Abdulnabi
i3Statprobe, Ann Arbor, MI, USA

Background: In clinical trial terms, an adaptive design is a design that can be modified based on the results of an interim analysis, without affecting the validity and integrity of the trial. One of the commonly used adaptive designs is a group sequential design. A group sequential design allows a trial to be stopped early if the treatment under consideration is futile (ineffective) or extremely effective, with the option of additional adaptations (e.g., sample size re-estimation) based on the results of an interim analysis.

Methods: A group sequential design is a multi-stage design with various stopping boundaries that are based on different boundary functions for controlling an overall type I error rate. In addition to prematurely stopping a trial due to safety, futility, efficacy, or all of these, as required by adaptive design rules, many clinical trials can be stopped for other reasons, such as slow recruitment of patients. In general, recruitment rates in clinical trials of central nervous system (CNS) drugs have been reported to be slow. Another reason for stopping a trial is for high dropout rates that can be especially high in a placebo group. Dropout rates in randomized clinical trials of CNS drugs have also consistently been reported to be high. Early stopping of a trial due to a low recruitment rate and/or a high dropout rate affects a group sequential design and therefore, clinical researchers will need to consider adjusting the planned design and the sample size estimation. SAS sequential design procedure (PROC SEQDESIGN) should be used to estimate the sample size and, sequential test procedure (PROC SEQTEST) should be used to estimate the treatment response.

Results: As a result of slow recruitment and a high dropout rate, the clinical trial missed its deadlines and therefore, led to the sponsor deciding to stop the trial. In this poster we propose the following two solutions:
- Changing the design assumptions, by reducing the number of patients for each stage
- Re-estimating the sample size? Using the newly-obtained trial data.

Conclusions: Low recruitment and high dropout rates in clinical trials can significantly reduce the sample size and power, resulting in a sponsor stopping the trial prematurely. As clinical trials for new drugs become more complex, pharmaceutical companies must use more efficient statistical methods to estimate sample sizes. Group sequential design is one of the commonly used adaptive designs to re-estimate a sample size.

15 The Power of Expectation Bias
Janet B.W. Williams\textsuperscript{1,2}, D. Popp\textsuperscript{3}, Kenneth A. Kobak\textsuperscript{3}, M Detke\textsuperscript{1,4}
\textsuperscript{1}MedAvante, Inc., Hamilton, NJ, USA; \textsuperscript{2}College of Physicians and Surgeons, Columbia University, New York, NY, USA; \textsuperscript{3}Center for Psychological Consultation, Madison, WI, USA; \textsuperscript{4}Indiana University School of Medicine, Indianapolis, IN, USA

Introduction: Expectation bias occurs when an individual’s expectations about a desired outcome influence one’s perceptions of one’s own or others’ behavior. Expectation bias can be an important contributor to placebo response in a treatment study. In psychiatry, expectation bias can have a significant effect on the outcome of a clinical trial. Subjects enter a trial expecting to improve; clinicians hope and expect that their subjects will improve. Double blind studies are designed to control for expectation bias by blinding the subject and the clinical assessor to whether the subject is taking placebo or drug. However, other unblinded factors, such as visit sequence and assessor-subject relationship, may also affect placebo response. One solution to reduce expectation bias is to use raters who are blinded to study protocol and visit number, and who are independent from the clinical trial site. Blinding to protocol details and study visit number eliminates the possibility that clinical ratings will be affected by an expectation of improvement as treatment progresses. Independence from the study site controls for the possibility of a relationship forming between rater and subject that could influence ratings (sometimes referred to as relationship bias).

Methods: In a study of acute Schizophrenia blinded independent raters evaluated subjects on the PANSS. Traditional
site raters (unblinded to study visit) evaluated subjects on the BPRS within 1 to 2 days of the blinded raters’ assessments. A study of psychosis in Parkinson’s Disease included patients rated by blinded independent raters in the US, and traditional unblinded, site-based raters in ex-US sites.

**Results:** In the Schizophrenia study, the blinded independent raters observed statistically significant differences between subjects on placebo, the active comparator, and one of two test arms at every time point. Site raters did not distinguish either of the two test arms at any time point. The blinded independent raters in the Parkinson’s psychosis study found a significant difference between placebo and drug from baseline to endpoint psychosis scores (p<.05) at week 2 and a trend (p<.09) at week 5 in the US sample, while no efficacy was detected by site raters outside of the US.

