

ISCTM 2011 Autumn Conference
3-4 October 2011
The Ritz-Carlton, Amelia Island

Poster Abstracts

Classification:

Statistical: 1-3

Strategic: 4-17

Clinical: 18-23

1 Statistical Considerations for the Analysis of Suicidal Ideation and Behavior Events in Clinical Trials

Gassmann-Mayer C¹, Nilsson M², Suryawanshi S³, McSorley P⁴, DuBrava S⁵, Jiang K⁶, Arani R⁷

¹Johnson & Johnson Pharmaceutical Research & Development, L.L.C., ²Eli Lilly & Company, ³Merck & Co. Inc, ⁴GlaxoSmithKline, ⁵Pfizer Inc, ⁶Sanofi-Aventis, ⁷Axiom Consulting Group, L.L.C.

Introduction: Suicidal ideation and behavior are commonly associated with certain types of mental illness such as major depression, bipolar disorder, and schizophrenia. In the last two decades, the potential association between an increased risk of suicidal ideation and behavior and use of pharmaceuticals has been the center of intense debate in medical research. More recently, analysis of suicidal ideation and behavior has been conducted on compounds used to treat illnesses outside psychiatric or neurologic areas, such as beta blockers, drugs for smoking cessation and weight loss. The FDA is developing guidelines for the prospective collection and analyses of suicide-related data in drug development. Prospective scales to collect suicide-related events, such as the Columbia Suicide Severity Rating Scale (C-SSRS), have become common instruments in the suicide research field. However, much still needs to be done to reach a consensus opinion on the analytical and statistical methods for the analyses and reporting of suicide-related events.

Methods: A cross-industry statistical team was originally assembled in 2008 within the Pharmaceutical Research and Manufacturers of America (PhRMA) suicidality working group, but afterwards convened independently to work on a scoring and statistical analysis guide for suicide-related events. One of the major challenges is that suicide-related events are relatively rare and the associated risks complex to quantify in individual clinical trials.

Results: This cross-industry team is developing a suggested “statistical analysis plan” focused on the C SSRS instrument, but the proposed statistical ideas and definitions are also applicable to other prospective scales. This statistical and analytical guide addresses the challenges of the definition of suicide-related composite statistical endpoints, treatment emergence, history/baseline reference, and the use of different statistical methods in the analysis of single-trial data. Additional medical input is required around the assumption of ideation in presence of observed suicidal behavior, and lack of a clinically interpretable hierarchy among suicidal behaviors.

Conclusions: This work illustrates a proposal from an industry statistical team to address the current challenges in the statistical analysis and reporting of prospectively collected data on suicidal ideation and behavior in clinical trials. The proposed statistical analysis guideline will evolve as the scientific debate continues among academia, pharmaceutical industry, and regulatory agencies to bring further clarity to the remaining unresolved issues. The development of broadly accepted standard definitions and use of different statistical approaches are necessary to better interpret suicide-related data and quantify the associated risks in the clinical evaluation of pharmaceuticals.

Disclosure: The authors report no conflicts of interest for this work while being employees and stockholders in their respective companies. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

2 Role of suicide history as a moderator of antipsychotic effect on depressive symptoms in schizoaffective disorder: a path analysis

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

¹Janssen Pharmaceutical Research and Development, LLC, Titusville, NJ, USA, ²Janssen Scientific Affairs, LLC, Titusville, NJ, USA

Introduction: When managing patients with schizoaffective disorder, it is difficult to determine whether improvement observed in depressive symptoms during treatment with antipsychotics occurs as a direct effect or indirectly through improvement in other symptom domains. This analysis examined this association and the impact of suicidal history as a moderator in subjects with schizoaffective disorder.

Methods: Post hoc database (N=614) analysis of two 6-week, randomized, placebo-controlled studies of antipsychotic treatment vs placebo in subjects with schizoaffective disorder. Subjects with baseline depressive symptoms (HAM-D-17 ≥ 16) were included. Structural equation models (path analysis) were used to separate total effects of antipsychotic treatment into direct and indirect effects on depressive symptoms. Change from baseline in HAM-D-17 score at the week 6 endpoint was the dependent variable; change in PANSS positive and negative factors and SAS (extrapyramidal symptom) scores at the week 6 endpoint were independent variables (possible mediators of depressive symptom changes). In each structural model equation, a factor for treatment was included for comparisons between antipsychotic treatment and placebo. Potential moderators of antipsychotic effect (ie, age, race, gender, suicide history, hospitalization, substance abuse, co-medication strata [antidepressants and/or mood stabilizers], and schizoaffective subtype) were analyzed by multiple regression models. Significant baseline moderators identified were used as additional factors in the path analysis.

Results: 333 subjects met inclusion criteria. Initial path analysis determined that 46.1% and 28.2% of treatment vs placebo effect on depressive symptoms was indirectly attributed to improvements in positive and negative symptoms, respectively; up to 26.3% was attributed to a direct treatment effect. Among potential moderators, a history of suicide attempts was found to be significantly associated with changes in depressive symptoms. For subjects with a suicide attempt history, treatment vs placebo improvement in HAM-D-17 was indirectly mediated through effects on positive (30.5%) and negative (19.4%) symptoms; up to 50.5% of the total effect represented a direct treatment effect. For subjects without suicide attempt history, improvement in HAM-D-17 was indirectly mediated through improvement in positive (60.5%) and negative (38.5%) symptoms; up to 1.8% of the total effect represented a direct treatment effect. Treatment effect was not attributed to extrapyramidal symptom changes in any model.

Conclusion: Results suggest that antipsychotic treatment may have a direct effect on depressive symptoms in subjects with schizoaffective disorder, with indirect effects mediated through positive and negative symptom changes. These effects appeared strongly modulated by a suicide attempt history.

Supported by 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 Janssen Scientific Affairs, LLC, and Johnson & Johnson stockholders.

3 Item Analysis of changes in the reliability and validity of scores on the Alzheimer's disease Assessment Scale (ADAS-Cog) and Mini Mental State Examination (MMSE) across time

Yavorsky C^{1,2}, DiClemente G², Opler M^{1,3}, Khan A^{1,5,6}, Liechti S¹, DeFries A^{1,4}, Jovic S¹

¹ProPhase, LLC, ²CROnos CCS, ³New York University School of Medicine, ⁴Johns Hopkins University Bloomberg School of Public Health, ⁵Nathan S. Kline Institute for Psychiatric Research, ⁶Manhattan Psychiatric Center

Introduction: For the assessment of Alzheimer's disease (AD), the most frequently used scales are the Alzheimer's Disease Assessment Scale (ADAS-Cog) and the Mini-Mental-Status Examination (MMSE). Though ADAS-Cog and MMSE publications demonstrate validity and reliability, the retention of psychometric traits across study visits has proven difficult in longitudinal multicenter trials. Despite intensive rater training there are concerns about their longitudinal validity and reliability. The goal of this study is to model the longitudinal changes in ADAS-Cog scores compared to changes in MMSE scores across time.

Methods: 2,618 subjects with AD participated in 1 of 7 studies obtained from the Critical Path Institute

Online Data Repository (CODR). Reliability was assessed by the intra class correlations (ICC) and Cronbach alpha (α). Cross-sectional construct validity was tested through Pearson correlations of the ADAS-Cog Test to the MMSE total/item scores. Item-total statistics measured the relationship of individual ADAS-Cog items to the MMSE. Data were stratified according to screening MMSE scores, as a marker of disease severity (mild AD = 21 - 26, moderate AD \leq 20).

