ISCTM 2013 Autumn Conference

30 September – 2 October 2013

The Hyatt, Penn's Landing, Philadelphia, PA

Poster Abstracts

Groupings:
Alzheimer's Disease: 1-8
Major Depressive Disorder: 9-15
Translational: 16-21
Psychoses: Schizophrenia and Bi-Polar-General: 22-29
Psychoses: Occupational and Social Cognition Outcome: 30-37
Working Group Reports: 38

1 Identification of Rater Errors on ADAS-Cog Training Exercises and when Conducting In-Study Assessments

Perez M, Bertzos KA, Brady CA, Marsh JL

inVentiv Health Clinical

Methodological Question Being Addressed: How can raters be trained and monitored on the ADAS-Cog to minimize rater error and maximize the ability to detect true drug effect?

Introduction: A concern in Alzheimer's disease (AD) clinical trials is that rating errors on the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) may undermine clinical trial results. Although most clinical trials employ experienced raters who have completed scale training, research has shown that raters still make errors when conducting study assessments. Identification of key problem areas for raters may inform development of training programs and in-study review programs that may ultimately minimize rater error, thus maximizing the ability to detect true drug effect. This study examined rater performance on an ADAS-Cog training exercise as well as in-study performance rating subjects.

Methods: Raters from 4 AD clinical trials were selected to participate in an ADAS-Cog training program based on self-reported educational status and scale experience. Training included a didactic session followed by an evaluation exercise requiring raters to score a video-recorded ADAS-Cog administration. Rater scores were then compared to pre-set criterion scores, and raters performing outside of the acceptable criterion range were remediated prior to testing study subjects. Once successfully trained, raters were allowed to conduct in-study ADAS-Cog assessments, which were submitted for central review by assessment scale specialists and evaluated for errors in scoring, administration, documentation, and for aberrant rating patterns. Identified errors/aberrant rating patterns were classified into pre-specified category types and raters were remediated for errors.

Results: Across the 4 studies, the ADAS-Cog training program was completed by 333 raters, of which, 95% percent made at least one error rating the video-recorded ADAS-Cog. Errors were identified in all ADAS-Cog subtests, with more frequent errors made on Spoken Language Ability, Orientation, Comprehension, Remembering Test Instructions and Word Finding Difficulty. Across the 4 studies, within a 15-month period, 217 of the trained raters administered 1,929 in-study ADAS-Cogs that were centrally reviewed for errors. A total of 873 errors/aberrant rating patterns were identified, with 80% of the raters making at least one error. On average, raters made less than 1 error per in-study ADAS-Cog assessment. As with the training exercise, errors were identified in all ADAS-Cog subtests, with more frequent errors made on Naming, Orientation and Word Recognition. Following central review and remediation, there was a 12% decrease in overall rater-errors/aberrant rating patterns over the 15-month period studied.

Conclusions: Although raters were experienced administering the ADAS-Cog, 95% made at least 1 error on the ADAS-Cog training exercise and 80% made at least 1 error rating in-study assessments. Since most AD trials are designed to detect a 1.5 to 4 point drug-placebo difference over 6-18 months, this findings reaffirms the concern that rater errors may be masking true trial outcomes. Raters are likely to benefit from more substantial training and rigorous centralized review programs to maximize the possibility of measuring true drug outcome.

Disclosures: All authors are employees of inVentiv Health Clinical and rater training activities are sponsored by pharmaceutical clients.

2 Opioid Modulation: a Novel Mechanism for the Treatment of Depression: Results of the ALKS 5461 Phase 2 Study

Marshall RD[1], Leigh-Pemberton RA[1], Memisoglu A[1], Martin W[1], DeSomer M[1], Ehrich E[1], Fava M[2]

[1]Alkermes, Inc., [2]Massachusetts General Hospital

Methodological Question Being Addressed: The sequential parallel comparison design (SPCD) has been proposed as a strategy to reduce both the overall risk in CNS drug trials and the sample size requirements. This SPCD study was a 2-stage randomized trial that re-randomized placebo nonresponders from stage 1 to either staying on placebo or going on active treatment during stage 2, and combined the estimates from both phases, including all subjects from stage 1 and only the placebo non-responders in stage 2.