**Conclusions:** Expectation bias may have an effect when raters’ scores and subjects’ reports are influenced by the expectation of improvement. Results suggest the use of site-independent raters who are blinded to protocol details and study visit may yield better signal detection. One or more authors report potential conflicts which are described in the program.

### 16 The Impact of Site Characteristics on Efficacy Measures

Bethanne Friedmann¹, Erin B. Kornsey¹, Kathryn Dawson¹, Henry J. Riordan¹, Michael F. Murphy¹, Neal R. Cutler²

¹ Worldwide Clinical Trials, King of Prussia, PA, USA; ² Worldwide Clinical Trials, Beverly Hills, CA, USA

**Introduction:** The success of psychiatric trials hinges on the quality of data collected by sites. Data errors and discrepancies between the key efficacy measures of a double-blind, randomized, placebo-controlled, multicenter, phase II clinical trial designed for adult attention deficit/hyperactivity disorder (ADHD), were closely monitored by a panel of clinical experts.

**Methods:** Data “flags” were based on various scales including the Clinical Global Impression of Severity (CGI-S), Clinical Global Impression of Improvement (CGI-I), Conners’ Adult ADHD Rating Scale – Observer: Screening Version (CAARS-O:SV), and Conners’ Adult ADHD Rating Scale –Self Report: Short Version (CAARS-S:S). Reports were generated twice a week based on information sites electronically entered since the last report was sent. These reports contained discrepancies between the scales, data entry errors, and rater errors. The clinical manager contacted the sites to gather more information about these data flags via telephone and email.

**Results:** Sixteen sites with randomized subjects for the trial and were included for analysis (mean number of flags 11.38, standard deviation 12.47). The overall number of data flags per site was negatively correlated with both the number of subjects screened (r = -0.45) and number of subjects randomized per site (r = -0.19). Prior studies have suggested that the number of flags per site were proportional to the number of subjects enrolled per site. Because there is evidence this was not the case, sites were divided into 2 categories: sites with a high rate of flags per randomized subject and sites with a low rate of flags per randomized subject. A significant difference in the mean number of subjects randomized by the data flag rates was determined (t = 2.43, p = 0.03) with a higher mean number of randomized subjects for site with lower flag rates. The severity of the data flags were also assigned numeric ratings from 1 (least severe, e.g. data entry error, missing data needs to be entered; n = 116), 2 (moderately severe, e.g. possible incorrect rater completing assessment; n= 59), to 3 (most severe, possible scale discordance; n = 7). This has important implications for the appraisal of sites with lower patient screen and randomization numbers.

**Conclusions:** By investigating and tracking the frequency and severity of the various flags over the course of the study it is possible to enhance the overall quality of data and ultimately lead to increased effect sizes.

**Disclosure:** No conflicts of interest exist in the research and development of this poster.

### 17 Does the Correlation between CGI, the HAM-D and the MADRS Improve When the Same Rater Administers all Three Scales?

Theresa A. Bromley, Nicolas Troyana, Marian Ormont
ePharmaSolutions, Conshohocken, PA, USA

**Introduction:** The Clinician Global Impression (CGI) is brief assessment of a patients overall symptoms and severity of the disease under study. The second part of the CGI is an assessment of the patient’s improvement in symptoms and functioning over time. The CGI is frequently used in CNS trials. This scale is typically administered by the most experienced clinician who is familiar with the disease indication, the collaborating information about the subject from other scales and study data and has had the opportunity to interview the patients. In some studies, the CGI is blinded to all study data and asked to independently assess the severity and improvement of a patient’s symptoms and functioning. Often the CGI rater is not the same rater as other measures in the study due to the increased rigor placed around minimum education and experience criteria for the CGI rater. In this study, The MADRS, HAM-D and CGI were given
over 10 study visits. Six hundred and thirty-eight visits were included in this study. This study predicted that when the CGI rater also administered the HAM-D and MADRS scales, the correlations between the ratings would improve. Preliminary results indicate that having the same rater conduct all three interviews does improve the correlation between the scale scores.

