Novel Method for Maintaining Ratings Quality in Global Schizophrenia Clinical Trials

David G. Daniel, MD, Alan Kott, MD

1Bracket Global, LLC

Introduction: Accurate and reliable diagnostic and rating procedures are challenging to maintain over time in global clinical trials. Rater drift, cultural and linguistic diversity, and relatively complex and lengthy measurement tools present special challenges in international schizophrenia trials. We describe the initial results of a system for monitoring the quality of evaluations and providing rapid feedback to local raters in geographically and culturally diverse global clinical trial settings.

Method: Multiple schizophrenia clinical trials involving North America, Europe, South America and Asia are underway utilizing the techniques described below. All raters were trained prior to study initiation utilizing highly interactive procedures including slide presentations, rating of videotaped patient interviews, and in some cases interview and rating of live actors trained to portray schizophrenia symptoms. After study initiation a video/audio recording system was installed at the sites for assessment of the accuracy and quality of site diagnostic and rating procedures. Sites uploaded videotaped diagnostic and ratings assessments for review by calibrated external reviewers of the same language and culture. The external reviewers provided feedback on an ongoing basis to the site and sponsor on diagnostic and scoring accuracy and interview quality. Interview quality was evaluated by the Research Interview Assessment Scale (RISA) (1).

Results: Early data and preliminary analyses are available from the ongoing studies. Additional data and analyses will be reported. Of 201 patients considered eligible at screening by sites 92% were considered eligible by external reviewers based on review of videotaped interviews. 155 videotaped PANSS administrations at sites were graded for interview quality by external reviewers using the RISA. 74.7% were regarded as excellent (RISA Score 28-30), 20.7% as acceptable (RISA score 24-27) and 4.6% as poor or unacceptable. Exact matches were obtained between the site and external rater on 60% or more of ratings for all 30 PANSS items. Mismatches between the site and external raters of 2 anchor points or greater in scoring any PANSS item were relatively uncommon and only exceeded 10% of ratings on items P2 (Conceptual Disorganization) and N7 (Stereotyped Thinking).

Discussion: External review of videotaped diagnostic and ratings interviews with timely feedback to sites and sponsors is feasible in global clinical trials settings. Interview quality at the sites and agreement between site and external PANSS ratings was relatively high in the preliminary analyses. New patterns may emerge and the results and conclusions may change as the size and cultural diversity of the sample increases. Future analyses will assess regional and cultural differences in measures of interview and ratings quality. In addition, the association between raters’ performance on assessments of interview and scoring competency at the investigators meeting vs. the subsequent quality of ratings of patients at the site will be examined.


Funding: Statistical analyses were funded by Bracket Global, LLC.
Assessment of Change in Body Weight After Antipsychotic Treatment is Confounded by Regression to the Mean

Antony Loebel1, Andrei Pikalov1, Jane Xu1, Cynthia Siu2, Josephine Cucchiaro1
1Sunovion Pharmaceutical Inc., 2 Data Power (DP) Inc.

Objective: The aim of this post-hoc analysis was to examine whether the previously observed correlation between initial BMI range and subsequent weight change reflected in part a statistical artifact, regression to the mean (RTM), rather than true effect modification.

Methods: Body weight and BMI were measured at baseline and at the 6-week endpoint in a double-blind, placebo- and active-controlled trial of lurasidone (LUR) and olanzapine (OLZ). Regression analysis was applied to estimate the magnitude of bias due to RTM on measurement of change in body weight at Week 6. To correct for the RTM bias, a control group was used in the ANCOVA model to estimate differences in weight change between treatment groups by baseline BMI ranges, using a statistical interaction test.

Results: Among placebo subjects, the magnitude of RTM bias in the obese subgroup (baseline BMI >=30) was -5 kg at week 6, due to the non-perfect correlation between baseline BMI and Week-6 measurements of body weight (r=0.87 < 1) and non-random selection (median baseline weight for the obese subgroup was 34.8 kg or 9 kg above the placebo group mean of 25.7 kg). A similar magnitude of RTM bias (-4 kg) was observed in the subgroup of obese subjects in both the olanzapine and lurasidone treatment groups. Compared to placebo, weight changes in the baseline obese, overweight, and normal groups were +4.4 kg, +5.3 kg, and +2.7 kg, respectively, for olanzapine-treated subjects (treatment-by-baseline BMI interaction tests, p=0.09); and +0.40 kg, +0.02 kg, and +0.68 kg, respectively, for lurasidone-treated subjects (treatment-by-baseline BMI interaction tests, p=0.72).

Conclusions: Contrary to previous findings, we found no evidence in this analysis supporting the argument that the magnitude of drug-induced weight gain was less in subjects with higher initial BMI than those subjects with average or low baseline BMI. Our findings suggest the previously observed inverse relationship between baseline BMI and weight change following antipsychotic treatment reflects, in part, RTM bias. Antipsychotic drugs appear to cause similar weight change in both high and low baseline BMI groups, when an appropriate control is incorporated in treatment comparisons.

Meta-Analysis of Structural Magnetic Resonance Imaging (sMRI) Biomarkers in Schizophrenia Medication Trials

Henry J. Riordan, PhD1, Kathy Dawson, PhD1, Neal Cutler, MD1, Chwen-Cheng Nathan Chen, MD, PhD2, Steven Potkin3, MD, Susan De Santi, PhD4
1Worldwide Clinical Trials, King of Prussia, PA and Beverly Hills, CA, 2Pfizer Inc, Groton, CT, 3UC Irvine, Irvine, CA 4GE Healthcare Princeton, NJ

Background: In order to determine the quality of evidence to support the application of imaging biomarkers in clinical trials of schizophrenia relevant sMRI literature was reviewed. A prior qualitative review following criteria by Altar et al (2008) suggested an evidence rating of “A-/B+” for sMRI in relation to medication effects. In order to provide a more quantitative evaluation a meta-analysis was undertaken to examine approved antipsychotic medication trials with at least one pre and post sMRI assessment.

Methods: A structured search of the literature utilizing PsycInfo, MEDLINE, PubMed, and Google Scholar databases yielded 265 papers for initial appraisal. Of these 100 abstracts were selected for more in depth review and of these 35 were reviewed more thoroughly. Careful examination revealed that six publications were suitable for meta-analytic review containing adequate data to calculate effect sizes (ES). Brain regions examined included whole brain, lobar, basal ganglia and ventricular volumes, as well as grey and white matter. Subject numbers for individual studies ranged from 10 to 73 with 145 schizophrenic subjects included overall. To control for differences in sample size, studies were weighted according to their inverse variance estimates. Weighted ESs (Cohen’s d) were calculated for change from baseline for each brain region along with 95% confidence intervals.

Results: Analysis of medication effects on brain regions across different medications, failed to reveal an overall significant effect (Cohen’s d = 0.07). However, there was considerable heterogeneity between brain regions with two of the nine regions yielding significant results. The weighted ES for whole brain measures was significant (d = 0.14) as was the ES for the putamen (d = 0.23); both suggestive of increases in sMRI measurements associated with
antipsychotic treatment. Important moderator variables included subtype of schizophrenia, typical versus atypical antipsychotic medication, chronicity and study length.

**Discussion:** This meta-analysis is preliminary and additional ES data will be calculated, as will an assessment of the relationship of volumetric changes to schizophrenia symptomatology. While these initial results do not fully support the qualitative evidence grading in that they do suggest that it is possible to more confidently select specific brain regions that are more likely to be sensitive to antipsychotic medication effects; and that the degree of this sensitivity appears to be moderated by important disease-related variables.


## 4 Computerized Functional Capacity Assessment in Schizophrenia: Evidence for Convergent Validity

Philip D. Harvey¹, Stacy Ruse ², Vicki Davix², Richard SE Keefe³

¹University of Miami Miller School of Medicine ²Neurocog Trials, Chapel Hill, NC ³Duke University

**Background:** Assessment of functional capacity is an intrinsic part of the treatment of cognitive impairments in schizophrenia. Current methods are found to be highly and consistently correlated with performance on neuropsychological (NP) tests, but the current assessment of functional capacity is still a role-playing exercise and more ecological validity would always be welcome. We have developed a computerized virtual reality assessment that contains all of the components of a shopping trip, including searching the pantry at home, making a list, taking the correct bus, shopping in a store, paying for the purchases, and getting home. We administered this assessment, as well as the MCCB and the standard functional capacity measure, the UCSD Performance-Based skills assessment, to a sample of people with schizophrenia concurrent validity.

**Methods:** As part of the VALERO study, part II, we are examining a large sample of people with schizophrenia with the MCCB, the UPSA-B, and real-world functioning measures. We added the virtual reality shopping trip at one of our sites. We present data on the first 36 patients with schizophrenia examined with this procedure and will be collecting more data until the time of the SIRS meeting. Dependent variables for the virtual test consisted of two different factor scores, one indexing the level of organization of the efforts (factor 1) and the other composed of the dependent measures for the success of completion of the tests (factor 2). We calculated a single global score for our analyses.

**Results:** Person correlations found that the VR composite factor was significantly correlated with UPSA-B total scores (r=.53) and with the MCCB composite score (r=.64). When a stepwise regression analysis was computed, the VRFCAT factor was entered into a regression analysis first, accounting for 41% of the variance in MCCB scores, and the UPSA-B scores did not enter the analysis. When the UPSA-B was forced in first, the VR factor score still accounted for 15% incremental variance beyond the effect of the UPSA-B.

**Implications:** A virtual reality functional capacity assessment was correlated with both paper and pencil functional capacity measures and with the MCCB. The correlations were quite high and the VR assessment was more strongly associated with the MCCB than was the UPSA-B. These results provide encouraging support for the possibility of remotely deliverable functional capacity assessment. The sample size will be larger by the time of presentation.

## 5 The Evaluation of Negative Symptoms by Videoconferencing in a Clinical Trial

JBW Williams¹,², D Popp¹, DA Osman¹, EA Cohen¹, MJ Detke¹,³

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

**Background:** Negative syndromes in Schizophrenia are of increasing interest to drug developers, and several assessment strategies have emerged for identifying negative symptoms. Among these is the Negative Symptom Assessment (NSA-16), as well as the PANSS negative symptom subscale and the Marder subscale. Assessment of patients with schizophrenia by videoconferencing has been shown to yield results equivalent to those obtained when the scale is administered face-to-face. Advantages of remote assessment include blinding to protocol details and visit number, effectively eliminating enrollment and expectation biases. Videoconferencing can also be used to facilitate calibration of a global cohort of raters, as interviews can be observed “live” by remote trainers. Examination of the ability to assess negative symptom scales reliably by videoconferencing is timely.
Methods: The PANSS and the NSA-16 were administered to subjects with schizophrenia in a randomized clinical trial via live videoconferencing by blinded independent central raters. Subjects were interviewed at screen, at 11 more visits over 36 weeks, and at endpoint or 1 year. On a subset of subjects, a senior clinician observed and independently rated the PANSS and NSA as a quality control measure.

Results: The PANSS and NSA-16 were administered at all visits (n=1127) to 224 subjects by 17 different blinded independent central raters. The mean duration of the NSA was 16 min. (SD=7); each followed a PANSS that was on average 36 min. (SD = 15). All total and subscale scores were normally distributed at screening. ICCs between raters and observing trainers were .98 on the NSA total score (N = 65 pairs) and .96 on the PANSS total score (N = 69 pairs). ICCs of individual NSA items ranged from .72-.1.0, with a mean ICC of .91. ICCs of PANSS subscales ranged from .94 -.96 with ICCs of .95 for the Marder subscale and .94 for the negative subscale. Inter-item correlations for the NSA and Marder subscale will be presented as well as relationships between items measuring similar constructs across the two scales.

Conclusion: Excellent item-level ICCs for the NSA suggest that negative symptoms can be rated reliably by videoconferencing using well-calibrated blinded independent raters.


Sharp IR, Kobak KA, Osman DA: The use of videoconferencing with patients with psychosis: a review of the literature. Ann Gen Psychiatry 2011;10(14)

6 NSA-16 Revisited: Identifying Latent Factors of Negative Symptoms in Schizophrenia

Popp, D1

1MedAvante, Inc.