Results: Mean age was 74.20 (SD = 8.13; range = 50-97) years, with 55.46% males and 44.54% females. Illness duration ranged from < 1 month to 136 months. Cronbach's α and test-retest intra class correlation coefficients for the ADAS-cog scale were moderate to high between baseline and subsequent visits (Visit 2, 0.895 and Visit 7, 0.666 (early and late visits within the trials), supporting their reliability. However, they show a progressive decrease in reliability. Visit level ICCs range 0.392 (Visit 7) to 0.806 (Visit 2) again showing a progressive decrease in correlations across time. There was also a progressive decrease in item correlations, with the lowest correlations for Visit 7 observed for Commands (ICC = 0.148), Comprehension (ICC = 0.092) and Spoken Language (ICC = 0.044). Correlations between ADAS-cog and MMSE scores across visits were moderate to high ($r = -.576$ to $r = -.732$) for the MMSE \leq 20 group and ($r = -.396$ to $r = -.811$) for the MMSE \geq 21 group.

Conclusions: The findings suggest that there is a decrease in reliability over time between the total and item scores on the ADAS-Cog. Our study highlights the significance of fully testing measures and rater competencies prior to, and throughout treatment trials. In particular, it emphasizes the value of training raters across the continuum of AD severity and guiding raters on scale and item level functioning across longitudinal studies.

4 Detecting Antidepressant Efficacy: The Impact of Study Design and the Value of Academic Sites

Boadie W. Dunlop, MD¹, Michael E. Thase, MD², Chuan-Chuan Wun, PhD³, Rana Fayyad, PhD³, Christine Guico-Pabia, MD³, Jeff Musgnung, MS³, Philip T. Ninan, MD³

¹Emory University, Department of Psychiatry and Behavioral Sciences, Atlanta, GA, USA; ²University of Pennsylvania, Department of Psychiatry; Philadelphia, PA, USA ³Pfizer Inc, Collegeville, PA, USA

Introduction: Placebo response, defined as demonstrable symptomatic improvement despite inactive treatment, varies widely in clinical trials employing subjective outcomes, particularly in psychiatric disorders. In antidepressant studies, placebo effects complicate the detection of true treatment effects and appear to be increasing over time. This meta-analysis of placebo-controlled studies of venlafaxine and desvenlafaxine (administered as desvenlafaxine succinate) was conducted to identify trial and patient characteristics that may impact placebo response and signal detection in studies of major depressive disorder (MDD).

Methods: Trial design characteristics and patient-level data from all Phase II-IV clinical trials of MDD performed by the manufacturers of venlafaxine and desvenlafaxine were analyzed. Trials were selected if they were conducted with adult outpatients, used a double-blind, placebo-controlled design, and were 6 to 12 weeks in duration. Trials limited to inpatients or geriatric patients were excluded. Data that could potentially affect treatment outcomes were analyzed. Study design characteristics included: number of post-baseline study visits, treatment arms, and assessments per visit, duration of the trial, fixed vs flexible dosing, and drug. Characteristics of the research environment included: year of study, percent of US and academic sites, and proportion of completers. Patient characteristics included: age, sex, race, and mean and median baseline MDD severity. Treatment outcomes included: differences in standardized mean Hamilton Depression Rating Scale (HAM-D) score and the risk ratio for response (i.e., $\geq 50\%$ reduction in HAM-D at endpoint) between antidepressant and placebo, mean change in HAM-D and response in placebo patients, and study outcome (significant drug-placebo separation at endpoint). Univariate meta-regression models assessed the effect of predictor variables that may be associated with the defined outcomes. To adjust for confounding effects among predictor variables, multivariate meta-regression models were also performed.

Results: Thirty trials that comprised 8933 depressed patients were included. Drug-placebo separation was most strongly predicted ($\beta = 3.74$, $P=0.0002$) by the proportion of academic sites in the trial. Other factors predicting significant drug-placebo separation included lower completion rate, greater number of post-baseline visits, earlier year of study, greater percentage of whites, and studies conducted with venlafaxine vs desvenlafaxine. Lower median baseline depression severity predicted significant separation in response rates. Only the proportion of academic sites stayed as a significant predictor in the multivariate meta-regression modeling for both continuous

change scores and response rates.

Conclusions: Academic sites provide significant value to clinical trials of psychiatric disorders. Recent efforts to reduce placebo response through increasing minimum severity criteria may be misguided.

5 Study Site Experiences & Attitudes Toward Prospective Assessments of Suicidal Ideation and Behavior in Clinical Trials: Results of an Internet-based Survey

Authors: ISCTM Suicidal Ideation and Behavior Assessment Working Group; Adam Butler¹, Michelle Stewart², (Co-chairs) and members Larry Alphs³, Phillip Chappell², Douglas Feltner⁴, William Lenderking¹, Atul R Mahableshwarkar⁵, and Clare Makumi⁶

¹United BioSource Corporation, ²Pfizer, Inc., ³Janssen Scientific Affairs, ⁴Douglas E. Felmer, LLC, ⁵Takeda Global Research and Development, ⁶GlaxoSmithKline

Introduction: The recent release of the FDA draft guidance for the prospective assessment of suicidal ideation and behavior (SIB) has led to the widespread implementation of SIB assessments in clinical trials. The ISCTM SIB Working Group (WG) was formed to understand the effects of the guidance on clinical and research practice, and to identify challenges to its implementation. To build a better evidence base on what impact the inclusion of SIB assessments has had on clinical research sites, the WG designed a brief internet-based survey regarding site experiences and attitudes regarding these assessments.

Methods: A 20-item survey was developed based on WG discussions. An invitation to participate from the ISCTM Secretariat and a link to the survey were distributed via email to 6058 sites that had participated in at least one CNS clinical trial in the prior two years. Responses were collected through an internet-based survey form and summarized descriptively.

Results: A total of 1004 responses were collected over 3 weeks with 57% coming from outside the US. Respondents included principal investigators (35.8%), raters (28.0%), coordinators (25.4%), and others (10.7%). The majority (79.2%) had personally conducted prospective SIB assessments. Psychiatrists and neurologists (42.5% and 10.4% of respondents, respectively) had experience using these assessments across a wide spectrum of indications. Overall, respondents indicated that the benefit of prospective assessment is “worth the additional burden” (72%), that it has been “easy to incorporate” (72%) and has “improved subject safety” (74%). The greatest challenge was getting accurate baseline lifetime history (54.5%), while the greatest benefit was identifying subjects at risk of suicide (85%). Other implementation challenges and benefits were identified. Open-ended responses revealed specific challenges to the use of these assessments, particularly in cognitively impaired populations.

Conclusions: This study, to our knowledge the first to look at this topic, found that the inclusion of prospective SIB monitoring is generally viewed positively. Nevertheless, approximately a quarter of respondents reported important implementation challenges. Study limitations include internet-based survey methodology (self-report with no independent verification), recruitment of sites based on CNS trial experience, responses may not fully capture the sites’ total experiences, and no comparison to previous standard practice. These findings may help guide stakeholders’ use of SIB assessments in clinical trials. Additional analysis is warranted, e.g., how results may differ geographically or among sites with different demographics or training.

Disclosure: ISCTM supported the implementation of the survey with staff time and funding; the affiliations of all authors are given above; the authors report no other conflicts of interest.

6 Assessing the sources of unreliability (rater, subject, time-point) of placebo responders using items of the Positive and Negative Syndrome Scale (PANSS)

Khan A^{1,6}, Yavorsky C^{1,2}, Liechti S¹, DiClemente G², DeFries A^{1,4}, Opler M^{1,3}, Jovic S¹

¹ProPhase, LLC, ²CROnos CCS, ³New York University School of Medicine, ⁴Johns Hopkins University Bloomberg School of Public Health, ⁵Nathan S. Kline Institute for Psychiatric Research.

Introduction: Failed trials are problematic for the development of pharmacological treatments. Contributing factors include: escalating placebo response rates, dosing regimens, low sensitivity in clinical scores and high statistical variance. Interim monitoring of assessment tools, clinical outcomes, and measurement errors is important for early decision-making. The goal of this study is to identify sources of unreliability in a failed clinical trial by assessing scores for placebo responders on the Positive and Negative Syndrome Scale (PANSS).