**Methods:** The scores for the MADRS, HAM-D and CGI were collected for 638 study visits. Eight-two raters and 175 different subjects were included in this study. CGI raters were required to have a doctorate degree and to serve as the PI or SubI. Results: Correlations between total MADRS, HAM-D and CGI-Severity and CGI-Improvement scales were examined. Results indicate that when the same rater conducted all three interviews the correlations between the three scales were significantly better than when an independent rater conducted only the CGI scale.

**Conclusions:** The CGI has been demonstrated to be sensitive to change and a robust measure of patient’s severity. Some criticism of the CGI has included its lack of structure and the in-depth knowledge of the patient required to rate this scale. When the rater interviews the patients to complete other assessments, that rater may be more accurate in rating the CGI scale.

18 **A Comparison of the Performance of Japanese and English Volunteers on the CDR System**
Keith A. Wesnes
United BioSource Corporation, Goring on Thames, UK

**Introduction:** The CDR System is a set of automated tests of cognitive function which has been widely used in worldwide Phase I studies. The CDR System comprises tests of attention, vigilance, working memory, and episodic memory. The purpose of the present analysis was to determine the comparability of the Japanese version of the CDR System data to the original English version.

**Methods:** In this analysis, 91 male Japanese volunteers aged 20 to 35 were compared to 1146 male English volunteers in the same age range, on a range of CDR System tests: simple and choice reaction time; digit vigilance; numeric working memory; immediate and delayed word recall; word recognition and picture recognition. All volunteers had participated in Phase I studies, and had been trained on four occasions on the CDR System tests prior to the first study day. The performance on the four training sessions and the first study day pre-dosing baseline was compared between the two populations. Test retest reliability was assessed, and Principal Components analysis with varimax rotation was conducted to compare the factor structure of System between the two populations.

**Results:** There were no consistent or notable differences between the two populations over the five testing sessions for any of the measures. For measures which showed training effects, performance in both populations had stabilised after the initial two or three training sessions. The test retest reliability of the measures showed the same patterns over the measures between the two populations. Finally, in both populations, factor analysis of the attention and episodic memory tasks identified 4 factors with Eigen values greater than 1. Varimax rotation, showed that the same measures loaded on four common factors in both populations, as has been seen in previous work.

**Conclusions:** The various language versions of word learning tasks on the CDR System are created individually in each language based on rules of length, frequency and imageability. Previous work in elderly stroke patients has shown equivalence of Korean and Chinese word lists to various European languages. This study extends this language equivalence to Japanese. The similar patterns of data on the various other measures between the two language versions, plus the comparable factor structure further confirms the equivalence of Japanese version of the CDR System to the English version.

19 **Is an In-Study Surveillance Program Effective at Reducing Error Rates for Both Experienced and Novice Raters?**
David S. Miller, Cynthia McNamara, Priscilla Samuelson, Daniella Mulder, Amanda Young
United BioSource Corporation, Wayne, PA, USA

**Aim:** Challenges exist to adequately identify, train and certify potential investigators and to ensure that they properly administer and score the efficacy instruments over the course of global Alzheimer's disease clinical trials. Previous research has shown that when less experienced raters are given enhanced pre-investigator's meeting (IM) training (previously described) on the Alzheimer's Disease Assessment Scale - Cognitive subscale (ADAS-Cog), they certify to participate at rates equivalent to their more experienced colleagues. We evaluated the impact of an In-Study rater surveillance program, which identifies rater errors in scale administration and scoring, and remediates raters when those errors occur, on the raters' subsequent performance in the course of a clinical trial. The performance of those raters who required and received the enhanced training was compared to their more experienced colleagues.

**Methods** Two separate, multi-national Alzheimer's disease clinical trials where such a customized surveillance program...
was employed were evaluated to assess the program's impact. All raters received extensive didactic training on the proper administration and scoring of the ADAS-Cog and also had to demonstrate their competence to rate by successfully watching and scoring a video of the scale being administered to a patient. Rater performance (proper ADAS-Cog administration and scoring) was assessed by worksheet review at 2 time points - baseline and 1 year. At each time point under consideration, a review of the ADAS-Cog worksheets occurred for each patient visit. Worksheets were assessed for errors by a calibrated Clinician. Error type was further characterized as being either clerical in nature or reflective of a deviation in the scale administration and/or scoring conventions. In both cases, raters were remediated when an error was identified.