Introduction: Negative symptoms in schizophrenia are inadequately treated and difficult to assess accurately. The Negative Symptom Assessment (NSA-16) was developed to evaluate the presence and severity of negative symptoms. Previous research suggests a multidimensional structure consisting of five factors: Communication, Emotion/Affect, Social Involvement, Motivation and Retardation.1

The current study re-examines the multidimensional nature of the NSA-16, when jointly administered with the PANSS, using Confirmatory Factor Analysis (CFA). Interrater reliability and convergent and divergent validity of the NSA-16 are examined.

Methods: Blinded independent central raters administered the PANSS and NSA-16 to subjects (n=223) with schizophrenia as part of a randomized clinical trial. Assessments occurred at screening, baseline, and 12 additional visits over one year.

Results: The intraclass correlation coefficient (ICC) for the NSA-16 total score was .98. Item score ICCs ranged from .72-1.0, indicating high interrater reliability.

A CFA replicating the proposed five factor structure of the NSA was conducted on data from screening visits. Results suggest poor discriminant validity among factors with six between-factor correlations >.90.

Re specification of the model was undertaken by progressively collapsing factors that were highly correlated and examining model fit statistics. A three factor model, collapsing Communication and Social Involvement into a single factor, as well as Emotion/Affect and Retardation, had an acceptable fit (CFI=.85; NFI=.80; RMSEA=.11).

Examination of factor loadings suggested loading the item Reduced Social Drive on the Motivation factor (CFI=.89; NFI=.85; RMSEA=.09).

Finally, to account for variance due to correlated measurement error, error variances for related items were correlated both within and between factors. The resulting modified three factor model with correlated error variance was the best fitting model (CFI=.92; NFI=.87; RMSEA=.08).

The three factors (Communication/Social Involvement, Emotion/Retardation and Motivation) demonstrated convergent validity with the PANSS negative (r=.79, .78 and .54 respectively) and Marder subscales (r=.74, .79 and .54), and divergent validity with the PANSS positive subscale (r=-.23, -.23 and .08).

Discussion: Results suggest a three factor multidimensional structure of the NSA-16. Replication of the five-factor
structure revealed discriminant validity concerns. Differences in results may be due to the NSA-16 being administered following the PANSS in the current study, which is common in clinical trials. Identifying reliable factors of the NSA-16 is important for developing new treatments to address the multidimensional nature of negative symptoms. Reliance on total scores that treat negative symptoms as unidimensional may obscure the efficacy of treatments on specific domains of negative symptoms and result in failed trials.


7 Using Generalizability Theory to Estimate the Effect of Raters, Subjects and Timepoints on the Reliability of Symptom Ratings on the Positive and Negative Syndrome Scale (PANSS)

Rothman, B1, Yavorsky C1,2, Khan A1,6, Liechti S1, DiClemente G2, Opler M1,3, DeFries A1,4, 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.

Introduction: In recent years, there is increasing attention to studying failed psychiatric clinical trials. The most commonly used scale in schizophrenia is the Positive and Negative Syndrome Scale (PANSS). The convention of reporting intraclass correlations (ICCs) assume that variance in scores is adequately explained by differences between subjects and error from differences in raters’ scores. An ICC fails to account for all relevant sources of unreliability, and PANSS results should be examined beyond ICCs. The goal is to identify the sources of unreliability in a failed clinical trial by assessing PANSS scores for placebo responders, placebo non-responders, and the treatment group.

Methods: This is a substudy from a failed Phase II double-blind, placebo-controlled trial of schizophrenia, using generalizability theory, to assess reliability on 3 conditions: raters, timepoints (8 visits), subjects.

Results: 141 and 71 subjects completed all visits for the treatment group and the placebo group, respectively. Using 20% improvement at endpoint, the placebo response rate was 40.07% and treatment response rate was 60.28%. For Positive subscale items, the most variability was for raters (range: 33% to 72%) for placebo responders, 31% to 68% for placebo nonresponders, 29% to 60% for the treatment group. The interaction of rater and timepoint was the second source of unreliability, averaging 12.28%, compared to 12.00% for placebo nonresponders, and 10.00% for the treatment group. All Negative subscale items showed the most variability for raters, for all groups. For General Psychopathology (except Preoccupation), raters accounted for the most variability in the scores for placebo responders averaging 51.00%. Ep2 is similar to the internal reliability coefficient α and denotes consistency in PANSS scores. For Positive subscale, Ep2 ranges from poor (Delusions) to excellent (Hostility), for all 3 groups, with an average of 0.727 for placebo responders, 0.662 for placebo nonresponders, and 0.680 for the treatment group. For Negative subscale, Ep2 ranges from poor (e.g., Passive Apathetic Withdrawal) to excellent (e.g., Difficulty in Abstract Thinking), averaging 0.647 across items for placebo responders, 0.667 for placebo nonresponders and 0.675 for the treatment group, denoting moderate consistency in scores.

Conclusions: Results confirm the efficacy of generalizability theory for the estimation of reliability and demonstrate a relationship between low rater reliability and a failed trial. Although there are other factors contributing to failed trials, PANSS training conducted at initial investigator’s meetings may not be effective enough for assessing psychopathology in longitudinal trials. Findings can guide data monitoring and rater training.

8 Reliability of a Structured Training Program for the Global Assessment of Functioning Scale in Patients with Excessive Sleepiness Associated with Shift Work Disorder

Yavorsky C1,2, Rothman, B1, DiClemente G2, Opler M1,3, Jovic S1, Khan A1,5, DeFries A1,4

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

Background: The Global Assessment of Functioning Scale (GAF) is widely used to assess psychological, social and occupational functioning. The GAF scale identifies the lowest and highest levels for a hierarchy of an illness. After the introduction of the GAF, a number of studies have been published that claim the GAF to be a
reliable scale, but these studies have all been conducted in research settings and have not been studied in patients with excessive sleepiness associated with shift work disorder. Additionally, the GAF lacks structure and depends greatly upon clinical judgment and conjecture. The directions on the GAF do not explicitly state how to integrate the three areas of functioning into a single composite rating.

**Objectives**: In view of the lack of studies on the GAF in patients with shift work disorder (SWD), the objective of this study was to assess the reliability of the GAF in a sample of outpatients with SWD in a clinical trial setting where the GAF is part of the outcome measure. We aimed to discover whether the GAF meets its purposes, when used in SWD, after a comprehensive training and structured guidelines was implemented.

**Methods**: Data was obtained from a Phase IV randomized placebo controlled clinical trial for 382 patients with clinically diagnosed shift work disorder. For this secondary analysis, only pre-treatment (screening and baseline) GAF scores are assessed as the aim of the study is to assess reliability and present baseline characteristics according to tertiles (low, medium, high scores) of the GAF in patients with SWD. Patients included in the study experienced late-in-shift sleepiness between 4 AM and 8 AM and were functionally impaired (Global Assessment of Functioning < 70).

**Results**: At screening, the mean GAF score by the clinician was 62.99 (SD: 4.48) and ranged from 46 to 70. The mean GAF score at baseline (1 week apart) was 62.78 (SD: 4.34) and ranged from 50 to 70. There were no significant differences across study sites. Pearson’s correlation coefficients between the baseline GAF score and the screening GAF score was 0.895 (P ≤ 0.001). Similarly, the concordance seen by Cohen’s Kappa between the GAF score at screening and at baseline when using categories (low ≤ 50, moderate 51 – 55, high ≥ 56) was statistically significant (kappa = .16, p = 0.03). Internal consistency between screening and baseline was also high with Cronbach alpha of 0.944.

**Conclusions**: The results of this study present new findings that the GAF can be rated reliably when used with mild-to-moderately impaired patients with SWD and when presented with specific guidelines for rating the GAF. Current guidelines for rating GAF are not comprehensive. Theoretical and empirical studies, as well as development of a manual with more information about scoring the GAF for various populations are warranted.
ASENT: Effect of Intramuscular Interferon beta-1a on Gray Matter Atrophy in Relapsing-Remitting Multiple Sclerosis

ASENT: Comparison of Brain Atrophy Measurement Methods in the Context of a Clinical Trial

ASENT: 5HT1A Binding and 5HT Transport in Frontal and Temporal Lobe Epilepsy

The Relationship Between Urine Cannabinoid Concentration and Choice Reaction Time in Chronic Marijuana Users

Boyd AF¹, Morgan, DW¹, Goddard SG², Gray KM³

¹CNS Vital Signs, LLC; ²Texas A&M University Department of Statistics; ³Medical University of South Carolina

Introduction: The aim of this study was to investigate whether urine cannabinoid levels in a group of chronic marijuana users were associated with choice reaction times as measured by a computerized neurocognitive test able to measure reaction time with millisecond (mS) precision.

Methods: At the baseline visit before treatment randomization in the parent marijuana cessation trial, 114 participants completed a comprehensive computerized neurocognitive testing battery (CNS Vital Signs). A component of the battery is the Continuous Performance Test (CPT), a well-known test measuring choice reaction time. Participants are asked to perform a task, responding only when a pre-defined stimulus (target) appears across a field of foils. As the task is relatively simple, CPTs typically have a ceiling effect (perfect “correct vs incorrect” scores are common). Data for participants scoring a perfect “correct vs incorrect” score on the CPT were analyzed by comparing the reaction times to the urine cannabinoid levels. A urine creatinine-normalized cannabinoid test was used to reliably compare quantitative levels (urine cannabinoid level/urine creatinine level). Regression analysis of the CPT reaction time compared to urine creatinine-normalized cannabinoid level was performed.

Results: Of the cross-section of 114 participants tested at baseline, 18 achieved a perfect score on the CPT. The mean reaction time for the sample was 381 mS while the median reaction time was 384 mS. The range was 303 mS to 457 mS. Further analysis of that group revealed a significant linear relationship between reaction time and urine creatinine-normalized cannabinoid level (p=.0091), with a correlation coefficient of 0.596 (95% CI 0.1784-0.836).

Conclusions: Participants with higher urine cannabinoid concentrations had slower reaction times. This finding suggests tests that precisely measure reaction times may identify subtle levels of marijuana-related impairment not observable using routine methods.

This research was supported by NIDA grant R01DA026777 and NCRR grant ULIRR029882.

Do Psychiatric Registries Include All Persons with Schizophrenia in the General Population? A Population-based Longitudinal Study

Mark Weiser MD¹,², Nomi Werbeloff, PhD¹, Bruce P. Dohrenwend PhD³, Itzhak Levav MD⁴, Rinat Yoffe MA⁴, Michael Davidson MD¹,²

¹Department of Psychiatry, Sheba Medical Center, Tel-Hashomer, Israel; ²Department of Psychiatry, Sackler Faculty of Medicine, Tel-Aviv University, Israel; ³Department of Psychiatry, Columbia University, New York; ⁴Division of Mental Health Services, Ministry of Health, Jerusalem, Israel

Background: Psychiatric hospitalization registries are utilized to investigate the incidence and prevalence of schizophrenia for both research and administrative purposes. The assumption behind this is that most individuals with schizophrenia will be hospitalized at least once in their life-time.

Method: In an epidemiological survey conducted in the 1980’s, a population-based sample (n=4,914) of Israel-born individuals then aged 25-34 were screened in the community, and 29 (0.6%) were subsequently diagnosed by psychiatrists using SADS/RDC criteria. Twenty four years later we linked data from the epidemiological survey with the Israeli National Psychiatric Hospitalization Registry.

Results: Twenty seven of the 29 individuals (93%) diagnosed with schizophrenia in the survey were identified in the hospitalization registry with the same diagnosis. Fifty-two (1.0%) participants not diagnosed during the
survey with schizophrenia were identified in the psychiatric hospitalization registry 24 years later with schizophrenia. The majority of them were diagnosed with other psychiatric disorders in the survey. If all diagnoses of schizophrenia are accepted at face value, the lifetime prevalence rate would be 1.8% for this cohort.

**Conclusion:** The overwhelming majority of individuals diagnosed with schizophrenia at ages 25-34 in an epidemiological survey were present in the Psychiatric Hospitalization Registry. However, the assessment of life-time rates of schizophrenia at these ages is problematic because some future cases are asymptomatic, others have premorbid non-psychotic disorders, while in others it is difficult to differentiate between affective disorders and schizophrenia. Availability of psychiatric services and hospitalization policy must be considered when generalizing these findings to other countries.