Methods: This is a substudy from a failed Phase II double-blind, placebo-controlled schizophrenia trial. Using a unique statistical technique, Generalizability Theory, we assessed the extent to which each facet (raters, subjects, timepoints) contributed to the variability (and inconsistency) in PANSS scores.

Results: The placebo response rate was defined as $\geq 20\%$ improvement in PANSS total score. The placebo response rate was 21.13% (15 of 71 subjects). *Positive Subscale:* The most variability was observed for raters (range 33% - 72%), followed by variability of rater by timepoint where 5 items (Delusions, Conceptual Disorganization, Excitement, Grandiosity, Suspiciousness/Persecution) showed variability from 11% to 15%. The E_p^2 is similar to the internal reliability coefficient α and denotes the consistency in the relative PANSS item scores. The E_p^2 ranges from poor (Delusions) to excellent (Hostility), with an average of 0.727 across the 7 items. *Negative Subscale:* The most percent variability was for raters, ranging from 54% to 69%, followed by 4 items (Blunted Affect, Emotional Withdrawal, Poor Rapport, Passive/Apathetic Social Withdrawal) with percent variability from 12% to 17% for raters by timepoint. The scores for Difficulty in Abstract Thinking, Stereotyped Thinking and Blunted Affect show good to excellent reliability, whereas the remaining items are poor. *General Psychopathology:* 15 items showed the highest percent variability for raters, ranging from 33% to 70%. The E_p^2 ranges from poor (Tension, Anxiety, Somatic Concern, Unusual Thought Content, Preoccupation, Poor Impulse Control) to excellent (Disorientation, Depression, Motor Retardation, Guilt Feelings).

Conclusions: The significance of rater training for reliability is recognized as vital. Training conducted at initial investigator's meetings may not be effective enough for assessment of psychopathology in longitudinal trials. Our results confirm the efficacy of using a technique which is not dependent on sample size, i.e. Generalizability Theory, to identify sources of unreliability. Although further studies using samples from more than one trial is warranted, the current findings can be used to guide data monitoring, rater training and identification of PANSS items requiring supplementary training.

7 Evaluation Of A Rule Switching Test Designed To Assess Executive Control

Wesnes KA¹², Dries J³, Edgar CJ¹.

¹Bracket, Goring on Thames, UK, ²Swinburne University, Melbourne, ³Bracket, Wayne, USA

Introduction: Executive function is a set of cognitive abilities that control and regulate other aspects of function to facilitate goal-directed behaviour. One experimental paradigm for assessing aspects of executive control is requiring subjects to switch between two tasks. We evaluated a test in which such switching has previously been shown to involve activation of the medial and dorsolateral areas of the frontal cortex (DiGirolamo et al, 2001).

Methods: In this rule switching test (RST), strings of identical digits of varying length are presented on a computer screen (eg 888, 111111, 3333). In one condition, the subject has to determine whether the number of digits in each string is greater or smaller than 5, and in another condition whether the value of the digits is greater or smaller than 5. There are never 5 digits in the string and the digit 5 is not used. Each condition is associated with a particular colour which is presented before each string. The subject initially performs each condition separately in a block of 36 trials. Then in the 'switching' phase, the subject is required to apply the rule according to the colour presented prior to each successive string; the colours being presented in an unpredictable sequence, forcing the subject to decide which rule to apply trial by trial. In the present study, 41 healthy volunteers aged 21 to 55 years performed the RST on three study days alongside Parts A and B of the trail making test (TMT). Part A of the TMT involves a single rule, while Part B involves rule switching. Analysis of variance was used to compare the performance between the various phases of the RST and correlations were run to establish test-retest reliability and examine relationships between the two tests and performance with the age of the volunteers.

Results: In the RST, the subjects took significantly longer to respond in the switching phase than when performing the task using either one of the rules ($p < 0.001$), this performance change reflecting the demands of rule switching and providing an assessment of executive control. The test-retest reliability of the speed scores in the switching and non-switching conditions of the RST was good ($r = 0.58$ to 0.89), as was that for Parts A and B of the TMT ($r = 0.60$ to 0.87). There were also statistically significant correlations between the RST switching condition and Part B of the TMT ($r = 0.37$, $p < 0.02$), as well as between the extra response time in the switching condition of the RST and the longer completion time of the TMT Part B ($r = 0.32$, $p < 0.05$). Further, the extra response time in the switching condition of the RST correlated with the age of the volunteers ($r = 0.5$, $p < 0.001$), indicating that the test is sensitive to age-

related declines.

Conclusions: The RST appears to be suitable for repeatedly assessing executive control over time in clinical trials with volunteers, and work is underway to confirm these findings in clinical populations.

Reference: DiGirolamo GJ et al. (2001). General and task-specific frontal lobe recruitment in older adults during executive processes: A fMRI investigation of task-switching. *NeuroReport* 12: 2065-2071.

8 Suicidality Rating Scale Validation in Schizophrenia or Bipolar I Disorder Patients in Community Behavioral Health Organizations

Lian Mao¹, William Olson², Steve Ascher¹, Jessica Panish², John Fastenau², Larry Alphas²

¹Janssen Pharmaceutical Research and Development, LLC, Titusville, NJ, USA, ²Janssen Scientific Affairs, LLC, Titusville, NJ, USA

Objective: To evaluate the measurement properties of a suicidality rating scale, the International Suicide Prevention Trial (InterSePT) Scale for Suicidal Thinking (ISST)-Plus, in patients with schizophrenia or bipolar I disorder receiving treatment in community behavioral health organizations (CBHOs).

Methods: The ISST-Plus was administered in the Research and Evaluation of Antipsychotic Treatment in Community Behavioral Health Organizations Outcomes (REACH OUT) registry. This registry obtains information on the use of antipsychotics for the treatment of schizophrenia and bipolar I disorder in CBHOs. Data obtained at baseline from REACH OUT were used to evaluate the measurement properties of part I of the ISST-Plus, which is devoted to suicidal thinking. Part I comprises 12 three-point Likert scale items, each thought to measure suicidal ideation. Part III includes a seven-point Likert scale global suicidality rating. Cronbach's alpha and correlation analysis were used to assess the internal consistency and construct validity of ISST-Plus part I. In addition, a factor analysis will be conducted to examine the dimensional structure of this instrument.

Results: Interim data of 244 subjects were included in this sample, of which 82.4% were "not at all suicidal," 16.8% were "mildly suicidal," and 0.8% were "moderately suicidal," according to the global suicidality rating. Cronbach's alpha for the ISST-Plus part I was 0.86, indicating good internal consistency of the scale. The correlation between the part I total score and the global suicidality rating was 0.76, indicating good construct validity for this scale. Results of the factor analysis will be reported.

Conclusion: This analysis demonstrated good internal consistency and construct validity of the ISST-Plus part I as a measure of suicidal thinking. Additional validation of the instrument is needed to better understand its underlying structure and its ability to validly assess change and suicide risk.

Supported by Janssen Scientific Affairs, LLC

9 Does Pharmacogenetic Testing in Psychiatric Populations Influence Clinician Treatment Selection and Confidence?

Jay Lombard¹, Roy Perlis², Bryce Kasuba¹, Rachel Dicker¹

¹Genomind, ²Massachusetts General Hospital

Introduction: Response to psychotropic therapy is highly variable and can be attributed, in part, to heritable traits. The extent to which testing for genetic variations in clinical contexts influences clinical practice has not been established. This study aims to demonstrate characteristics of patients selected by their providers to receive pharmacogenetic testing and to examine effects of availability of genetic results on clinician treatment decisions.