**Results:** Of the 1131 certified, participating raters, 214 (19%) required enriched training, while 917 (81%) did not. Overall, 62% fewer raters required a support contact at 1 year than at baseline. This reduction was a statistically significant (p <0.001). At each time point, there was no statistically significant difference in the percentage of raters from each group, enriched and non-enriched, who required a support contact (27.1% vs. 23.7% at baseline, and 12.1% vs. 8.6% at 1 year).

**Outcomes of Clinical Trials from USA and Europe – A Meta-Analytic Comparison of Double-Blind Randomized Multi-Center Studies in Restless Legs Syndrome.**

Ralf Kohnen¹, Hanna Scholz², Claudia Trenkwalder³, Magdolna Hornyak²

¹Research Pharmaceutical Services Inc., Fort Washington, PA, USA and University of Erlangen-Nuremberg, Nuremberg, Germany; ²Interdisciplinary Pain Centre, University Medical Centre, Freiburg, Germany; ³Paracelsus-Elena Klinik, Centre for Parkinson Syndromes and Movement Disorders, Kassel and Department of Clinical Neurophysiology, University of Goettingen, Germany

**Introduction:** Restless Legs Syndrome (RLS) is a frequent and mostly chronic neurological disease leading to substantial impairment in sleep, mood, and daily life. Epidemiological data show that up to 10% of the general population is afflicted by RLS, and 1% to 3% are in need of drug treatment. Dopamine agonists are recommended as first-line treatment for the disorder. We performed the most comprehensive meta-analyses to date investigating the efficacy and safety of dopamine agonists in RLS. In this poster, we compare studies performed in the USA with those conducted in Europe with regard to changes in the International RLS Study Group Severity Rating Scale (IRLS), the gold standard of patient reported outcome measures in RLS research.

**Methods:** We included double-blind randomized controlled trials (RCTs) of dopamine agonist treatment versus placebo or versus other drug treatment for RLS with levodopa. Findings. We identified 35 placebo controlled and 3 active controlled RCTs investigating overall 7365 patients (mean age 55 years, 64% females). Of 30 studies using the IRLS, 17 were conducted in Europe and 12 in the USA with a total of 6380 patients. Studies compared placebo or levodopa to oral treatment with either cabergoline, pergolide, pramipexole, ropinirole or with the patches rotigotine and lisuride. The likelihood for bias in the meta-analyses was considered low and results showed low to substantial heterogeneity. Treatment duration was 9.8 ± 7.2 weeks with few treatment durations up to seven months. Overall, dopamine agonists compared to placebo improved RLS symptoms on the IRLS by -5.7 points (95% CI -8.23 to -5.31) than in trials conducted in USA (Mean Difference -4.13; 95% CI: -5.31 to -2.95).

**Conclusions:** Dopamine agonists are effective in the treatment of restless legs syndrome by improving symptom severity in trials conducted both in the USA and Europe. Differences between dopamine agonists and placebo or active control were larger in Europe. The following reasons for these differences will be discussed: (1) the most effective drugs, cabergoline, pergolide, lisuride and rotigotine, were mainly investigated in European trials, (2) patients in Europe were mainly recruited from clients of neurological hospitals and private practices whereas the majority of patients in USA were recruited via advertisements, (3) European investigators were more experienced in RLS than those in USA. Our results are of interest for globalization of clinical research.

**Comparing Measures of Negative Symptoms of Schizophrenia in Clinical Trials: The Investigators’ View**

David G. Daniel¹, Dawn I. Velligan², Nicholas Greco IV³, John J. Bartko⁴

¹United BioSource Corporation, Mclean, VA, USA; ²UTHSCSA, San Antonio, TX, USA; ³Abbott Laboratories, Abbott Park, IL, USA; ⁴Independent Consultant, Newville, PA, USA

**Introduction:** The Scale for the Assessment of Negative Symptoms (SANS), the Negative Symptom Assessment Scale (NSA-16), and subscales from the Positive and Negative Symptoms Scale (PANSS) vary in length, complexity and the domains of negative symptoms evaluated. All three scales are reliable and valid measures of negative symptoms for clinical trials. However, little recent direct comparative data is available on raters’ views of the three scales.