**Keywords:** Psychosis, Schizophrenia, Registry, Prevalence, Hospitalization

### 22 Rationale and Methods of a Trial of Lisdexamfetamine Dimesylate as Adjunctive Treatment for Negative Symptoms of Schizophrenia

Dirks, B1, Lasser, R1, Nasrallah, H2, Kirsch, C1, Gao, J1, James S1, Lindenmayer JP3

1Shire Development Inc., Wayne, PA, USA 2University of Cincinnati College of Medicine, Cincinnati, OH, USA 3New York University, New York, NY, USA

**Introduction:** Primary (enduring) negative symptoms of schizophrenia (NSS), associated with low mesocortical dopaminergic activity, respond poorly to current antipsychotics. Since amphetamines increase synaptic dopamine (DA) in striatum and prefrontal cortex, there is little recent research with amphetamines in NSS due to the potential for positive symptom (PS) worsening. Rationale and methodology are discussed for an adjunctive lisdexamfetamine dimesylate (LDX; prodrug of d-amphetamine) trial for carefully selected participants with stable PS, predominant NSS, and maintained on atypical antipsychotics that block mesolimbic DA-2 receptors and may prevent PS recurrence.

**Methods:** This 28-site trial of adjunctive LDX included a novel design of 10-week open-label (OL) and 4-week double-blind, placebo-controlled randomized withdrawal phase. Participants (18-55 years) with clinically stable schizophrenia (≥2 years), predominant NSS, low PS, and maintained for ≥12 weeks on antipsychotic monotherapy were enrolled. Participants with NSS were included with scores ≥55 (items 1-6, 8-12, 14-16, 18-21) on the Scale for Assessment of Negative Symptoms (modified-SANS) and scores ≥3 on ≥2 SANS Global Ratings (affective flattening, alogia, avolition-apathy, or anhedonia-asociality). Required low PS were assessed with Positive and Negative Syndrome Scale (PANSS) positive subscale scores <20; exclusion with scores ≥4 on 3 PANSS positive items (delusions, hallucinations, or suspiciousness/persecution). Excluded were participants who scored ≥9 on Calgary Depression Scale for Schizophrenia (CDSS) to avoid confounding depression effects on outcomes. Participants with current suicide risk, recent substance abuse history, or uncontrolled comorbid psychiatric disorder were also excluded. OL LDX (20-70mg/d) was optimized over 7 weeks and continued through OL maintenance (weeks 7-10). Participants with a reduction at week-10 in modified-SANS-18 total score were randomized to continue dose-optimized LDX or placebo for 4 weeks. Efficacy assessments included modified-SANS-18 total and global ratings; PANSS total and subscores. Safety assessments included adverse events, vital signs, electrocardiograms, and CDSS. Clinical stability/suitability to continue enrollment was confirmed at 3 weekly screening visits and with specific withdrawal criteria of ≥25% increase in PANSS total or ≥2-point increase in PANSS items (as above), increased suicidal/harmful behavior, hospitalization, caregiver loss, or risk of medication nonadherence.

**Conclusions:** Multiple design features—including trial duration; specific entry criteria for selecting individuals with predominant NSS, low PS, and low depression; and rigorous, continuous PS monitoring—aimed to optimize the potential for safe adjunctive LDX administration to antipsychotic medication and to assess treatment-related NSS improvement free of potential confounds. This methodology may be useful for future research on treating NSS with DA agonists such as amphetamines.

**Acknowledgments:** Clinical research was funded by the sponsor, Shire Development LLC. Under the direction of the authors, Michael Pucci, PhD, an employee of SCI Scientific Communications & Information (SCI), provided writing assistance for this poster. Editorial assistance in formatting, proofreading, copy editing, and fact checking was also provided by SCI. Shire Development LLC provided funding to SCI for support in writing and editing this poster.

**Disclosures:** One or more authors report potential conflicts which are described in the program. Bryan Dirks, MD is a Shire employee and holds stocks and/or stock options in Shire. Robert Lasser, MD is a Shire employee and holds stocks and/or stock options in Shire. Henry Nasrallah, MD has received research support and/or
consultant fees from AstraZeneca, Forest, Janssen, Merck, Novartis, Otsuka, Pfizer, Shire, and Sunovion. Courtney Kirsch, BS is a Shire employee and holds stocks and/or stock options in Shire. Joseph Gao, PhD is a Shire employee and holds stocks and/or stock options in Shire. Steven James, MD is a Shire employee and holds stocks and/or stock options in Shire. Jean-Pierre Lindenmayer, MD receives/has received advisory board/consultancy fees from the following companies: Janssen, Lilly, Merck, Multi Health Systems, Roche, and Shire; receives/has received income sources and equity from Office of Mental Health, NY State; receives/has received grant support from AstraZeneca, Janssen, Lilly, Otsuka, Roche, and Sunovion.

23 Recruiting Strategies for First Episode Schizophrenia: Staging Tactics

Elmouchtari A1, Schooler NR1, Weiden PJ1,2
1SUNY Downstate Medical Center, Brooklyn NY 2University of Illinois at Chicago, Chicago IL

Background: In general, enrolling psychiatric patients in clinical trials is a challenging task. Targeting a population of patients with schizophrenia at their first episode of illness, residing in an inner city Afro-Caribbean community, and belonging to families with low socio-economic and educational levels further complicates matters. And when your study protocol includes a long term prescription of a psycho-active medication that will be delivered after randomization to either an oral on an intra-muscular injection, the mission becomes close to impossible. We faced exactly this set of challenges. We designed a comprehensive and creative approach to recruit subjects with the goal to recruit while adhering to high scientific standards and the ethical principles (Weiden et al 2009).

Methods: A two-stage recruiting strategy was developed. Patients with provisional diagnoses of schizophrenia, schiz-ffective or schizophreniform disorder and less than 16 weeks of antipsychotic medication exposure were invited to evaluation and clinical stabilization (Stage 1) for up to 12 weeks. Stage 1 included one formal family psychoeducation session. They were told about the possibility of later RCT participation. Those whose diagnosis and treatment history were confirmed were invited to participate in an RCT to compare a second generation long-acting injectable antipsychotic to second generation oral medications.

Results: Seventy four subjects consented to Stage 1; 28 (38%) were not eligible to participate in the RCT. The most common reason was failure to engage in outpatient treatment, followed by medical reasons, poor clinical response, administrative reasons and failure to meet diagnostic criteria. Forty six (62%) met criteria for the RCT and 38 (83% of those eligible) consented.

Comment: The recruitment strategy allowed us to engage patients and form a therapeutic alliance before formally inviting them to RCT participation. Family psychoeducation was a key element in this process. We have not formally tested the alternative strategy of immediately inviting patients to participate in the RCT, but initial efforts at recruitment did not involve the strategy that required two separate consent interviews and consent forms. We believe that the process was successful because it built upon a developed therapeutic alliance and trust. The method may be useful in other patient populations who are early in their illness. The strategy has implications for generalization of trial results to patients who do not consent to trial participation because it allows collection of information about patients who do not consent to randomization but who may receive the treatments being studied.


24 Exploring the Neurobiology of Glutamatergic and Cholinergic Dose-response in Schizophrenia Using Mechanistic Disease Modeling as a Novel Systems Pharmacology Approach

Hugo Geerts1,2, Athan Spiros1, Patrick Roberts1
1In Silico Biosciences, Berwyn, PA 2University of Pennsylvania, Philadelphia, PA

Background: Inverse U-shape clinical dose-responses are sometimes caused by off-target effects at higher doses, but often are a consequence of the underlying neurobiology at the target or at the circuit level.

Methods: Computer-based mechanistic disease modeling is a systems pharmacology approach that uses mathematical algorithms to simulate the firing dynamics of neuronal cells in biophysically realistic networks. The approach uses combines preclinical neurophysiology of relevant brain circuits and targets with human imaging studies, pathology and neuropharmacology and is therefore a real translational tool. The model is
calibrated by optimizing the correlation between retrospective clinical trial results (71 drug-dose combinations with 24 different antipsychotics) and the outcome of the same drug-dose combinations in the computer model. The model captures threefold more variance than the simple D2R occupancy measure. The mGluR2 affects synaptic glutamate levels as a presynaptic autoreceptor.

A detailed receptor state mathematical model of the a7 nAchR, based upon preclinical data captures the dynamical balance between activation and desensitization and the outcome is subsequently introduced in a cortical network of cognitive performance at the presynaptic Glu membrane and on GABA interneurons. For cognition, the model is calibrated using the result of 45 different interventions on cognitive readouts such as N-Back working memory tests or ADAS-Cog in Alzheimer’s disease.

Results: Stimulating the presynaptic mGluR2 autoreceptor leads to an inverse U-shape dose-response, because of the balance between excitation and inhibition in cortical network activity projecting to the N accumbens. Sensitivity analysis reveals that the optimal activation level and the shape of the dose-response curve depend upon a number of processes that can be modulated by various genotypes. Positive modulation of the a7 nAchR also leads to an inverse U-shape dose-response in the neuronal network because of the balance between activation and desensitization at the level of the receptor; the shape is dependent upon the ambient basal level of free Ach, which might be modulated by the comedication or the disease state. The modeling approach also identifies non-invasive endophenotypes, such as BOLDfMRI that might reflect the dose-response.

Discussion: The use of computer-based mechanistic disease-modeling in schizophrenia is a relatively inexpensive way to explore the issues associated with inverse U-shape dose-responses in CNS indications, based upon the underlying neurobiology. Identification of the processes that affect this neurobiology can lead to better clinical trial design and probably a higher success rate in clinical trials.

Feasibility of Assessing the Pragmatic vs. Explanatory Design of a Clinical Trial

Cynthia A. Bossie1, Joseph F. Hulihan1, Larry Alphs1

1Janssen Scientific Affairs, LLC., Titusville, New Jersey

Introduction: Effectiveness or pragmatic trials seek to answer the question “Does this intervention work under usual conditions?” whereas efficacy or explanatory trials focus on “Can this intervention work under ideal conditions?” The PRECIS (Pragmatic: Explanatory Continuum Indicator Summary) instrument has been developed to assist researchers in designing trials that are more pragmatic or explanatory (Thorpe et al, 2009). This pilot project explored the feasibility and limitations of adapting PRECIS to retrospectively rate a recently published antipsychotic trial along the pragmatic:explanatory continuum.

Methods: PRECIS identifies 10 study domains to be considered when designing trials as explanatory or pragmatic. For this work, PRECIS was modified to aid the post-hoc rating of studies on these domains along the pragmatic:explanatory continuum. To this end, a 7-point rating scale (0=extremely explanatory to 6=extremely pragmatic) with a total scoring range of 0 to 60 was added. To examine the construct validity of this instrument, a recently published trial in schizophrenic patients anticipated to have both effectiveness and pragmatic characteristics (Grimaldi-Bensouda et al, 2011) was rated independently on the modified PRECIS instrument by 3 raters (the authors). Average domain and total study scores were determined. After independent rating, the authors discussed the rationales for their ratings, identified areas of dispute, and identified scale limitations.

Results: Total modified PRECIS scores for the Grimaldi-Bensouda study by the 3 raters were 46, 48, and 52, with an average total score of 48.7. Six domains were rated as more pragmatic than explanatory with average ratings of 5.3-6.0; four domains showed elements of both designs with average ratings of 3.7; no domains were rated as explanatory. Variability of ratings among raters for each of the domains was limited. Limitations identified for this adapted instrument include: a lack of clear definition of each domain; insufficient definition of the anchors for scoring these domains; and inability of the instrument to assess the quality of study design, conduct, analysis, and interpretation relative to the trial objectives.

Conclusion: This pilot project demonstrated the feasibility of using a modified PRECIS to rank completed studies on the pragmatic:explanatory continuum. While scores were similar across raters for this study, this project revealed several limitations (e.g., more complete domain descriptors and anchors for rating) which should be addressed in follow-up work.
26 A Qualitative Evaluation on Evidence Level for Selective Imaging Biomarkers in Schizophrenia Clinical Trials

ISCTM Working Group for Workshop of Imaging Biomarkers in Schizophrenia: Chwen-Cheng Nathan Chen1,2, Susan De Santis3,4, Dong Jing Fu5, Daniel van Kammen6, Christine Moore7, Theo van Erp8, Henry Riordan9, Steven Potkin10

1PCBU, 2Pfizer Inc, Clinical Sciences, 3GE Healthcare, Princeton, NJ, 4Department of Psychiatry, NYU Langone Medical Center, New York, NY, 5Janssen Scientific Affairs, Titusville, NJ, USA, 6CNS Pharma Consulting, LLC, 7PharmaNeuTis, Cary NC, 8Department of Psychiatry & Human Behavior, School of Medicine, University of California Irvine, Irvine, CA, 9Medical & Scientific Affairs, Worldwide Clinical Trials, 10Department of Psychiatry & Human Behavior, University of California, Irvine, CA

Introduction: The main objective of this qualitative evaluation was to provide a review of imaging biomarkers in schizophrenia trials with regard to the utility of these markers as a measure of treatment response. Schizophrenia treatment studies that included the structural or functional imaging modalities including vMRI, fMRI, FDG-PET, receptor occupancy PET & SPECT, and DTI were identified, categorized and evaluated. We report on the results of this qualitative evaluation at the “evidence level” for these putative imaging biomarkers.