Methods: This is a retrospective cross-sectional analysis of clinician survey responses. A select group of clinician members of the Neuroscience Education Institute (NEI) was recruited to participate in a program involving the use of a genetic assay in psychiatric patients. Genes tested included the serotonin transporter linked protein gene (SLC6A4), voltage gated calcium channel gene (CACNA1C), dopamine receptor subtype two gene (DRD2), catechol-O-methyl transferase gene (COMT) and methylenetetrahydrofolate reductase gene (MTHFR) as well as cytochrome P450 2D6 (CYP2D6) and cytochrome P450 2C19 (CYP2C19). Analytic results reports were provided to clinicians. Decisions concerning appropriate treatment regimen for patients were solely the responsibility of the clinicians. Clinicians were asked to complete a Clinical Decision Survey (CDS), which included questions ranging from patient's chief complaint to influence on treatment. CDS responses were analyzed using SAS/STAT software, version 9.2 of the SAS/STAT system for Windows.

Results: CDSs from 69 clinicians were collected with information relating to 105 patients. Clinicians reported, as a result of receiving a genetic report, an influence on treatment for 76.1% of patients, and a change in medication treatment for 67.6% of patients. Clinicians also reported increased confidence in treatment decisions, as a result of receiving assay results, for 87.6% of patient cases. Clinicians reported at least two years since the last month-long period of euthymic mood in 47.1% of patients, more than two adequate treatment trials in 74.5%, more than four adequate treatment trials in 39.2% and more than 10 adequate treatment trials in 7.8% of patients. 41.7% reported a chief complaint representative of depression, 10.6% representative of bipolar, 25% representative of anxiety disorders, 20.3% representative of reduced motivation/fatigue, and 3.8% representative of cognitive dysfunction.

Conclusion: The results of this study suggest that this genetic assay is clinically acceptable. Preliminary clinician responses indicate clinical value through added confidence in treatment decisions for a patient population which has a significant resource burden related to treatment failures. Prospective research is needed to further establish utility of this genetic assay as it relates to treatment guidance .

Disclosure Statement: Bryce Kasuba and Rachel Dicker are employed by Genomind, Jay Lombard is an employee of and has equity interests in Genomind, and Roy Perlis serves on the Scientific Advisory Board for Genomind.

10 Utilizing site-based raters and computer administered assessments to identify placebo responsive subgroups in a Bipolar Depression RCT

Sachs GS¹, Jacoby D², Arkow M³, DeBonis D³

¹Bracket Global and Massachusetts General Hospital, ²Repligen Corporation, ³Bracket Global

Introduction: High placebo response is associated with clinical trial failure. The search for interventions to reduce placebo response is important to successful drug development.

Methods: We examined patterns of placebo response to explore hypotheses regarding the failure of a double blind bipolar depression study. The trial included tandem assessments of diagnosis and the primary outcome measure (MADRS) administered by site-based raters (MADRS_{SBR}) and independently assessed by computer (MADRS_{COMP}) Placebo response was compared for subgroups defined by clinical criteria (e.g. Baseline Depression severity) and study operational criteria (e.g. academic vs independent site).

Results: Overall no significant differences were found between the placebo group (n=68) and the active treatment group (n=72) in the 140 subjects for whom treatment assignments and tandem ratings were available. The Active vs placebo difference ($AP\Delta$) from the subsample enrolled at academic sites favored active treatment, but only reached marginal statistical significance. At independent sites, however, a trend favoring placebo was found MADRS_{SBR} vs MADRS_{COMP} $AP\Delta = -1.3$ vs -2.0 , respectively. While active treatment performed comparably at academic and independent sites mean placebo response was greater at independent sites (Placebo MADRS_{SBR} = -10.1 vs 14.2 and Placebo MADRS_{COMP} = -8.1 vs 14.3). Among the placebo treated subjects response was greater in those with no family history of mood disorder (FH-) vs those with any family history of mood disorder (FH+, Placebo MADRS_{SBR} = -12.2 vs -18.5 , $p < 0.10$ and Placebo MADRS_{COMP} = -11.7 vs -18.0 , $p < 0.10$). In the FH- subgroup (n= 22) placebo was superior to active treatment on both MADRS_{SBR} and M MADRS_{COMP} $AP\Delta = +11.3$ vs $+11.2$ respectively.

Conclusion: Placebo response was not uniform across site and subject variables. Our results suggest higher rates of placebo response associated with enrollment at independent sites and subjects without family history of mood disorder may be causal factors in failure of the efficacy study. Further studies are needed to clarify if these variables are better understood as correlates of other aspects of clinical trial operations such as diagnostic validity or enrollment rate. Recognition of differential placebo response across subgroups may provide a basis for improving RCT design.

11 Parkinson's Disease Sleep Scale – Validation of the revised version PDSS-2

Kohnen R¹, Trenkwalder C², Chaudhuri R³ and the PDSS-2 study group

¹ReSearch Pharmaceutical Services Inc., Fort Washington, PA, USA, and Psychology Department, University of Erlangen-Nuremberg, Nuremberg, Germany, ²Paracelsus-Elena Hospital, Kassel, Germany and Clinical Neurophysiology Department, University of Goettingen, Germany, ³National Parkinson Foundation Centre of Excellence, Kings College Hospital and University Hospital Lewisham, Institute of Psychiatry, London, UK

Introduction: The Parkinson's Disease Sleep Scale (PDSS) is the only disease-specific sleep scale in Parkinson's Disease (PD). Its previous version consisted of a 15-item visual analogue scale that assesses the profile of nocturnal disturbances in PD patients.

Objective: To extend the scale so that it becomes a frequency measure scale with five categories (0=never to 4=very frequent) and encompasses unmet needs such as symptoms of restless legs syndrome, akinesia, pain and sleep apnea.

Methods: For validation of the PDSS-2, PD patients' ratings and investigators' interviews were compared to ratings from a semi-structured interview with a caregiver/partner, and to related scales. The PDSS-2 was repeated for test-retest-reliability after 1-3 days. For validation, the PDSS-2 was compared with the UPDRS (Unified Parkinson's Disease Rating Scale) severity ratings, PDQ-39 quality of life questionnaire, and the MOS sleep scales. Discriminative validity used stratification according to the CGI severity rating, the Hoehn & Yahr staging and reports from caregivers on sleep problems of the patients.

Results: A total of 113 PD patients from 3 sites in Austria, Germany and UK showed a mean (SD) total score of 16.5 (\pm 8.9) (range: 2–40) indicating mild to moderate sleep disturbances on average. PDSS-2 item-total correlation for proving internal consistency was satisfactory (correlations $>$ 0.30). From a factor analysis 3 subscales were derived: 1: "motor problems at night", 2: "PD symptoms at night" and 3: "disturbed sleep". Cronbach's alpha coefficient for the total score was 0.73, its test-retest-reliability (ICC) was 0.80. With a few exceptions (MOS sub-scales "snoring" and "optimal sleep", PDQ-39 sub-scale "social support", UPDRS IV.B "Clinical Fluctuations") significant correlations were observed between the PDSS-2 total score and all convergent sub-scales of the four instruments used for the assessment of convergent validity indicating that the PDSS-2 total score is qualified to measure sleep problems (MOS) and consequences of disturbed sleep on quality of life (PDQ-39). These correlations also showed that the more severe the sleep disturbances were described by the patients the more severe their overall PD symptoms were rated by the investigator (UPDRS, CGI). For discriminative validity, significant differences were found in the PDSS-2 total score depending on CGI, Hoehn and Yahr severity levels, and caregiver ratings.

Conclusions: According to the results of this validation study, the PDSS-2, with an extended spectrum of nocturnal disabilities and easier use for patients, is a reliable and valid tool for measuring sleep disturbances in PD. It is currently utilized in several clinical trials.