**Method:** Thirty nine raters participating in an industry sponsored schizophrenia clinical trial completed a survey
ascertaining experience with the scales (1-no experience to 8 - greater than 5 years experience), critiquing the clarity of the anchor points (0- not at all clear to 5- extremely clear), how well each scale measured the concept of negative symptoms (0- very ineffectively to 5 -very effectively) and the rater’s preference among the scales.

**Results:** 97% 93% and 81% of the raters were experienced using the PANSS, SANS and NSA-16, respectively. Only the raters experienced in using each scale were included in the following analyses. Only 17.2%, 27.6% and 29.6%, respectively, of the raters felt the PANSS, SANS and NSA-16 were “very effective” in measuring negative symptoms. 3.4%, 0 % and 0% respectively of the raters felt the PANSS, SANS and NSA-16 were “very ineffective” in measuring negative symptoms. With respect to their effectiveness in measuring negative symptoms, mean scores for the PANSS, SANS and NSA-16 were 3.59, 3.97 and 4.19 respectively, (F (2,82) = 3.53, p < 0.03). 10.3%, 7.4% and 10.7% of the raters, respectively, felt the PANSS, SANS and NSA-16 anchor points respectively were “very clear”. 6.9% 0% and 0%, respectively felt the PANSS, SANS and NSA-16 anchor points were “very unclear”. With respect to clarity of the anchor points, mean scores for the PANSS, SANS and NSA-16 were 3.66, 3.67 and 3.79 respectively, (F(2,81) = 0.29, p < 0.75 (NS). Of the 21 raters who used all three scales, there was equal preference among them (33.3%, 33.3% and 33.3%) respectively for PANSS, SANS and NSA-16.

**Conclusions:** In this survey of 39 clinical trials raters, all three scales were judged to be relatively effective in rating negative symptoms, but the PANSS was felt to be significantly less effective in rating negative symptoms than the SANS or NSA-16. There were no significant differences among raters with respect to their perception of the relative clarity of the anchor points of the three scales only a relatively small minority viewed the anchor point of any of the scales as “very clear”. The small sample size and possibility of sample bias limit the conclusions that may be drawn from this analysis.

**Disclosure:** The statistical analyses for this study were paid for by United BioSource Corporation.

**References:**
Andreasen, N.C., 1981. Scale for the Assessment of Negative Symptoms (SANS). University of Iowa, Iowa City, Iowa.

### 22 Open Label Trial of Carbamazepine-ER for Bipolar Depression Using the Interactive Computer Interview for Outcome Assessment

Lori Davis1,2, Xiaohua Li2, Andrew Leon3, Dan Debonis4, Gary Sachs5

1VA Medical Center, Tuscaloosa, AL, USA; 2Department of Psychiatry, University of Alabama School of Medicine, Birmingham, AL, USA; 3Department of Psychiatry, Weill Cornell Medical College, New York, NY, USA; 4United BioSource Corporation, Chevy Chase, MD, USA; 5Department of Psychiatry, Harvard Medical School, Boston, MA, USA

**Introduction:** Patients with bipolar disorder spend significantly more time depressed than manic. Several small open-label and double-blind trials have suggested antidepressant effects of immediate-release carbamazepine in acute bipolar depression. Extended-release capsulated carbamazepine (ERC-CBZ; Equetro) is FDA-approved for the treatment of bipolar acute mania and mixed episodes. As a next step in understanding its broader efficacy, we evaluated the antidepressant effects of ERC-CBZ in the treatment of bipolar I or II depression using a novel approach to the primary outcome assessment, i.e. the interactive computer interview (ICI). Methodological and technological advantages of ICI rating for bipolar depression are discussed in this poster presentation.