Methodology: Articles from human clinical trials, published after 1995 on the selected imaging biomarkers were selected from various databases including PsycInfo, MEDLINE, PubMed, and Google Scholar. A standardized information template was utilized during review of each article for relevant data collection including type of imaging biomarker, trial design, intervention, patient population characteristics, and major efficacy findings for image region of interest (ROI) as well as tolerability assessment. A qualitative evaluation of the study and biomarker was performed using grading of evidence level as outlined by Altar et al’s (2008) prototypical process for creating evidentiary standards for biomarkers and diagnostics. An evidence matrix was applied for evaluating articles including 1) theory of biological plausibility, 2) interaction with pharmacologic target, 3) pharmacologic mechanistic response, 4) linkage to clinical outcome of a disease 5) mathematical replication confirmation, 6) analytic validation, and 7) relative performance.

Results: 850 articles fulfilled the predetermined criteria. 81 articles received detailed review for evaluation. Evidence grading varied with imaging biomarkers, ranging from A- for receptor occupancy measured by PET or SPECT to D- for diffusion tensor imaging (DTI) with various supporting reasons for grading.

Conclusion: Different imaging biomarker modalities are required at the various phases of clinical development of schizophrenia treatment. Matching the imaging modality to the research question should result in better utilization of imaging biomarkers in clinical research that are more sensitive to treatment effects in a clinical trial setting.


27 A Tale of Two Trial Designs: Evaluation of Efficacy vs Effectiveness in Schizophrenia

Ferziger, R1, Mao, L2, Huilihan, J1, Bossie, CA1, Alphs, L1

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

Introduction: Public health decision making is informed by data on how interventions work under actual conditions of use (“real-world” or effectiveness data), which are captured in pragmatic studies. These studies differ from explanatory studies, which examine whether an intervention can work in more controlled circumstances (efficacy data). In any given trial, different design elements may be independently more pragmatic or more explanatory—an important consideration for the interpretation and generalizability of results. This analysis compares two studies that evaluate the same medication, using an instrument that categorizes various design domains along a pragmatic–explanatory continuum.

Method: Two studies that differed appreciably in terms of their objectives and overall approach were compared: An ongoing study of antipsychotic treatment for people with schizophrenia who have been incarcerated (study 1; PRIDE) and a placebo-controlled trial used for regulatory approval (study 2; Pandina et al, 2010). Studies were scored on a modification of the Pragmatic–Explanatory Continuum Indicator Summary (PRECIS) tool (Thorpe et al, 2009). Modified PRECIS rates 10 key design domains, each on a 7-point rating scale from 0 = extremely pragmatic to 6 = extremely pragmatic (total study score range 0-60). Five individual raters (the authors) independently scored the studies.
Results: The average (SD) modified PRECIS total score was 34.0 (6.36) for study 1 (rater scores: 26, 29, 35, 40, 40) and 5.8 (2.17) for study 2 (rater scores: 3, 5, 6, 6, 9). In study 1, three of the 10 domains had mean scores suggesting that they were predominantly pragmatic (~5), four had features of both approaches (~3), and three were predominantly explanatory (~2). All 10 mean domain scores for study 2 were predominantly explanatory (0-1) and were more explanatory than those for study 1.

Conclusion: This analysis demonstrates that study 1 (PRIDE study; antipsychotic treatment for people with schizophrenia who have been incarcerated) captures more pragmatic (effectiveness) information, while study 2 (controlled study supporting regulatory approval) captures more explanatory (efficacy) information. Some of the variability in the ratings was likely due to imprecise definitions of certain domains and inadequate anchoring for the ratings. Evaluation of studies using a modified PRECIS instrument can inform the use of clinical trial results for better public health decision making.

Support: Janssen Scientific Affairs, LLC
L Alphs, C Bossie, R Ferziger, and J Hulihan are employees of Janssen Scientific Affairs, LLC, and Johnson & Johnson stockholders. L Mao is an employee of Janssen Research and Development, LLC, and a Johnson & Johnson stockholder.

28 Methodological Issues Affecting Signal Detection in a Clinical Trial for Cognitive Impairment in Schizophrenia

Michael Hufford¹, Maria Gawryl², Nancy Dgetluck², Vicki Davis¹, Stephen Murray², Richard Keefe¹,³, Dana Hilt²

¹NeuroCog Trials, Inc. (Durham, NC), ²EnVivo Pharmaceuticals (Watertown, MA), ³Duke University (Durham, NC)

Introduction: A variety of methodological issues can significantly impact signal detection in clinical trials of novel pharmacotherapies for schizophrenia. We examined two methodological issues affecting signal detection in a clinical trial of EVP-6124, a selective, potent, oral nicotinic alpha-7 agonist being developed for cognitive impairments in schizophrenia.

Methods: A Phase 2b trial evaluated the safety and efficacy of two doses of EVP-6124 (0.3 mg and 1.0 mg qd) versus placebo among 319 chronic schizophrenia patients on stable second-generation antipsychotic drugs enrolled in the U.S., Russia, Ukraine and Serbia. Each patient was treated for 3 months.

Results: EVP-6124 was safe, well tolerated, and had clinically meaningful effects compared to placebo on cognitive performance (CogState and MATRICS Consensus Cognitive Battery [MCCB - in US subjects only]) and interview-based assessments (Schizophrenia Cognition Rating Scale [SCoRS]) as well as negative symptoms. We explored the temporal patterning of responses on the MCCB by examining what percentage of total change had occurred at the Day 44 vs. Day 84 visit. Placebo patients manifested 94% of their total change by the Day 44 visit, compared to only 61% and 56% for the 0.3 and 1.0 mg EVP-6124 groups, respectively. In short, nearly all of the placebo effect’s total impact on the MCCB was detected by Day 44, whereas both EVP-6124 groups continued to show substantial cognitive improvements in the latter half of the trial. We also examined whether the presence of an informant would help to provide a more sensitive test of cognitive impairments on the SCoRS. Consistent with previous research, we found that the presence of an informant helped to increase the effect sizes (ES) (1.0 mg EVP-6124 subjects vs. placebo with no informant: ES = .36; with an informant: ES = .51).

Conclusions: These findings have important methodological implications for the design and conduct of future clinical trials of cognitive enhancing agents in schizophrenia.

29 Measuring Quality of Life, Adiposity, and Sedentary Behavior in Antipsychotic-Treated Patients with Early Schizophrenia

Martin Strassnig MD¹

¹University of Miami

Objective: To examine how adiposity and sedentary behavior relate to quality of life (QoL) in patients with early schizophrenia receiving newer antipsychotic medication.

Methods: Cross-sectional study to assess adiposity by dual-energy x-ray absorptiometry (DXA) scans, habitual physical activity with the International Physical Activity Questionnaire (SF-IPAQ), and QoL by the RAND Medical
Outcomes Study SF-36. QoL scores were compared with age-adjusted normative population data. ANCOVA and regression analyses were used to examine determinants of QoL in early schizophrenia.

**Results**: There were 36 participants, with twenty-nine (72.5%) males. Age among participants was 25.1 ±3.6 years, range [19-34] and duration of illness was 30 ±18 months. Mean body mass index was 28.3 (±5). Patients had significantly lower physical functioning (p=0.0034), role physical (p=0.0003), general health (p=0.0001), vitality (p=0.03) and physical component scores (PCS) (p=0.003) than population comparisons. Habitual sedentary time, more than activity- or adiposity levels, was associated with QoL in early schizophrenia. Type of antipsychotic was not associated with QoL but with certain metabolic markers.

**Discussion**: Measures of QoL, adiposity, self-reported activity level and physical functioning are readily ascertained in patients with early schizophrenia. QoL is lower in early schizophrenia, and predominantly experienced as a physical problem. QoL is more strongly related to sedentary behavior than activity- and adiposity levels, or choice of antipsychotic. Targeted efforts to improve sedentary behavior therefore, can be hypothesized to improve QoL, a key clinical outcome.

### 30 Dose-Response Model of Lurasidone Treatment in Schizophrenia

Sunny Chapel, Ph.D.¹, Andrei Pikalov, M.D., Ph.D.², Yu-Yuan Chiu, Ph.D.², Jay Hsu, Ph.D.², Jane Xu, Ph.D.², Josephine Cucchiaro, Ph.D.², Antony Loebel, M.D.²

¹Ann Arbor Pharmaceutometrics Group, Inc (A2PG), Ann Arbor, MI. ²Sanovion Pharmaceuticals, Inc, Ft. Lee, NJ

**Introduction**: Characterization of dose-response relationships for new psychotropic agents may be difficult to determine based on results of individual clinical trials, due to various confounds such as variability in attrition and placebo-response rates. As a consequence, post-marketing changes in recommended therapeutic dosing ranges for antipsychotic drugs are not uncommon. The goal of this exposure-response analysis was to more precisely clarify dose-response effects for lurasidone.

**Methods**: Data were pooled from five 6-week, randomized, double-blind, placebo-controlled, once-daily, fixed-dose studies of lurasidone in the dosing range of 40-160 mg for the treatment of an acute exacerbation of schizophrenia. The PANSS and exposure data were fitted using the nonlinear mixed effects modeling methodology implemented in the NONMEM software (Version VI).

**Results**: In the final exposure-response model, LS mean change-from-baseline in PANSS exhibited a linear trend relative to dose of lurasidone. The 160 mg dose provided the greatest clinical benefit in terms of PANSS reduction relative to lower doses. In addition, the 120 mg dose produced improvement in PANSS that was intermediate between 80 mg and 160 mg. LS mean change in PANSS exhibited a linear trend relative to dose on treatment days 14, 28, 35, and 42. A time effect rate analysis indicated that 50% of the reduction in PANSS total score observed during acute treatment for each dose group occurred at 9 days. Between-study variability in clinical response was evident in the placebo group, but not in the lurasidone group, and was contributed to by demographic covariates (eg, age, weight, race). A log-linear hazard model indicated that patients were more likely to drop out when baseline PANSS scores were higher, and during the initial hospitalization period. However, dropout rate was not correlated with dose of lurasidone.

**Conclusions**: The effect of lurasidone was described using a linear dose-response model for drug effect, with increased treatment response observed at higher doses of lurasidone. Attrition was not correlated with lurasidone dose. Funded by Sunovion Pharmaceuticals Inc.

### 31 Augmenting SSRIs with an α4β2nAChR Partial Agonist: Lack of Efficacy in Insufficient Response Major Depressive Disorder

Fava, M.¹, Ramey, T.², Bell, J.², Li, X.², Boyer, S.², Davidson, W.², Billing, B.², Arroyo, S.²

¹Massachusetts General Hospital. ²Pfizer Inc., Neuroscience Research Unit

**Introduction**: This proof-of-concept study was aimed to investigate the efficacy, safety, and tolerability of an α4β2nAChR partial agonist (PA) in the augmentation of antidepressant therapy (ADT) with SSRIs in adults with Major Depressive Disorder (MDD) using a prospective study design.

**Methods**: A 6-week, double-blind (DB), outpatient, parallel group study was conducted in which the 6-week DB phase was preceded by a prospective 8 week open label (OL) SSRIs treatment. Subjects were in a current episode of MDD at least 8 weeks prior to enrollment into OL phase and with insufficient response to treatment. At the end of week 8, those
exhibiting a <50% improvement in the HAM-D17 and a total score ≥16, were randomized into the 6-week DB phase. Insufficient response to SSRIs treatment was augmented with the α4β2nAChR PA or placebo while continuing on ADT. Numerous design improvement initiatives were utilized and will be described. The primary efficacy endpoint was the change from DB baseline (week 8) in the MADRS at week 14. Independent remote interviews were utilized to confirm eligibility at screening and at week 8.