12 Diagnostic Test Accuracy Systematic Reviews in schizophrenia: optimizing screening of articles

Grabowski S¹, Maayan N¹, Soares-Weiser K², O'Blenis P³, Roberts S⁴, Adams CE⁵

¹Research Associate, Enhance Reviews Ltd, Oxford, UK ²CEO, Enhance Reviews Ltd, Oxford, UK ³Director, Evidence Partners Inc, Ottawa, Canada ⁴Search specialist, Cochrane Schizophrenia Group, University of Nottingham, UK ⁵Coordinator Editor, Cochrane Schizophrenia Group, University of Nottingham, UK

Introduction: We are currently evaluating the accuracy of Schneider's First Rank Symptoms and Operational Criteria (DSM, ICD, Feighner or RDC criteria) as a diagnostic tool to differentiate schizophrenia from other psychotic disorders. Identifying the relevant studies is an issue, as most of these studies are not indexed adequately and literature searches are very broad.

Methods: The Cochrane Schizophrenia Group conducted searches in MEDLINE, EMBASE, and PsychINFO (OvidSP), without applying any restrictions based on language or study design. References were independently screened by two reviewers using specific forms in DistillerSR systematic review software. This software allowed for blinded title and abstract screening and reference categorization as well as providing management and reporting functions. Reviewers used iPads to conduct screening as we found that this was faster and more comfortable than paper or PC-based methods for large volumes of references.

Results: We identified 32874 potentially relevant references. By screening titles we discarded 29800 irrelevant references. Of the remaining 3074 abstracts, we discarded 72% as irrelevant. Full text reports of 781 references are being ordered for further inspection. For 15% of references there were disagreements between the two reviewers. Distiller easily identifies these conflicts and assists in their resolution by making it easy for a senior reviewer to check them.

Conclusions: This huge task has been informative. Distiller software has been invaluable to refine, record, and make accurate the drudgery of this work. There is a complete lack of standards by which diagnostic test accuracy studies are referenced. Broad search strategies are currently impossible to avoid and content indexing should be improved. These reviews leave authors with large volumes of titles to screen for a relatively few potentially ‘true positive’ studies. Even at the abstract level, broad inclusion criteria have been necessary as it was often impossible to discern if the paper was really a report of a study of diagnostic test accuracy. As a result, there remains a large volume of full-text articles to acquire and screen.

13 **SOME URBAN LEGENDS OF CNS CLINICAL TRIAL METHODOLOGY: Unsuccessful Solutions to the Problem of Failed Trials**

Williams, JBW^{1,2}, Popp, D¹, Reines, S¹, Detke, M^{1,3}

¹MedAvante, Inc. ²Department of Psychiatry, Columbia University ³Indiana University School of Medicine

Introduction: As the rate of failed trials in CNS has grown, drug developers have attempted strategies to improve signal detection and reduce failures. We present five common strategies and evaluate their effectiveness.

Methods:

1. Increasing sample size. If statistical power increases with sample size and effect size is fixed, it appears reasonable that increasing sample size will increase effect size. Liu et al. (2008) examined four depression trials to evaluate this assumption.
2. Choosing “proven” sites. Some believe that selecting sites with proven effectiveness across several studies will continue to yield positive results. Gelwicks et al. (2002) analyzed data from sites that participated in at least two trials with at least 30 subjects.
3. Using experienced raters. It seems logical that more experienced raters will minimize variability, thereby improving signal detection. Kobak et al. (2009) examined interrater agreement across three cohorts of raters: experienced and calibrated, experienced but non-calibrated, and inexperienced.
4. Increasing training. The potential for variability across raters in a trial negatively affects study power and signal detection, leading some to believe increasing rater training will reduce variability. Demitrack et al. (1998) trained raters in an intensive session with videotapes and discussion.
5. Using certain regions. Many believe greater signal detection can be obtained outside the US. Khin et al. (2011) conducted a meta-analysis of FDA data on 81 US and ex-US antidepressant trials.

Results:

1. Increasing sample. In three positive studies, treatment effect was observed before the first 100 subjects per treatment arm were enrolled. Continuing to enroll subjects did not maintain significance in most cases. Treatment effect size decreased over time despite increases in sample size.
2. Choosing “proven” sites. Site performance across consecutive studies was inconsistent (all correlations <.50).
3. Using experienced raters. Calibration appears to improve reliability over and above experience alone. Experienced and calibrated raters had the highest ICC (.93) whereas experienced and non-calibrated raters had the lowest ICC (.55).
4. Increasing training. ICCs did not improve across six hours of training.
5. Using certain regions. Analysis revealed increasing placebo response across US and ex-US regions, and a similar decrease in US and ex-US effect size.

Conclusion:

Strategies for improving signal detection are often used, despite a lack of clear evidence of their effectiveness. These “urban legends” are widely touted, but evidence to support them is mixed at best. This review highlights the importance of examining effectiveness of methodological solutions to improve trials.

14 **Scientific and Economic Benefits of Sequential Parallel Comparative Design (SPCD), a Novel Clinical Trial Methodology**

Knable MB¹, Bowman M²

¹Bethesda Behavioral Sciences, ²RCT Logic LLC

Introduction: In order for a New Drug Application to be approved by the FDA, a study drug must often be superior to placebo in at least 2 pivotal trials. Also, many trials are sometimes required before 2 positive trials are obtained. It is estimated that the placebo response rate in trials for Major Depressive Disorder (MDD) is 35-45%.

Sequential Parallel Comparative Design (SPCD) is a novel clinical trial method that may be able to reduce placebo response rates and the expense of clinical trials.

Methods: An SPCD trial involves two phases of treatment. Phase 1 is aimed at: (1) comparing drug and placebo, and (2) generating a cohort of placebo non-responders. Phase 2 is aimed at comparing drug and placebo, as in a conventionally designed trial, but utilizing individuals who were placebo non-responders in Phase 1. Sample sizes and power for standard design and for SPCD trials were computed using 1 and 2 degree of freedom score tests and various assumptions regarding differences between drug and placebo response rates (see: "Sample Size and Power Calculations for the Sequential Parallel Comparison Design" Ivanova et al *Statistics in Medicine* 2011, <http://onlinelibrary.wiley.com/doi/10.1002/sim.4292/abstract>)

Results: When estimating sample size at assumed levels of statistical power, the SPCD method typically allowed for a 20 - 50% reduction of sample size. Likewise, when estimating statistical power with given sample sizes, the SPCD method allowed for a 10-25% increase in statistical power.

Conclusions: It is estimated that the SPCD design would be associated with significant savings in direct clinical trial costs, decreased time to market for new drugs, and reduced costs associated with "failed" trials.

15 Prior Trial Experience and Stipend Amount Not Associated With Improvement on Placebo in Antipsychotic RCTs in Schizophrenia

Robert E. Litman, MD¹, Jelena Saillard, MS, MBA¹, Susan Szymialis, BS¹

¹CBH Health, LLC

Introduction: Positive treatment response to placebo in antipsychotic randomized, controlled trials (RCTs), a major confound in antipsychotic development, has been attributed to the inclusion of subjects who repeatedly seek trial participation and high reimbursements (aka "professional patients").

Methods: We retrospectively reviewed a sample of 76 schizophrenia research subjects randomized to placebo and 40 subjects randomized to active comparator treatment from 8 antipsychotic RCTs ranging in length from 5-9 weeks to determine if there were differences in symptom response to placebo measured by the change in total PANSS ratings (Δ PANSS, endpoint-baseline, LOCF) between subjects who were naïve to clinical trials and patients who had already participated in at least one prior clinical trial. We also examined the relationship between stipend amount, Δ PANSS, and study completion for all subjects.

Results: There was no difference in Δ PANSS between trial-naïve and experienced trial patients (trial-naïve Δ PANSS = 6.63 + 13.86; experienced Δ PANSS = 5.12 + 15.57; $t = 0.4$, $p = \text{NS}$). In addition, 15 patients with placebo treatment response (Δ PANSS >5% improvement) and 34 patients with placebo worsening (Δ PANSS >10% worsening) did not differ in terms of prior trial participation (placebo responder: 33.3% trial-naïve; placebo worsener: 41.1% trial-naïve; $\chi^2 = 0.27$, $p = \text{NS}$). Number of prior trials and stipend amount did not correlate with Δ PANSS (placebo group: Δ PANSS x number of prior trial participation $r = 0.05$, $p = \text{NS}$; Δ PANSS x stipend amount $r = 0.03$, $p = \text{NS}$; active comparator group: Δ PANSS x number of prior trial participation $r = 0.22$, $p = \text{NS}$; Δ PANSS x stipend amount $r = -0.30$, $p = \text{NS}$); however, the stipend amount positively correlated with study completion ($r = 0.34$, $p < 0.01$). Δ PANSS was significantly different between placebo and active comparator patients ($t = 3.17$, $p < 0.01$); significance was not affected when prior trial completion, stipend amount, or length of trial were entered as covariates.