**Methods:** Subjects with type I or II bipolar depression were prospectively enrolled in an 8-week, open label, monotherapy trial of ERC-CBZ. Subjects were assessed biweekly using an ICI rating system, called the Rater Station™, to simulate and calculate the Montgomery-Asberg Depression Rating Scale (MADRS) as the primary outcome and the Young Mania Rating Scale (YMRS) as a safety outcome. After study personnel logged into the customized laptop computer, the Rater Station™ guided subjects to answer questions from a yes/no or multiple choice format. The computer selected follow-up questions and used a scoring algorithm to map subject responses to rating scale anchor points have been validated in comparison with experienced human raters administering the MADRS and YMRS. Research staff provided instructions at the first visit and written and computerized instructions were made available at each visit. Subjects answered the ICI questions independently, i.e. without prompting by an investigator or research assistant. The Hamilton Rating Scale for Anxiety, and Clinical Global Impression-Bipolar (CGI-BP) were clinician administered and evaluated as secondary outcomes.

**Results:** Thirty-two evaluable subjects were considered in the efficacy analysis. Paired samples t-tests were used to compare changes in scale scores from baseline to week 8. The MADRS revealed a significant improvement in depression between the baseline and week 8 assessments (p < .001). Except for CGI-BP mania severity scores all secondary outcome measures showed significant improvements. Subjects easily completed the ICI assessment and there were no missing data due to technological difficulties.
Conclusions: Extended-release carbamazepine is well tolerated and significantly improves depression in patients with bipolar disorder. The ICI rating system offers a novel approach to outcome assessments. Serving as a pseudo-single blind approach, the ICI interview improves objectivity and may improve efficiency by reducing rater training.

23 Diagnosis Criteria, Inclusion and Outcomes: An Analysis of Criteria Beyond DSM and Disease Severity in Ziprasidone Bipolar Mania Trials

Francine S. Mandel\textsuperscript{1}, Douglas Vanderburg\textsuperscript{1}, Gary Sachs\textsuperscript{2}

\textsuperscript{1}Pfizer Inc. New York, NY, USA; \textsuperscript{2}Massachusetts General Hospital, Boston, MA, USA

Introduction: The implication of many protocol specified inclusion criteria in bipolar mania trials are not well understood. Generally studies specify baseline disease severity per scores on rating scales and require subjects meet current DSM-IV criteria for mania. We examined three additional inclusion criteria and their relationship to each other.

Methods: Adult subjects from 2 double-blind trials with bipolar 1 disorder, YMRS scores >14, and scores >2 on at least 4 items at screening and baseline received ziprasidone (80-160 mg/d) or placebo for 21 days. The present post hoc analysis examines three additional inclusion criteria thought to correspond to bona fide patients with bipolar mania seen in routine clinical practice: #1: age at onset (< 30 years), #2: number of prior hospitalizations (>2) and #3: elevated mood and/or grandiosity (score >3). Additionally, a “marginal inclusion bracket” was evaluated to eliminate potential inflation of disease severity at baseline (YMRS 14-17). The distribution and relationship among these 4 factors were examined.

Results: Based on data entered by site-based personnel 295/403 (73.2\%) of subjects met inclusion criterion #1, 281/402 (69.9\%) met #2, and 150/404 (37.1\%) met #3. In addition, 7.9\% met none of the 3 measured criteria at baseline, 24.5\% met 1 criterion, 47.5\% met 2 criteria and 20\% met all 3 criteria. There was a significant relationship (P<0.0001) between average baseline YMRS score and the number of additional criteria met (linear regression equation estimated baseline YMRS = 21.36 + 2.98 x number of inclusion criteria). Thus, for each additional criterion met, an increase of 3 points in baseline YMRS score was seen. Forty six subjects (11.4\%) in these studies were within 4 points of the minimum YMRS score at baseline (marginal inclusion bracket). No subjects met all 3 criteria, whereas 22.9\% of the subjects not in the marginal inclusion bracket (score > 4 points above the minimum MRS score) met all 3 criteria for diagnosis. There was a significant interaction between elevated mood/grandiosity and ziprasidone treatment (P<0.014) but no interaction for criteria #1 or #2.

Conclusions: This analysis demonstrates that selecting patients for whom diagnostic confidence is higher may find a sample with higher baseline severity as measured by the YMRS. This may be an important consideration for study design in bipolar mania clinical trials. Information on the impact of these additional criteria on outcomes will be presented in the poster.

Disclosure: Funded by Pfizer, Inc.