Results: In the OL phase, 297 subjects were treated with SSRIs, of whom 162 (54.5%) subjects were qualified to be randomized in the DB phase. When 113 of 198 planned subjects in double-blind phase (57%) have either completed or discontinued the study, an unblinded interim analysis for efficacy and safety was conducted and the stopping criteria for efficacy futility were met. In the final analysis, the treatment difference presented in LS mean ± S.E. of the α4β2nAChR PA vs. PBO was -1.30±1.565 with 2-sided 80% confidence interval (-3.32, 0.71), (2-sided p=0.4062). Placebo response in MADRS change from double-blind baseline at Week 6 of the double-blind phase, where LS mean ± S.E. = -8.30±1.088, was determined not to be a factor in the lack of drug effect. Further exploratory post-hoc analyses are currently ongoing.

Conclusions: There was no evidence of a treatment effect of α4β2nAChR PA vs. placebo in this study. Futility was declared and the study was terminated early. The drug was safe and well tolerated in this study.

32 Clinical Trial Design Assessing Augmentation for Cognitive Function in Major Depressive Disorder

Madhoo, M1, Keefe, RSE2, Roth, RM3, Sambunaris, A4, Wu, J1, Trivedi, M5, Anderson, CS1, Lasser, R1

1Shire Development Inc., Wayne, PA 2Duke University Medical Center, Durham, NC 3Dartmouth Medical School, Lebanon, NH 4Atlanta Institute of Medicine & Research, Atlanta, GA 5University of Texas Southwestern Medical School, Dallas, TX

Introduction: Cognitive dysfunction in major depressive disorder (MDD) has been reported, even during remission. However, studies assessing cognitive dysfunction in MDD may be affected by various factors, including study methodology and population, making data interpretation difficult. For example, earlier studies examined small populations of differing depressed state and disease severity, which may interfere with detection of cognitive function changes. We describe a clinical trial design that addresses some limitations associated with earlier studies of cognition in MDD.

Methods: Common methodologic limitations (e.g., small sample size, inclusion of actively depressed individuals, use of disparate assessment tools) were identified based on a literature review of cognitive function in MDD. The trial described here measured the effects of lisdexamfetamine dimesylate (LDX) augmentation on cognitive function in fully- or partially-remitting individuals on selective serotonin reuptake inhibitor (SSRI) monotherapy with persistent cognitive dysfunction because data in mildly depressed and euthymic populations are limited.

Results: Key features of this randomized, placebo-controlled trial included a multicenter design (27 US sites), a 2-week screening period to ensure symptom stability and participant eligibility, a 9-week active treatment duration, and the use of augmentation in participants exhibiting full or partial remission with their current SSRI monotherapy. To facilitate detection of cognitive function changes independent of depressive symptoms (state- vs trait-dependent changes), individuals with mild depressive symptoms (Montgomery-Asberg Depression Rating Scale total score ≤18) and executive dysfunction (Behavior Rating Inventory of Executive Function-Adult Version [BRIEF-A] Global Executive Composite [GEC] T-score ≥60) were included. Multiple measures assessing subjective executive and cognitive function (BRIEF-A self-report [primary endpoint] and informant-report [secondary endpoint]) and objective neurocognitive function (CNS Vital Signs test battery [secondary endpoint]) facilitated broad examination of cognition; objective and subjective functional outcomes supported the study’s signal-finding ability.

Conclusions: This trial featured a targeted, subjective patient-reported outcomes design to enhance detection of cognitive improvement in fully- or partially-remitting individuals with MDD. Positive findings on the primary outcome (BRIEF-A GEC T-score) support the methodology’s robustness. However, the possible influence of affective symptoms on subjective assessment tools should be considered when interpreting the findings. Future trials employing this design are needed and key endpoints for assessment of cognition and depressive symptoms should be established.
A Virtual Study Design to Reduce “Real World” Challenges of a Depression Study

Bryce Kasuba1, Jay Lombard1, Roy Perlis2, Herb Harris1, Lauren Novasitis1, Keya Watkins3, Rachel Dicker1

1Genomind, 2Massachusetts General Hospital, 3ePharmaSolutions

Introduction: Collecting outcomes data from subjects in actual practice settings poses challenges not normally presented in typical research studies. This study employs design features to enable collection of patient-reported outcomes (PROs) data in real-world settings with clinicians not generally experienced in clinical research. This research study offers unique designs to overcome these challenges, including electronic consent and secure online capture of study assessments. Aims include characterization of the patient population selected for pharmacogenetic testing by their clinicians and demonstration of the impact of genetic testing on clinician treatment decisions and patient outcomes as measured by change in severity of depression.

Methods: This is an open label, prospective analysis of clinicians who order the Genecept™ Assay and patients for whom the test is ordered. The study will electronically consent and collect responses from clinician study participants and subjects and analyze genetic data from these subjects. IRB approved outreach is done when clinicians order a pharmacogenetic test for a patient meeting depression criteria. Analytic results reports will be provided to clinician study participants, who will complete a baseline survey which asks about the Assay’s influence on patient treatment. The genes tested include the serotonin transporter protein (SLC6A4), gated calcium channel (CACNA1C), dopamine receptor subtype two (DRD2), catechol-O-methyl transferase (COMT) and methylentetrahaldolate reductase (MTHFR) as well as cytochromes P450 2D6 (CYP2D6) and 2C19 (CYP2C19). Clinicians will record subjects’ psychiatric history and severity of illness using Clinical Global Impressions -Severity (CGI-S). A follow up assessment occurs at 3 months which includes an assessment of improvement in illness severity using CGI-I. Subjects will be asked to complete at baseline and 3 months an assessment of depression symptoms, and quality of life. All analyses will be performed in SAS and PROC MIXED will be used for the primary hypothesis.

Results: This study is registered on clinicaltrials.gov
http://www.clinicaltrials.gov/ct2/show/NCT01507155?term=genomind&rank=1
Recruitment begins February 2012; 100 Subjects are expected to be recruited by June 2012.

Conclusion: Conducting clinical trials in non-research settings is challenging, but affords the opportunity to collect data relevant to real-world practice. While clinicians are interested in being part of research and offering innovative treatment, many barriers exist including training, ICH/GCP regulations, IRB submission processes and study-sponsor agreements. This study design offers a novel paradigm to quickly operationalize a clinical trial through electronic Informed Consent and will be the model for our future disease specific studies.

Disclosure Statement: One or more authors report potential conflicts which are described in the program. Bryce Kasuba, Lauren Novasitis and Rachel Dicker are employed by Genomind, Jay Lombard is an employee of and has equity interests in Genomind, Herb Harris is a consultant for Genomind and Roy Perlis serves on the Scientific Advisory Board for Genomind. Keya Watkins is employed by ePharmaSolutions.

A Validation Study of Suicidality Scales: Consistency of Mapping to the Columbia Classification Algorithm of Suicide Assessment

Mao, L1, Fu, D-J2, Sheehan, DV3, Lindenmayer, J-P4, Alphs, L1

1Janssen Research and Development, LLC, Titusville, NJ, USA, 2Janssen Scientific Affairs, LLC, Titusville, NJ, USA, 3University of South Florida College of Medicine, Tampa, FL, USA, 4New York University, New York, NY, USA

Introduction: Assessment of suicidal ideation and behavior (SIB) represents a critical safety evaluation for psychiatric and neurologic treatments where suicidality is a concern. The FDA now mandates that information on SIB collected in clinical trials map to the Columbia Classification Algorithm of Suicide Assessment (C-CASA) to ensure consistency in SIB classification across studies. Currently, the Columbia Suicide Severity Rating Scale (C-SSRS) represents the “gold standard” for collecting SIB information that is mapped to the C-CASA. However, investigators need a wider variety of assessment instruments that will meet a diverse range of trial needs. Two valuable and operationalized alternatives to the C-SSRS are the InterSePT Scale for Suicidal Thinking-Plus (ISST-Plus), and the Sheehan Suicidality Tracking Scale (S-STS). This report describes a methodological approach to validation of the ISST-Plus and the S-STS as compared with the C-SSRS for comparative mapping of symptoms of suicidal ideation, self-harm, and suicidal behavior on the C-CASA.

Methods: Individuals aged >=19 years with a broad range of SIB are identified from acute visits to an
emergency room for psychiatric reasons, acute in-patient psychiatric unit, or psychiatry clinical research office settings. Consenting subjects are then interviewed by trained mental health professionals with experience working with suicidal patients, using semi-structured interviews developed for each of the 3 suicide assessment instruments (C-SSRS, IST-Plus, S-STS) employing standardized definitions of all critical SIB terms. Information and ratings from subjects who have a complete set of responses on all three instruments will be mapped to the C-CASA by computer coded mapping, or by a trained individual at the site using predefined C-CASA mapping algorithms for all three instruments.

**Results:** Approximately 40 subjects with various degrees of severity of SIB will be recruited for this study. Simple and overall kappa coefficients are computed for each mapping variable and are used to quantify the degree of agreement between the C-CASA and the IST-Plus, C-SSRS, and S-STS. Inter-rater reliability is assessed using kappa statistics as well as the intra-class correlation coefficients for each scale.

**Conclusion:** If mapping validity for all three efficacy instruments is established, this study will demonstrate that three alternative approaches to SIB evaluation are sufficiently consistent with the C-CASA to allow valid overall summarization of data while permitting a more diverse choice of alternatives for SIB assessment.

Supported by Janssen Scientific Affairs, LLC

**Disclosures:**
L Alphs and DJ Fu are employees of Janssen Scientific Affairs, LLC, and Johnson & Johnson stockholders. L Mao is an employee of Janssen Research and Development, LLC, and a Johnson & Johnson stockholder. J-P Lindermayer has received grant/research support from Janssen Scientific Affairs, LLC, Eli Lilly and Company, AstraZeneca, Johnson & Johnson, Pfizer, BMS, Otsuka, Sunovion, and Roche and reports receiving consultant fees from Janssen Scientific Affairs, LLC, Roche, and Shire. D Sheehan reports the following disclosures (1 – consultant, 2 – grant/research support, 3 – lectures/presentations, 4 – stockholder): Abbott Laboratories[1, 2, 3], Ad Hoc Committee, Treatment Drug & Assessment Review[1], Alexza[1], Alza Pharmaceuticals, Palo Alto, California[1], American Medical Association[2], American Psychiatric Association Task Force on Benzodiazepine Dependency[1], American Psychiatric Association Task Force on Treatments of Psychiatric Disorders[1], American Psychiatric Association Working Group to revise DSM III Anxiety Disorders Section[1], Ancoite Foundation[2], Anxiety Disorders Resource Center[1], Anxiety Drug Efficacy Case, U.S. Food & Drug Administration[1], Applied health Outcomes/XCENDA[1], AstraZeneca[1, 2, 3], Avera Pharmaceuticals[1, 2], Boehringer Ingelheim[3], Boots Pharmaceuticals[3], Bristol-Myers Squibb[1, 2, 3], Burroughs Wellcome[2, 3], Cephalon[1], Charter Hospitals[3], Ciba Geigy[3], Committee (RRC) of N.I.M.H. on Anxiety and Phobic Disorder Projects[1], Connecticut & Ohio Academies of Family Physicians[1], Cortex Pharmaceutical[1], Council on Anxiety Disorders[1], CPC Coliseum Medical Center[1], Cypress Bioscience[1], Distal Products Company[3], Division of Drugs & Technology, American Medical Association[1], EISAI[1, 2], Eli Lilly[2, 3], Excerpta Medica Asia3, Faxmed, Inc[1], Forest Laboratories[1, 2], Glaxo Pharmaceuticals[3], GlaxoSmithKline[1, 2, 3], Glaxo-Wellcome[2], Hickm Pharmaceuticals[3], Hospital Corporation of America[3], Humana[3], ICI, INC Research[1, 3], International Clinical Research (ICR)[2], International Society for CNS Drug Development (ISCDD)[1], Janssen Pharmaceutical[1, 2, 3], Jazz Pharmaceuticals[1, 2], Kali-Duphar[2, 3], Labopharm-Angellini[1, 2, 3], Layton Bioscience[1], Lilly Research Laboratories[1], Lundbeck, Denmark[1], Marion Bioer Dow[3], McNeil Pharmaceuticals[3], Mead Johnson[2, 3], Medical Outcome Systems[4], MediciNova[1, 2], Merck Sharp & Dohme[2, 3], National Anxiety Awareness Program[1], National Anxiety Foundation[1], National Depressive & Manic Depressive Association[1], National Institute of Drug Abuse[2], National Institute of Health (NIH)[2], Neurionetics[1], NovaDel[1], Novartis Pharmaceuticals Corp[2], Novo Nordisk[3], Organon[1, 3], Orion Pharma[1], Parke-Davis[2, 3], Pfizer[1, 2, 3], Pharmacia[1], Pharmacia & Upjohn[1, 3], PharmaNeuroBoost[1, 3], Philadelphia College of Pharmacy & Science[1], Pierre Fabre, France[1], Quintiles[1], Rhone Laboratories[3], Rhone-Poulenc Rorer Pharmaceuticals[3], Roche[1], Roerig[3], Sagen[1], Sandoz Pharmaceuticals[2, 3], Sanofi-Aventis[1, 2, 3], Sanofi-Synthelabo Recherche[1, 2], Schering Corporation[3], Septra[1], Shire Laboratories, Inc[1], SmithKline Beecham[1, 2, 3], Solvay Pharmaceuticals[1, 3], Takeda Pharmaceuticals[1], Tampa General Hosp[1], University of South Florida Psychiatry Center[2], University of South Florida College of Medicine, TAP Pharmaceuticals[2, 3], Targacept[1], TGH-University Psychiatry Center[3], Tikvah Therapeutics[1], Titan Pharmaceuticals[1], United Biocare[1, 3], The Upjohn Company[1, 2, 3], U.S. Congress-House of Representatives Committee[1], USF Friends of Research in Psychiatry, Board of Trustees[1], Warner Chilcott[2, 3], Wyeth-Ayerst[1, 2, 3], Zambon[1, 2, 3].
Hippocampal Volume Predicts Response to Riluzole in a Pilot Cohort of Patients with Generalized Anxiety Disorder