Conclusion: Our data do not support an association between professional patients and placebo response in antipsychotic RCTs, although they are limited by small sample size and reflect only a single site's practices. Additional research should include larger numbers of patients from multiple investigative sites and RCTs.

16 7 Deadly Sins: Guidelines for Reporting Clinical Trials Methodology Research

Popp D¹, Williams JBW^{1,2}, Detke M^{1,3}

¹MedAvante, Inc. ² Department of Psychiatry, Columbia University ³Indiana University Medical Center

Introduction: With clinical trial failure rates in several disease areas approximately 50 percent, many methodological approaches for increasing signal detection have been proposed, but reporting variations deter evaluations and comparisons. Standardization will alleviate this and reduce reporting bias in studies evaluating clinical trial methodologies.

Methods: We propose seven requirements for standardized reporting and illustrate how misuse or omission influences interpretation of results.

Results: Report interrater reliability (IRR). IRR should be reported in studies with multiple raters and observations. For severity assessments, the intraclass correlation coefficient (ICC) is required to accurately determine IRR. If individual items from a single severity scale are treated as separate observations, ICCs may be inversely related to reliability.

Use appropriate statistical tests. Kappa does not capture concordance on continuous variables. At times, a fixed criterion is used to indicate rater agreement with a “gold standard” score. This is highly influenced by the criterion and can inflate reports of IRR. Include effect size measures. Measures of effect size should be reported for all means comparisons regardless of statistical significance. This permits readers to determine clinical relevance, regardless of sample size, and allows comparisons both within and across studies with different outcome variables.

Identify a priori and post-hoc analyses. Methodological comparisons should be identified a priori in a statistical analysis plan, much like efficacy analyses, or be identified as post-hoc. Ideally, a primary analysis should be explicated. Failure to report this can result in over-interpretation of exploratory analyses performed on small subsets of data.

Acknowledge and correct for multiple comparisons. Authors should report all analyses if multiple comparisons are performed on a single sample, published or not. Reporting a significant result on a subset of data without indicating total comparisons made across the entire data set or correcting for multiplicity may lead to over-interpretation of false positives. Include inferential statistics for means comparisons. Statements concerning differences or patterns in means should be substantiated with inferential statistics. Correct interpretation of null hypothesis testing (NHT). NHT is commonly misinterpreted in methodology, such as concluding that smaller p values indicate more important effects, or that a non-significant p value represents a finding of no difference.

Conclusion: We propose guidelines for statistical reporting to standardize outcomes research on methodologies and reduce reporting bias. Empirical research evaluating the effectiveness of new methods to increase signal detection holds important consequences for clinical trial methodology and future drug development decisions facing sponsors and regulators.

17 **The Impact of Continuous Quality Improvement on an Early Phase Clinical Trial: Operational Methodologies and Best Practices**

Stacey Boyer¹, Joanne Bell¹, Wendy Davidson¹, Barbara Evans¹, Leslie Jacobsen¹, Santiago Arroyo¹, Maurizio Fava²

¹Pfizer, Inc., ²Massachusetts General Hospital

Introduction: To enhance the efficiency and quality of early phase trial design, execution and outcomes, strategic and adaptive continuous quality improvement (CQI) practices are recommended to ensure study fidelity. In order to facilitate timely delivery of the trial, quality advancement activities were executed to support clinical operations requirements for investigators relative to those of internal stakeholders. These methodologies have leveraged resources from diverse areas of expertise, yielding a common culture of CQI and the opportunity for application in future clinical trials.

Methods: In order to enable effective solutions, options were identified to facilitate effective planning, streamline processes and practices, and advance cross-line coordination. Design and operational elements were incorporated to maximize efficiency and ensure study fidelity. Through understanding and improving underlying processes and systems, application of CQI was implemented to drive quality and timeliness of pivotal study milestones and metrics.

Results: Through the development of CQI efforts in this trial, the interventions were able to be assessed by the study team to analyze the rate of return. The application of these methodologies has led to sustained focus on performance, quality, and compliance. The SAFER independent remote interview provided confirmation of eligibility for enrollment and randomization, enabling quality patient selection. An established clinical PK unblinding plan for monitoring drug exposures was sensitive to larger scale miss-dosing, minimizing risk of inappropriate termination of the study while ensuring patient safety and scientific integrity. Application of a failure mode and effects analysis yielded de-risking of dosing processes, while maintaining study compliance.

Additionally, screen and open label fail data was reviewed in an ongoing manner, yielding identification of opportunities to adjust entry criteria without compromising study integrity. Patient flow was further maintained by requiring all sites to have back-up raters present and qualified for ratings. Consistent site engagement was crucial as facilitated by a refresher investigator meeting, site-level meetings, and weekly calls with sites and monitors to identify barriers to recruitment and brainstorming of mitigation strategies.

Conclusions: It is imperative that clinical teams create and foster a culture of collaboration and communication to optimize the conduct and quality of their trials under specified timelines. Highlighting study progress, while focusing on information exchange is key. CQI application in clinical practice enables the success of clinical trials through the integration of collective operational creativity and scientific innovation, ultimately driving the delivery of therapies to fulfill substantial unmet medical needs.

18 Results of an H3 Receptor Antagonist Clinical Trial in Adults diagnosed with Excessive Daytime Sleepiness (EDS) Associated with Narcolepsy

Stacey Boyer¹, Donna Palumbo¹, Jing Liu¹, Yifan Huang¹, Santiago Arroyo¹, Gary Zammit², Andrew Krystal³

¹Pfizer, Inc., ²Clinilabs, Inc., ³Duke University Medical Center

Introduction: A potential alternative approach to the therapy of excessive daytime sleepiness (EDS) associated with narcolepsy, is the blockade of histamine H3 receptors. Histaminergic neurons arising from the hypothalamic tuberomammillary nucleus project to a variety of brain regions that regulate the sleep-wake cycle. Preclinical data indicate that numerous H3 antagonists are reported to increase wakefulness. H3 antagonism was also associated with insomnia in healthy subjects. However, there are no reports of the treatment of humans with narcolepsy with an H3 antagonist. This study aimed to investigate the efficacy and safety of a selective H3 receptor antagonist in improving EDS in adults with narcolepsy.

Methods: Patients with a primary diagnosis of narcolepsy according to the International Classification of Sleep Disorders (ICSD) criteria and excessive sleepiness associated with the diagnosis of narcolepsy as demonstrated on the Maintenance of Wakefulness Test (MWT) were included. The study was randomized, double-blind, placebo-controlled, multi-center, two-period crossover, investigating a fixed-titration, flexible-dose dosing scheme (derived from compound-specific knowledge gained in previous studies and via PK analysis) was conducted with an H3 antagonist and placebo. Each treatment period consisted in a 20 day Titration Phase and a 21 day stable Treatment phase. There was a 7-day washout phase between the two treatment periods. Target dose was 2 mg. The primary efficacy endpoint was the Maintenance of Wakefulness Test (MWT) and the key secondary efficacy endpoint was the Epworth Sleepiness Scale (ESS).