Chadi G. Abdallah¹,², Jeremy D. Coplan², Andrea Jackowski³, Xiangling Mao⁴, Dikoma C. Shungud⁴, Sanjay J. Mathew⁵

¹ Department of Psychiatry, Yale University School of Medicine ² Department of Psychiatry, SUNY Downstate Medical Center ³ Departamento de Psiquiatría, Universidade Federal de São Paulo, SP, Brazil ⁴ Departments of Radiology, Psychiatry and Biophysics, Weill Medical College of Cornell University ⁵ Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine

Background: Using riluzole in patients with generalized anxiety disorder (GAD), we previously found differential changes in hippocampal NAA concentrations, which covaried positively by the degree of therapeutic benefit. In the current study, we investigated for positive association between hippocampal volume and hippocampal NAA in the context of riluzole response in GAD.

Methods: Eighteen medication—free adult patients with GAD received 8-week of openlabel riluzole. Ten healthy subjects were matched control. Participants were submitted to magnetic resonance imaging and spectroscopy at baseline and at the end of week 8. Patients who completed all interventions were classified as remitters (n = 7) or minimal—responders (n = 6), based on achieving Hamilton Anxiety Rating Scale (HAM —A) scores ≤ 7.

Results: At baseline, GAD patients had significant reduction in total hippocampal volume compared to healthy subjects (F(1,21) = 6.55, p = 0.02). This reduction was most pronounced in the remitters, compared to minimal—responders and healthy subjects. Delta (final - baseline) hippocampal volume was positively correlated with delta NAA in GAD. This positive association was highly significant in the right hippocampus in GAD [r = 0.81, p = 0.002], with no significant association in healthy subjects [Fisher r —to—z p = 0.017]. Across all GAD patients, delta hippocampal volume was positively associated with improvement in HAM —A (rspearman = 0.62, p = 0.03).

Conclusion: The strong correlation between hippocampal volume, NAA, and response to riluzole provides supportive evidence for the utility of hippocampal NAA and volume as biomarkers to assess response to glutamate—based drugs, a finding that warrants replication in an expanded sample.

More Experienced Raters Demonstrate Better Concordance on the MADRS as Compared to MADRS-S Scores

Theresa A. Bromley PhD¹, Marian A. Ormont MD¹, Nicolas Troyano PhD¹

¹ePharmaSolutions

Background: The use of patient-reported outcomes as a secondary endpoint in clinical trials has been increasing. As patients become more involved in their own disease assessment and management, it is important that the clinician be sensitive to the ways in which patients describe their symptomatology and overall impression of illness. In a recent depression clinical trial, clinicians rated the Montgomery-Asberg Depression Rating Scale (MADRS-CR) at 10 study visits. During each visit, the subjects were asked to complete the self-administered version of the MADRS (MADRS-S). Pearson Correlations were completed on these scores for 638 visits. It was predicted that scores from more experienced MADRS-CR raters would have higher correlations with the subject-rated instrument. Results indicate that more experienced MADRS-CR raters have higher correlations with the MADRS-S, particularly in less depressed patients where more probing of responses may be needed.

Methods: The scores from 638 MADRS-CR and MADRS-S were analyzed from 10 study visits. Inclusion criteria for the study was a HAM-D-17 score of > 18. Sixty-two raters and 175 subjects were included from US clinical research sites. Scores were collected and analyzed. Pearson correlations were run on each comparison.

Results: Results indicate that raters with more than 2 years experience with the MADRS-CR had significantly higher correlations with MADRS-S scores. These correlations were higher when overall scores on both the MADRS-CR and MADRS-S were lower reflecting less depressive symptomatology.

Conclusion: Most raters were sensitive to subjects’ symptomatology; scores on the MADRS-CR tended to track
scores on the MADRS-S form week to week. More experienced raters demonstrated higher agreement with subjects’ self-ratings, particularly on subjects reporting less depressive symptoms. Correlations between Clinician-rated and Subject-rated scores on individual MADRS varied greatly, with the lowest correlations exhibited for sleep and appetite items. Some research has shown that alignment between clinician and patient reports of symptoms are more positively correlated with compliance. Therefore, clinicians that are sensitive and aligned to patient reported symptoms may improve drug compliance and retention in clinical trials.


The authors report no conflicts of interest for this work.

37 Neuroimaging Correlates of Serum BDNF Concentrations and Cognitive Function in Patients with Coronary Artery Disease

Swardfager W1,2, Herrmann N1,2, Saleem M1,2, Ranepura N1, Scott CS1, Black SE1,2,3 Oh PI1,2,3, MacIntosh BJ1,2 & Lancot KL1,2,3

1 Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada 2 University of Toronto, Toronto, Ontario, Canada 3 Toronto Rehabilitation Institute, Toronto, Ontario, Canada

Background: Treatment of depression in late-life is often complicated by poor remission rates and the involvement of cognitive symptoms, which can predict non-response to currently available therapies. Smaller temporal lobe volume has been implicated in both depressive and cognitive symptoms in late-life. A vascular etiology has been suggested to underlie these neurobiological findings; accordingly, patients with coronary artery disease (CAD) are at least two-fold more likely to suffer from depression, they demonstrate an increased rate of decline in cognitive performance and they show increased cerebral atrophy. This study examines possible volumetric and neuroimaging correlates of depressive and cognitive symptoms in patients with CAD. Moreover, serum brain derived neurotrophic factor (BDNF), a possible biomarker for depression, will be assessed as a correlate of neuroimaging findings.

Methods: Consecutive patients with CAD undergoing cardiac rehabilitation were approached. Depressive symptoms were assessed using the Centre for Epidemiological Studies Scale for Depression (CES-D) and depressive episodes were diagnosed using the Structured Clinical Interview for DSM-IV. A cognitive battery consisting of the Mini Mental Status Examination (MMSE), California Verbal Learning Test 2nd Ed., Stroop, Trail Making Test Part B and the Digit Symbol-Coding task was administered. Anatomical T1, fluid-attenuated T2, diffusion tensor (DT) and pseudocontinuous arterial spin labeling (pcASL) MRI sequences were performed to assess regional brain volume, extent of white matter disease, microstructural white matter integrity, and cerebral blood flow, respectively. Brain regions and tissue classes were segmented by a semi-automated procedure. Serum BDNF concentrations were assayed from fasting blood samples by enzyme-linked immunosorbent assay.

Results: Forty-six of 60 subjects (85% male, mean age 65.2±9.4 yr) have been recruited. The mean (±SD) MMSE score was 28.9±1.8 and the mean CES-D score was 8.1±9.1. Twenty percent of patients met DSM-IV criteria for depression. In 17 subjects for whom BDNF assays and tissue segmentation have been performed, BDNF concentrations were associated with the brain to cerebrospinal fluid ratio in the left anterior (p=.669, p=.003) and left posterior (p=.485, p=.045) medial temporal lobe.

Conclusions: Lower serum BDNF may be a risk factor for medial temporal atrophy in patients with CAD. Possible novel neuroimaging biomarkers will also be explored. These data suggest further studies to determine if baseline BDNF concentrations can predict outcomes in clinical trials targeting depressive and cognitive symptoms in patients where established vascular disease may be a factor.

38 Adaptive versus Traditional Design Approaches to Target Dose Estimation

Ibrahim Turkoz1, Vlad Dragalin2, Jose Pinheiro1, Judy Kando3, Radhi Abdulnabi1, Tom Parke5, Olga Marchenko6, and Ginger Haynes7 for the ISCTM Adaptive Design Working Group

1Janssen R&D, LLC, 2Aptiv Solutions, 3Sunovion, 43i Global, 5Tessella, 6Quintiles, 7Eli Lilly and Company

Introduction: A number of approved effective treatments for schizophrenia exist. Developers must better understand
Methods: A phase II study for a hypothetical schizophrenia compound was characterized as follows. The primary outcome was change from baseline to endpoint in PANSS total score. The study was placebo-controlled, 6 weeks in duration, with 12 months enrollment. Minimal effectiveness was defined as a 10 point or greater difference from placebo (Standard Deviation=20). Four dose response curves, three active dose responses and one flat dose response, were specified to compare traditional and adaptive designs (TD and AD).

The TD included 3 active doses, 3 mg, 9 mg, and 15 mg with 64 subjects per arm (N=256) to achieve 80% power for identifying the minimal effective dose (MED), resulting in 4 scenarios, one for each curve. A single one-way ANOVA model was used to estimate the mean response by dose.

The AD included 6 active doses, 1.5 mg, 3 mg, 6 mg, 9 mg, 12 mg, and 15 mg, for a total 448 subjects. AD simulations included 2-5 interim looks, resulting in 16 scenarios. The AD used the Normal Dynamic Linear Model to estimate the dose-response relationship and change the allocation ratios at interim looks. The trial started with equal allocation across doses. At each interim, the study would be stopped for futility if the posterior probability that treatment effect is smaller than 5 units is less than 0.1 for all doses. The allocation ratio to each dose was aimed to minimize the variance of the estimated mean response at the minimum effective dose. The allocation to placebo remained fixed at 64 subjects throughout the trial.

Two hundred datasets were simulated for each scenario under each design. The designs were compared as to their efficiency and accuracy in identifying the MED.

Results: A MED was identified more frequently with AD (87-99%) than TD (39-93%). AD was more efficient than TD in identifying the MED as 32% to 107% more subjects would be required with TD to achieve the same estimation accuracy at the target dose as with AD.

The identified MED PANSS score was within 10% of the true MED dose PANSS score more frequently for AD than TD (67%).

Conclusion: In this simulation, AD was more efficient and accurate in identifying the MED in most scenarios. Results of additional simulations will be presented at the poster session.

Draft FDA Decision Tree for Assessment of Abuse Potential

Bonson KR1, Sun SW1

1Controlled Substance Staff, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD

Introduction: For central nervous system (CNS)-active drugs, an important aspect of evaluating safety is determining if the drug has abuse potential. If abuse potential can be identified early in the development process, it will be possible to plan for, describe and manage the inherent safety issues while the drug is being brought to market.

FDA has drafted a decision tree to provide a model that will improve regulatory efficiency, consistency, and transparency in the assessment of drug’s abuse potential.

The tree is based upon the draft Guidance for Industry: Assessment of Abuse Potential of Drugs (2010) and aligns with real-world clinical development and allows for active dialogue with and feedback from stakeholders.

The decision tree is not intended to determine whether a drug has abuse potential and whether it should be recommended for scheduling under the Controlled Substances Act. These decisions cannot be made until data submitted in an NDA have been fully reviewed by FDA.

Methods: As drug development progresses from nonclinical to clinical, various scientific questions related to abuse potential are answered through the outcomes of each of the studies.

Throughout this process, there are three major “Abuse Decision Points” in which the “abuse signals” resulting from study outcomes help answer the following questions:

1) Is the drug (or major metabolite) CNS-active?
2) Is a human abuse potential study needed?
3) Do the abuse-related data in the NDA show that the drug has abuse potential?

This decision tree is designed for and limited to the evaluation of new molecular entities, as well as other drugs that have not previously undergone an abuse assessment in the U.S.

**Results**: The majority of studies listed in the decision tree are required for all drugs as part of the general safety evaluation.

There are only three studies that are dedicated to the assessment of abuse potential, including:

- Two nonclinical studies (self-administration and drug discrimination)
- One clinical study (human abuse potential study).

**Conclusions**: The assessment of abuse potential occurs throughout the nonclinical and clinical evaluation of a drug’s safety and efficacy during the drug development process.

A simplified and transparent process such as a stepwise decision tree will aid stakeholders in having a common understanding of the necessary evaluations in the assessment of abuse potential.

---

**Assessing Longitudinal Changes in Cognitive Function in the Elderly**

Keith A Wesnes¹, Brian K Saxby², Gary Ford²

¹Bracket Global, Goring-on-Thames, RG8 0EN, U.K. & Centre for Human Psychopharmacology, Swinburne University, Melbourne, Australia
²Institute for Ageing and Health, Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne, NE4 5PL, U.K.

**Introduction**: The deficits to cognitive function which occur in normal ageing can potentially be treated with pharmaceutical products. Further, as criteria have now been proposed for pre-clinical dementia (Sperling et al., 2011), trials are now being planned with compounds designed to prevent or reduce cognitive decline in groups of ‘healthy volunteers’ identified to be at risk of developing Alzheimer’s disease. However, in order to conduct such trials, cognitive tests need to be employed which can reliably assess such change. In the present study a cohort of healthy elderly volunteers was assessed yearly over a 5 year period with a computerised test battery.

**Methods**: The CDR System is a computerised set of 9 tests of attention, working and episodic memory which has been widely used in trials of potential cognition enhancers in healthy volunteers, age-related cognitive decline, MCI and the dementias. 256 normotensive volunteers (113 females), mean age 76 years (range 70 to 90), mean MMSE 28.8 (range 23 to 30), were trained on the CDR System twice before a baseline was established, and then retested yearly for up to 5 years.

**Results**: Validated composite factor scores were derived from the various test measures. Performance was found to decline significantly over the study period on four of the five scores: power of attention (p<0.0001), quality of episodic recognition memory (p<0.0002), quality of working memory (p<0.015) and speed of retrieval of information held in memory (p<0.0001). Power of attention showed significant deficits from year one onwards, two other measures showed deficits by year one, and all showed significant deficits from year three onwards.

**Conclusions**: This study has demonstrated that the use of validated and sensitive tests of cognitive function can detect decline over a 5-year period in healthy elderly volunteers. Such testing is therefore fit for purpose for the evaluation of treatments aimed at preventing or even reversing age-related declines in cognitive function, as well as treatments which may delay the onset of Alzheimer’s disease in high risk but otherwise healthy populations.

**Disclosure**: No funding was received for the conduct of this research. The main author reports potential conflicts which are described in the program.


---

**Assessing Cognitive Function via the Internet**

Keith A Wesnes¹

¹Bracket Global, Goring-on-Thames, RG8 0EN, U.K. Centre for Human Psychopharmacology, Swinburne University, Melbourne, Australia

**Introduction**: Many large scale registry trials in normal and pathological aging are being planned or conducted. In Alzheimer’s disease, interest is turning to prevention studies which may be conducted in healthy populations identified
to be at risk of developing Alzheimer’s disease. Automated cognitive function testing may be a practical solution for such work, and the present study investigates the utility of administering such testing via the internet.

**Methods:** Four tests from a computer based cognitive methodology, the CDR System, were internet enabled: simple reaction time, choice reaction time, digit vigilance and delayed picture recognition. Participants logged on to a website, entered their age and gender, and performed the tests on-line. Their data were compared to normative data from the standard administration of these CDR tests.

**Results:** A total of 52,237 individuals aged 18 or over performed at least one of the tests. 46.5% of the population was aged 18 to 40 years, 41.7% 41 to 60, and 11.8% 61 or older. There were highly significant declines with increasing age on the measures of speed on all tasks, as well as for the ability to correctly identify the pictures. Further, variability in reaction times increased with age, as did cognitive reaction time (the difference between choice and simple reaction time). The declines from 18 to 25 years to successive five year bandings (eg 26 to 30, 31 to 36 etc) were generally comparable between the internet based testing and the standard administration. Also performance on a number of measures was directly comparable between the two forms of administration.

**Conclusions:** This study has shown that large cohorts can be assessed using internet based cognitive tests, and that the general performance on these tests is directly comparable to that from the same tasks administered in the standard fashion. Notably, rates of decline with aging were directly comparable, as were the patterns of declines on various measures. These findings suggest that internet based cognitive testing is a viable technique in large patient trials, and should prove a useful and convenient means of longitudinally assessing cognition in patient registry studies or large long-term clinical trials.

**Disclosure:** The main author reports potential conflicts which are described in the program.

---

**Ratings Variability of UPDRS part III Sub-items in Multinational Parkinson’s Disease Clinical Trials**

Swartz JE¹, Xu Y², McNamara C²

¹Bracket, Goring-on-Thames, UK, ²Bracket, Wayne, PA, USA

**Introduction:** Training and certification on the UPDRS part III, the gold-standard primary motor assessment endpoint in Parkinson’s disease (PD) clinical trials, is notoriously difficult. Therefore implementation of a meticulously structured training and certification programme before commencing any multinational PD study, to ensure uniformity, reliability and consistency of performance across regions, sites and investigators, is essential. We present data from two Phase III trials, to assess whether UPDRS part III item subtype, degree of severity of items trained upon and method of training delivery (live versus remote) influenced the overall concordance of ratings.

**Methods:** Rater training consisted of didactic and interactive presentations on the UPDRS. Case examples and videotaped patient vignettes were scored and feedback provided. Raters were considered “qualified” to rate if they met predefined experience criteria relating to educational background, experience with the PD population and scale, and demonstrated acceptable concordance with expert scores on a UPDRS part III videotaped patient interview. Acceptable scores were established by a panel of five PD experts.

The frequency of discrepancy of all UPDRS part III sub-items was compared initially and further classified by certification status and training delivery. Chi-square test was used for testing differences of discrepancy by certification status. Within each certification status, the effect of training delivery on rating agreement was examined by Fisher’s exact test.

**Results:** Certification data on 608 raters from two different PD studies were pooled and analysed. Of the 22 UPDRS part III sub-items assessed, seven items showed marked scoring discrepancies, with > 20% of raters disagreeing with the expert-agreed score range. All seven items demonstrated intermediate severity range. Those raters initially listed as ‘questionable’ demonstrated a significantly greater disagreement rate upon scoring items when compared to ‘qualified’ raters (p-values < 0.05 for all items). This was particularly evident in several of the intermediate severity range sub-items. Live training, conducted at Investigators’ meetings, significantly improved agreement in both ‘qualified’ and ‘questionable’ raters for these intermediate severity items, compared to remote (on-line) training methods.

**Conclusions:** Certain UPDRS part III sub-items are subject to greater assessment and scoring variability than others. Intermediate severity range sub-items generally appear especially difficult to score, generating the most discrepant
ratings.

These results highlight the benefits of a rigorous training programme and the need to focus on items in the more controversial intermediate range of severity, when training and certifying on the UPDRS part III. Further consideration to training delivery is warranted and may decrease ratings variability.

**Low-Contrast Visual Acuity Correlates with Cognitive Performance in Relapsing-Remitting and Secondary-Progressive Multiple Sclerosis**

Todd Feaster PsyD

1Bracket Global

**Background**: Visual loss is one of the most common and disabling physical symptoms in multiple sclerosis (MS). Several clinical trials are researching low-contrast letter acuity charts in conjunction with optical coherence tomography (OCT) for their utility in tracking disease progression via retinal nerve fibre layer thickness and the process of neurodegeneration in MS. However, little if any research has been conducted on the potential utility of low-contrast visual acuity as a means of examining cognition in MS.

**Objectives**: The primary objective is to determine whether there are any relationships of low-contrast visual acuity with cognitive performance within a comprehensive neuropsychological battery.

**Methods**: Sixty-five patients with relapsing-remitting MS (RRMS) and secondary progressive MS (SPMS) were administered a comprehensive neuropsychological evaluation assessing emotional functioning, memory, executive functioning, processing speed, attention, visuospatial abilities, motor skills, and verbal fluency. Visual examination included an assessment of low-contrast visual acuity (5%, 2.5%, and 1.25% illuminated charts), near visual acuity and visual acuity.

**Results**: Sloan low-contrast letter charts (SLCLC) exhibited correlations with visual and non-visual neuropsychological tests as well as correlations with physical performance in RRMS and SPMS independent of age. This relationship was strongest among tests of motor functioning and processing speed. For example, SLCLC significantly correlated with performance on the Nine-Hole Peg Test \( r = 0.51, P < .001 \), 25-Foot Timed Walk \( r = .48, P < .001 \), Symbol Digit Modalities Test (SDMT) \( r = 0.51, P < .001 \) and Verbal Fluency \( r = .35, P < .01 \).

**Conclusions**: Low-contrast visual acuity plays an important role in everyday life. It has been suggested that a measure of visual function should be strongly considered for addition to the Multiple Sclerosis Functional Composite Test (MSFC). Low-contrast visual acuity may be promising tools for detecting both physical and cognitive changes in patients with RRMS and SPMS and should be examined more closely in larger, longitudinal studies.

**Symptom Severity Rating in Restless Legs Syndrome (RLS): Validation of the RLS-6 Scales**

Ralf Kohnen1, on behalf of the German RLS Expert Group*

ReSearch Pharmaceutical Services, Forth Washington, PA, USA & Psychology Department, University Erlangen-Nuremberg, Nuremberg, Germany

**Introduction**: Restless Legs Syndrome (Willis-Ekbom Disease, RLS) is a sensori-motor disorder which affects approximately 10% of the population in Western countries. Patients are suffering from an urge to move with or without dysesthesias which is engendered by rest, relieved by movement and follows a circadian rhythm (worse in the evening and at night). Disturbed sleep, impaired mood, cognitive impairment as well as low quality of life are long-term consequences of this chronic disease. The RLS-6 scales have been used as efficacy endpoints to assess relief from symptoms in various clinical trials globally for different RLS treatments. New data on the validity and reliability of the RLS-6 scales are presented.

**Methods**: The RLS-6 scales are 6 global self-rating scales using 11 numerical categories from 0 = “no symptoms” to 10 = “maximum” severity. Four scales assess the severity of RLS “at bedtime”, “during the night”, “during daytime when at rest”, and “during daytime when engaged in activities”. Additionally, two scales request patients’ “satisfaction with sleep”, and severity of “daytime tiredness/sleepiness”. All scales are evaluated separately, no total or sub-scale scores are intended.

**Findings**: Evaluation was performed with the pooled data of four clinical trials on efficacy of different dopamine agonists (893 subjects). Convergent validity: Mean baseline correlations between the RLS-6 items and the International RLS Severity Rating Scale (IRLS) total score were highest for “severity during the night” (0.65) and smallest for “severity during the day when engaged in activities” (0.38). For change from baseline to LOCF correlations were
similar, ranging from 0.30 (severity during the day when engaged in activities) to 0.74 (severity during the night). Discriminant validity: All RLS-6 scores were different (p< .0001 for all tests) for three severity classes according to the Clinical Global Impression severity scale at baseline (moderate, severe, very severe). Reliability: There was a high test-retest-reliability on the basis of a one day interval, ranging from 0.63 (satisfaction with sleep) to .96 (severity at bedtime).