Results: Ninety-five subjects were randomized to the study. Fifty-five completed active treatment and sixty completed placebo. The study subjects average age was 32.1 years. The least-squares means for the change in average MWT nap latency from baseline to the end of 3-week stable dosing phase were 1.27 (0.546) minutes for active treatment and 1.12 (0.514) (ns) minutes for placebo. The difference (80% CI) between placebo and active treatment in this change was 0.16 (-0.81, 1.12) minutes. The active treatment vs. placebo difference in least-squares mean ESS total score change from baseline was small and insignificant at days 5, 10, 15 and 20 of the titration phase, and days 7, 14 and 21 of the stable dosing phase. Headache (10 (12.8%)) and insomnia (8 (10.3%)) were the all-causality adverse events seen in > 5% with the highest frequency of occurrence in subjects receiving active treatment. The observed serum H3 antagonist concentrations in narcoleptic subjects were as expected from healthy subjects.

Conclusions: While the selective H3 receptor antagonist was generally safe and well tolerated, the study did not exhibit evidence of a significant treatment effect in patients with narcolepsy and excessive daytime sleepiness.

19 eC-SSRS Assessments of Lifetime Ideation and Behavior are Predictive of Suicidal Behaviors Occurring During Trial Participation

James C. Mundt¹, Kelly Posner², John H. Greist³, Michael Federico⁴

¹ Center for Psychological Consultation, ² Columbia University, ³ Healthcare Technology Systems, ⁴ ERT

Introduction: Safety concerns regarding suicidality in clinical trials resulted in the release of FDA draft guidance in September, 2010. Prospective assessment of suicidal ideation and behavior in clinical trials was

recommended, and the Columbia-Suicide Severity Rating Scale (C-SSRS) was identified as an acceptable instrument. Alternative administration methods, such as interactive voice response (IVR) technology, are also acceptable. The feasibility, reliability, and validity of the eC-SSRS have been demonstrated and it has been incorporated into many clinical trials. This poster examines the importance of lifetime ideation and behavior assessed at baseline in relation to prospective risk for emergent suicidal behavior during trial participation.

Methods: 35,224 eC-SSRS records from ongoing and completed studies were extracted from a centralized database in May 2011. Data from 14 studies (7 Major Depression; 3 Insomnia; 2 Epilepsy; 1 Post-traumatic Stress Disorder; and 1 Fibromyalgia study) were merged and 217 records (0.6%) were excluded due to incompleteness, early system implementation errors, and under-representation of fibromyalgia subjects. Each record included study, site, and subject IDs, date/time stamps for each assessment, and subject responses to the eC-SSRS queries.

Results: Each eC-SSRS assessment (6,308 baseline/lifetime; 28,699 prospective follow-ups) was scored with respect to reported ideation and/or behavior. Lifetime ideation with an intention to act was reported at baseline by 14.1% of subjects; 27.9% reported prior suicide-related behavior. A baseline and one or more prospective follow-ups were provided by 3,776 subjects (Mean of 5.9 visits and 63.7 days of follow-up). The percentages of subjects prospectively reporting suicidal behaviors during study participation, related to lifetime ideation and behavior are shown below.

These data show that subjects who report lifetime suicidal ideation that includes a method or plan with intent and/or prior suicidal behavior at baseline are four to eight times more likely to prospectively report a suicidal behavior during study participation than subjects without lifetime ideation or behavior.

Baseline Report (Lifetime)	N	Prospectively Reported Suicidal Behavior	Relative Risk
Ideation (-); Behavior (-)	2,792 (73.9%)	2.4%	1.00
Ideation (+); Behavior (-)	75 (2.0%)	12.0%	5.00
Ideation (-); Behavior (+)	478 (12.7%)	9.6%	4.01
Ideation (+); Behavior (+)	431 (11.4%)	18.3%	7.64

Conclusion: Improved precision for suicide monitoring in clinical trials is critically important. The eC-SSRS is an efficient and effective tool for prospectively monitoring treatment safety and benefit. Subjects reporting lifetime suicidal ideation with intention to act, prior suicidal behavior, or both, at baseline assessments are at greater risk of prospectively reporting emergent suicidal behavior during trial participation.

20 Recovery to Neurocognitive Baseline After Acute Ethanol Intoxication

Chandiramani DV¹, Goddard SD², Boyd AF³, Udani J⁴

¹UNC Eshelman School of Pharmacy; ²Texas A&M University Department of Statistics, ³CNS Vital Signs, LLC, ⁴Medicus Research, LLC

Introduction: The aim of this study was to measure objectively the neurocognitive effects of acute ethanol toxicity and recovery to baseline over a two-day study period. The effects of ethanol are known to be dependent on blood alcohol concentration (BAC). The BAC can be measured objectively with local equipment (Breathalyzer or blood tests) and is an accepted marker of impairment by the courts. The metabolic pathways of ethanol elimination are well researched and are subject to inter-individual variation by gender, liver volume, ethnicity and genetic polymorphisms. Ethanol traditionally follows zero order kinetics where elimination occurs at a constant rate until most of the alcohol is cleared. On average, elimination rates for ethanol range between 10-25 mg/100ml/hour. Elimination rates are faster in women than in men, but not to an extent that warrants a separate gender-specific calculation.

Methods: 15 healthy volunteers (10 males, 5 females) aged 21-50 years took part in this confinement study. Subjects abstained from ethanol use for a week prior to the study. Exclusion criteria included abnormal liver enzymes, any evidence of liver disease, a family history of alcoholism, or any clinically significant organ system disorder. CNS Vital Signs, a computerized neurocognitive battery was administered to measure neurocognitive performance. Subjects were tested across 2 days after achieving a protocol defined BAC of 0.08 to 0.12. Analysis by paired T-tests with a Bonferroni correction applied was performed to compare each sample (Test 2,3,4,5) to the baseline sample (Test 1). After applying the Bonferroni Correction a p-value of 0.0125 is required for statistical significance.

Results & Discussion: Across both the traditional domains, as well as the pooled reaction times, impairment under the influence of ethanol was profound. Assuming the rate of elimination of 10-25mg/100ml/hr, each subject should have cleared the ethanol sometime between 2300 on day 1 and 0600 on day 2. Thus all subjects should have cleared the ethanol by the time Test 3 was administered. The Average Domain Standard Score mean change from Test 1 to Test 2 was approximately 15 points (1 SD). The Pooled Reaction Time mean change from Test 1 to Test 2 was 264 mS. From Test 2 to Test 3 the Average Domain Standard Score improved 12 points returning to a statistically equivalent baseline, $p=0.1639$. However, the Pooled Reaction Time change from Test 2 to Test 3 improved only 88 mS, yielding a p -value of 0.0293 when compared to baseline, suggesting residual impairment after the ethanol was cleared.

Conclusion: Ethanol BAC levels achieved in this study affected neurocognitive function. For domains, a statistical return to baseline was apparent at Test 3 (12 hours post-ethanol consumption). However, return to baseline for the Pooled Reaction Times at Test 3 almost approached a statistical significance from baseline with a p -value of 0.0293. This finding suggests reaction time recovery to baseline while recovering from ethanol intoxication may be slower compared to domain score recovery. For accurate assessment of reaction time deficits, a precision timing instrument such as a computerized neurocognitive battery should be considered.

21 Utility of the Intent to Attend Scale in a Long-term Schizophrenia Trial

Satoru Tsuchiya, M.S.¹, Jay Hsu, Ph.D.¹, Josephine Cucchiari, Ph.D.¹, Antony Loebel, M.D.¹

¹*Sunovion Pharmaceuticals, Inc*

Introduction: Attrition typically ranges from 25% to greater than 60% in randomized clinical trials (RCTs) in schizophrenia (Rabinowitz et al, 2009). As a consequence, attrition can introduce bias and decrease the statistical power, precision, and generalizability of study results. Although mixed-effects models permit use of incomplete data, MMRM analyses assume data is missing at random. Leon et al (2007) have developed the Intent to Attend instrument, a brief scale that assesses a subject's self-rated risk for study dropout at baseline and subsequent visits. This study evaluated the utility of this instrument in predicting dropout over the course of a long-term clinical trial and in addressing potential bias due to dropout in efficacy analyses.