**Conclusion:** The RLS-6 scales show a high convergent and discriminant validity. The scales are able to identify changes over time and their test-retest-reliability is high. In addition to the IRLS scale, the RLS-6 scales may point to special qualities of treatments, like influence on daytime symptoms and can be used in diaries.

*D German RLS Expert Group: Heike Beneš MD, Schwerin, Wolfgang H Oertel MD, Marburg, Karin Stiasny -Kolster MD, Marburg, Claudia Trenkwalder MD, Kassel, Germany and Birgit Högl MD, Innsbruck, Austria

**Disease-specific Quality of Life (QoL) in Restless Legs Syndrome (RLS): Validation of the QoL-RLS Questionnaire**

Ralf Kohnen¹, on behalf of the German RLS Expert Group*

¹ReSearch Pharmaceutical Services, Forth Washington, PA, USA & Psychology Department, University Erlangen-Nuremberg, Nuremberg, Germany

**Introduction:** Restless Legs Syndrome (Willis-Ekbom-Disease, RLS) is a sensori-motor disorder which affects approximately 10% of the population in Western countries. Patients are suffering from an urge to move with or without dysesthesias which is engendered by rest, relieved by movement and follows a circadian rhythm (worse in the evening and at night). RLS is associated with functional impairments of activities of daily living and quality of life. New data on the validity and reliability of the disease specific QoL-RLS questionnaire for measuring quality of life (QoL) are presented.

**Methods:** The QoL-RLS self-rating questionnaire includes 12 items, using 6 categories from 0 = “no impairment” to 5 = “very severe impairment (total score: 0-60). One item requests global QoL. Further items are related to (a) consequences of the RLS symptoms on sleep, activities of daily living, mood and social interactions, (b) the consequences of disturbed sleep on everyday life, (c) consequences of pain and side effects as well as to (d) the evaluation of coping behavior.

**Results:** Evaluation was performed with the pooled data of four clinical trials on efficacy of different dopamine agonists (893 subjects). Construct validity: A factor analysis resulted in a two factors solution (“impairment by symptoms” and “burden of symptoms”) explaining 56.67 % of the variance. Reliability: The QoL-RLS total score showed a high internal consistency (α= 0.89). Convergent validity: Mean baseline correlations to the IRLS total score (0.67) and its sub-scale “Symptom Impact” (.73) were high. Change from baseline to end of treatment correlated with change in the IRLS total score at 0.75. Discriminant validity: The QoL-RLS total score was different (p< .0001) for three severity classes according to the Clinical Global Impression severity scale at baseline (moderate, severe, very severe).

**Conclusion:** The QoL-RLS questionnaire is a valid and reliable instrument which reflects RLS patients’ subjective impairment of well-being and activities of daily living due to RLS symptoms and concomitant features. Because of high internal consistency and sensitivity for change over time the QoL-RLS total score is qualified to be used as an patient report outcome endpoint in clinical trials.

* German RLS Expert Group: Heike Beneš MD, Schwerin, Wolfgang H Oertel MD, Marburg, Karin Stiasny-Kolster MD, Marburg, Claudia Trenkwalder MD, Kassel, Germany and Birgit Högl MD, Innsbruck, Austria

**The Effect of Pharmokinetic Parameters on Euphoria, Drug Liking Following Different Oral Hydromorphone Formulations in Opioid-experienced, Non-dependent, Recreational Drug Users**

Howard Hassman¹, Steven Glass¹, Lawrence Brownstein¹, Lynn R Webster²

¹CRI Worldwide, Mount Laurel, New Jersey United States²Lifetree Clinical Research and Pain Clinic, Salt Lake City, Utah, United States

**Introduction:** One factor of opioid attractiveness with recreational abuse is the rate of drug absorption and corollary pharmacodynamic experience. In this study, the relationship between abuse potential (determined by “likability” and “euphoria”) and the pharmokinetic profile maximum plasma concentration [Cmax] and time to maximum plasma concentration [Tmax]) of differing formulations of hydromorphone and placebo was explored in recreational drug users.
**Method:** This post-hoc analysis of a double-blind, placebo-controlled, randomized, 2-phase, crossover study of subjects with histories of recreational opioid use who received single oral doses of placebo and of 2 formulations of hydromorphone: immediate-release (IR) hydromorphone 8 mg (Dilaudid®) and once-daily hydromorphone extended-release (ER) (EXALGO®) at doses of 16, 32, and 64 mg intact and 8 mg milled to disrupt the extended-release properties of EXALGO tablets.

**Results:** When adjusted for dose, Cmax for all intact once-daily hydromorphone ER treatments, yielded significantly lower Cmax compared with 8-mg IR hydromorphone (P<0.001). Tmax was delayed and AQ (Cmax/Tmax) was lower with all intact once-daily hydromorphone ER doses compared with 8 mg IR hydromorphone. In the 28 subjects completing all treatments and included in the final analyses, the intact once-daily hydromorphone ER formulations had significantly lower liking and euphoria effects for the first 4 hours (P<0.05) compared with the 8 mg IR and 8 mg milled hydromorphone ER tablets.

**Conclusions:** If liking and euphoria effects drive opioid use, then once-daily hydromorphone ER intact may be less attractive than IR hydromorphone or hydromorphone ER milled. Delaying the peak in plasma concentrations may be associated with reduced abuse liability.

---

**Design Features of a Real-World Effectiveness Study of Long-Acting Injectable Risperidone in Frequently Relapsing Bipolar Disorder**

Bobo WV, Epstein RA, Patton TD, Bossaller NA, Shelton RC

*Department of Psychiatry; Vanderbilt University School of Medicine; Nashville, TN (USA)*

**Introduction:** Real-world effectiveness studies of patients with bipolar disorder have to deal with real-world complexity, but retain methodologic rigor. Significant design challenges are thus faced when conducting longitudinal real-world effectiveness studies of bipolar pharmacotherapy.

**Methods:** We describe design features of a preliminary 52-week open, randomized trial that assessed the effect of long-acting injectable risperidone (RLAI) as an adjunct to treatment-as-usual (TAU), as compared with TAU alone, for preventing relapse events in frequently relapsing, community-treated patients with bipolar I or II disorder. Specific challenges included: recruiting a representative sample with greatest possibility of reaching primary endpoint; minimizing dropout; avoiding contamination of routine clinical practices by study procedures, yet ensuring rigorous data collection; adopting simple, clinically significant endpoints; and ensuring that clinical outcomes could be longitudinally assessed even after the first relapse event.

**Results:** Components of the study design relevant to achieving these goals included: (1) Minimum entry/exclusion criteria to enroll target population and maintain safety; (2) no exclusion based on comorbidity or prior adherence; (3) enrollment and follow-up occurred at a large, urban community mental health clinic (CMHC); study clinicians were CMHC employees; (4) clinical assessments (other than tolerability and clinical global state) conducted every 2 weeks by full-time on-site research staff; (5) person-years (p-y) used to track follow-up; thus, patients eligible for multiple relapse events; and (6) relapse definition incorporated DSM-IV criteria for an acute bipolar mood episode (LIFE), symptom ratings (YMRS, MADRS), clinical impression (CGI-S), hospitalization, necessary medication adjustments, dropout (due to ineffectivity), and use of urgent care/crisis services available at the CMHC. There were no significant between-group differences in the number or duration of relapse events, the primary endpoint, or in the number of acute depressive or manic episodes, per p-y of follow-up; however, urgent care referrals (p < 0.04) and necessary medication adjustments (p < 0.01) were significantly less frequent in the RLAI+TAU group than those who received TAU alone.

**Conclusions:** We demonstrated the feasibility of conducting a randomized study that incorporated design features of naturalistic trials in a highly-relapsing CHMC-treated population of patients with bipolar I or II disorder. Further validation of our approach is needed.

---

**Salsalate for the Treatment of Pre-diabetes in People with Schizophrenia**

William R. Keller, Robert P. McMahon, Robert W. Buchanan

*Maryland Psychiatric Research Center, University of Maryland School of Medicine, Baltimore, Maryland USA*

**Introduction:** Obesity is a major risk factor for developing type II diabetes mellitus (DMII) and is an underlying risk factor of insulin resistance for individuals with schizophrenia. Obesity, insulin resistance, and DMII share certain underlying features in common, especially increased levels of pro-inflammatory cytokines and other inflammatory measures, including C-reactive protein (CRP), interleukin-6 (IL-6) and tumor necrosis factor-α (TNF-α). Salicylates,
including aspirin, are effective in the treatment of diabetes, with a number of studies demonstrating these medications increase insulin sensitivity and decrease glucose levels. Several studies examined salsalate in obese, non-diabetic otherwise healthy individuals and found salsalate decreased fasting glucose levels and increased insulin sensitivity. This was the first study evaluating the use of salsalate in individuals with schizophrenia.

Methods: This was a 6-week open-label study of salsalate for individuals with a DSM-IV-TR diagnosis of schizophrenia or schizoaffective disorder at risk for diabetes. All individuals were required to have a body mass index (BMI) \( \geq 27 \text{ kg/m}^2 \), and to not meet criteria for type I or type II diabetes mellitus. The study included 10 outpatients with stable psychiatric symptoms who received salsalate 2 grams orally twice a day. The primary outcome measures were metabolic parameters (Hemoglobin A1c, fasting glucose, insulin C peptide, lipid profiles), cytokines (IL-1\( \beta \), IL-1\( \alpha \), IL-2, IL-6, TNF-alpha, IL-10), C-reactive protein, and adiponectin. Secondary outcome measures were psychiatric symptoms (Brief Psychiatric Rating Scale, Scale for the Assessment of Negative Symptoms) and cognitive functioning (Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)) measured before and after treatment.

Results: Triglycerides (185 mg/dL vs 119 mg/dL, \( p=0.03 \)) and hemoglobin A1C (5.7% vs 5.5%, \( p=0.04 \)) were significantly lower post-treatment. Insulin C peptide trended towards significance after treatment (2.79 ng/ml vs 2.04 ng/ml, \( p=0.1 \)). Fasting glucose was not significantly different before and after treatment. Adiponectin, IL-1\( \beta \), IL-2, IL-6, and C-reactive protein were lower after treatment and had effect sizes between 0.33 and 0.48, but the differences before and after treatment were not significant.

Conclusions: Salsalate was effective in lowering triglycerides and hemoglobin A1C for individuals with schizophrenia at risk for diabetes. Cytokines were generally lower after treatment, but there were not enough participants to detect a difference in cytokines. This study is limited by the open-label design, small sample size, and lack of comparison with a placebo group. Participants consisted of stable outpatients. This may explain why there was no observed difference in psychiatric symptoms or cognitive functioning.

Exploratory Moderators of the Treatment Effect and Implications in the Design of Subsequent Trials

Lori L. Davis, MD\(^1\), Andrew C. Leon, PhD\(^2\)

\(^1\)VA Medical Center, Tuscaloosa, AL, \(^2\)Weill Cornell Medical College, New York City, NY

Background: In a randomized controlled trial, unemployed veterans with PTSD (N=85) were prospectively randomized 1:1 to Supported Employment (SE) or VA conventional Vocational Rehabilitation Program (VRP) for one year. Compared to VRP, those participants who received SE were 2.7 times more likely to gain competitive employment (the a priori primary outcome). During the 12-month study, 76% of the IPS participants gained competitive employment, compared with 28% of the VRP participants. Veterans assigned to IPS also worked substantially more weeks than those assigned to VRP (42% versus 16% of the eligible weeks, respectively) and earned more during the 12-month period.

Methods: This poster presentation will focus on the exploratory analyses of moderators of the treatment effect and how the results may be used to design future studies. A moderator is a baseline characteristic that is positively or negatively associated with the treatment versus control effect size. In exploratory moderator analyses we examined the potential barriers to employment, including inadequate transportation, inadequate housing, family care burden, and inadequate financial means.

Results: We found that there was a greater supportive employment benefit in those with inadequate transportation, inadequate housing, and for those without a family care burden. These results will be presented in detail. We will examine sample size requirements for design subsequent trials that apply these results. The N required to detect moderator by treatment interaction is approximately 4x that of a study recruiting an enriched sample (i.e. recruiting subjects with target level of the moderator).

Conclusion: Exploratory analyses of moderators of the treatment effect in an RCT are not intended guide clinical decision making, but rather are an excellent strategy used to design subsequent trials. These methodological approaches apply to pharmaceutical interventions as well as vocational rehabilitation interventions.