Methods: Data were analyzed from a study of subjects meeting DSM-IV criteria for chronic, stable schizophrenia who had been randomized to 12 months of double-blind treatment with lurasidone or risperidone. Efficacy assessments included the PANSS and CGI-S. The intent to complete (ITC) was a single item that asked the subject at baseline to rate "how likely is it that you will complete the study" on a 10-point scale ranging from 0 (not at all) to 9 (extremely likely). The intent to attend (ITA) item was completed at each visit, and asked "how likely is it that you will attend the next visit" on the same 10-point scale.

Results: Subjects with lower ITA scores (higher self-rated risk for dropout) had a higher likelihood of study discontinuation at the next visit, based on a Cox proportional hazards analysis. A trend to increased risk for dropout prior to study completion (12 months) was observed for subjects with lower ITC scores (higher self-rated risk for not completing the study) at study baseline. Use of ITC or ITA scores as a covariate did not affect results of MMRM analyses of the PANSS and CGI-S in this study.

Conclusions: This analysis of Intent to Attend data from a long-term, double-blind schizophrenia trial suggests that ITC/ITA is a useful measure to predict risk for subject dropout. ITC/ITA did not affect the results of a mixed effects model efficacy analysis, perhaps in part because the study enrolled stable subjects and thus change from baseline in efficacy endpoints was small.

References: Leon AC, Demirtas H, Hedeker D. Bias reduction with an adjustment for participants' intent to dropout of a randomized controlled clinical trial. *Clin Trials* 2007;4:540-7. Rabinowitz J, Levine SZ, Barkai O, Davidov O. Dropout rates in randomized clinical trials of antipsychotics: a meta-analysis comparing first- and second-generation drugs and an examination of the role of trial design features. *Schizophr Bull* 2009;35:775-88.

Funded by Sunovion Pharmaceuticals Inc.

22 A Double-Blind, Placebo-Controlled Study of Cariprazine as Adjunctive Therapy in Major Depressive Disorder

Hayes, R¹, Bose, A¹, Lu, K¹, Greenberg, W¹, Németh, G², Laszlovszky, I²

¹Forest Research Institute, Jersey City, New Jersey, ²Gedeon Richter Plc., Budapest, Hungary

Introduction: Cariprazine is an orally active and potent dopamine D3/D2 receptor partial agonist. Based on known properties and localization of D3 receptors, it has been hypothesized that a compound with higher selectivity for D3 versus D2 receptors may have treatment advantages including a better side-effect profile (eg, fewer extrapyramidal symptoms). Low-dose cariprazine exhibits antidepressant-like effects in various animal models. It was hypothesized that cariprazine would exhibit antidepressant efficacy at doses that would keep D2 receptor antagonism below the threshold for EPS in patients with nonpsychotic major depressive disorder (MDD). This study evaluated the efficacy, safety, and tolerability of cariprazine adjunctive to standard antidepressant therapy (ADT).

Methods: A randomized, double-blind, placebo-controlled trial in outpatients aged 18-65 years who met DSM-IV-TR criteria for MDD, with HAM-D₁₇ score ≥ 18 and failure to respond to adequate trials of 1 or 2 previous ADTs. Participants received open-label ADT (citalopram, duloxetine, escitalopram, sertraline, or venlafaxine ER) plus placebo for 8 weeks. Nonresponders were randomized to double-blind adjunctive treatment (placebo, low-dose cariprazine [0.1-0.3 mg/day], or high-dose cariprazine [1.0-2.0 mg/day]) plus ADT for an additional 8 weeks. A fixed-flexible dose design restricted possible dose increase to the 4-week time point. The primary efficacy parameter was MADRS change from baseline to Week 16; additional efficacy measures included HAM-D₂₄, CGI-I, and CGI-S. Third party, computer-based rater monitoring and diagnostic validation was conducted to control placebo response.

Results: 502 patients entered the prospective ADT phase and 231 were randomized to double-blind treatment. The least squares mean change from baseline to Week 8 in MADRS was -8.0, -7.5, and -9.8 for placebo, low-dose cariprazine, and high-dose cariprazine, respectively. The cariprazine high-dose group, but not the low-dose group, demonstrated consistent nonsignificant numerical improvements across the primary and additional efficacy parameters. Dose increases at Week 4 appeared to have no added benefit. Cariprazine was generally well tolerated; no new safety concerns were revealed, although changes in many safety parameters were greater with high-dose cariprazine relative to other groups.

Conclusions: High-dose cariprazine demonstrated consistent numerical improvement suggesting potential for efficacy at higher doses; it was generally well tolerated in patients with MDD. Lack of signal in the low-dose group failed to validate the antidepressant-like efficacy observed at low doses in animal models. The fixed-flexible design appears to have limited the interpretability of data from the higher dose in each treatment group and no obvious benefit was derived from third party rater monitoring and diagnostic validation.

23 A Multicenter, Randomized, Double-Blind Trial to Evaluate the Effect of Cariprazine in Bipolar Depression

Ahuja, S¹, Bose, A¹, Lu, K¹, Greenberg, W¹, Németh, G², Laszlovszky, I²

¹Forest Research Institute, Jersey City, New Jersey, ²Gedeon Richter Plc., Budapest, Hungary

Introduction: Cariprazine, a D3/D2 preferring dopamine D3/D2 receptor partial agonist antipsychotic candidate, is in Phase III clinical development for the treatment of schizophrenia and bipolar mania. Low-dose cariprazine has previously demonstrated efficacy in animal models of depression. This study was designed to investigate low-dose cariprazine in bipolar depression. Third party, computer-based rater monitoring and diagnostic validation (Concordant Rater Systems) was conducted to control placebo response.

Methods: An 8-week double-blind, randomized, parallel-group, flexible-dose study of cariprazine 0.25 to 0.75 mg/day (low dose), cariprazine 1.5 to 3.0 mg/day (high dose), or placebo was conducted. Male and female outpatients (age, 18 to 65 years) who met DSM-IV-TR criteria for bipolar I or II disorder, had a current major depressive episode >4 weeks and ≤ 12 months duration, scored ≥ 20 on the 17-item Hamilton Depression Scale (HAM-D-17), scored ≥ 2 on HAM-D-17 Item 1, and scored ≤ 12 on the Young Mania Rating Scale (YMRS) were included. Patients were randomized 1:1:1 to receive low-dose cariprazine, high-dose cariprazine, or placebo. Computerized patient rating and site-based rating assessments (Montgomery-Asberg Depression Rating Scale [MADRS], 24-item HAM-D, and YMRS) were completed weekly. The primary efficacy parameter was MADRS

change from baseline to Week 8 using mixed model repeated measures (MMRM) analyses.

Results: 233 patients were randomized. Of the Safety Population (n=227), 76% completed the study. More high-dose cariprazine patients (9%) compared with low-dose cariprazine (4%) and placebo patients (3%) discontinued due to adverse events (AEs). The least squares mean change from baseline to Week 6 (Week 8) in MADRS were -15.5 (-16.2), -15.7 (-16.9), and -17.3 (-16.2) for placebo, low-dose cariprazine, and high-dose cariprazine, respectively. Cariprazine was generally well tolerated; the most common AEs in the cariprazine groups ($\geq 10\%$ and twice rate of placebo) were insomnia, akathisia, dry mouth, nausea, and diarrhea.

Conclusions: This study failed to establish cariprazine efficacy at Week 8; Week 6 effect was not sustained at Week 8. The effect is similar to that seen with aripiprazole in patients with BP depression [Thase et al. J Clin Psychopharmacol. 2008;28 13-20]. In addition, improvement in the placebo group was notably high which could be attributed to use of fixed-flexible design or the titration scheme used. Use of third party rater monitoring using concordance between clinician and patient ratings had little effect in minimizing placebo response. Both doses of cariprazine were generally well tolerated.