Introduction: The endogenous opioid system is thought to play a key role in the regulation of mood. The contemporary use of opioids for depression is limited by abuse potential, presumably a result of mu opioid agonism. ALKS 5461 is a co-formulation of buprenorphine (BUP), a partial mu agonist, combined with ALKS 33, a counter-acting mu antagonist designed to yield a nonaddictive opioid modulator. The ALKS 33 component was optimized to be highly potent and sublingually bioavailable with the latter two properties being essential for sublingual coformulation. A phase 2 study was conducted to confirm preliminary evidence of anti-depressive efficacy observed in an earlier phase 1/2 pilot study of ALKS 5461.

Methods: 142 patients with major depressive disorder (MDD) and inadequate response to SSRI or SNRI therapy in the current depressive episode were enrolled. A blinded sequential parallel comparison design (SPCD) with two 4-week treatment stages was utilized, and the estimates from all subjects in stage 1 were pooled with the estimates from the placebo non-responders re-randomized to treatment with active therapy or placebo in stage 2. Treatment groups included 2 mg/2 mg BUP/ALKS 33, 8 mg/8 mg BUP/ALKS 33, and matching placebo. All subjects, including those assigned to placebo treatment, remained on background SSRI/SNRI therapy.

Results: Results from the study showed that ALKS 5461 significantly reduced depressive symptoms across a range of standard measures, including the study's primary outcome measure, the 17-Item Hamilton Depression Rating Scale ([HAM-D17]; p=0.026), as well as secondary measures including the Montgomery-Asberg Depression Rating Scale ([MADRS]; p=0.004) and the Clinical Global Impression – Severity Scale (p=0.035). The most common AEs observed in the study included nausea, vomiting and sedation, typical of opioid therapy.

As predicted, drug-placebo differences on response rates, HAMD-17 scores and MADRS scores were greater for both doses in stage 2 vs. stage 1.

Conclusions: ALKS 5461 demonstrated significant and clinically meaningful improvement in depressive symptoms among patients with an inadequate response to SSRI/SNRI therapy. These significant findings in a relatively small sample for a POC study testing two separate doses against placebo were driven by comparatively greater drug-placebo differences in stage 2, supportive of the efficiency of the SPCD design and its ability to de-risk CNS trials.
Combined with findings from a prior pilot study, these results indicate that opioid modulation may be a novel and important new treatment approach for this serious and chronic disease.

Disclosures: This study was sponsored by Alkermes, Inc. One or more authors report potential conflicts which are described in the program.
Fava M: Patent for Sequential Parallel Comparison Design (SPCD), which are licensed by MGH to RCT Logic, LLC; and patent application for a combination of azapirones and bupropion in Major Depressive Disorder (MDD).

References: Bodkin JA, et al. J Clin Psychopharmacol. 1995;15:49-57. Fava M, et al. Psychotherapy and Psychosomatics 2003;72:115-127; and "Erratum" 2004; 73: 123.

3 Histamine H3 Antagonists are not Effective in Alzheimer's Disease: Exploring the Translational Disconnect between Rodents and Humans with QSP

Geerts H[1,2], Roberts P[1,3], Spiros A[1]

[1]In Silico Biosciences, Lexington MA, USA, [2]University of Pennsylvania, PA, USA, [3]Oregon Health and Science University, Portland, OR, USA

Methodological Question Being Addressed: Many clinical trials in CNS fail despite preclinical studies suggesting otherwise, but there is no systematic study on the underlying reasons for this outcome, although a lot can be learned from negative studies. Possible issues include inadequate target engagement, harmful off-target effect, interference of

comedication or target selection mismatch. Virtual patient trials run on a Quantitative Systems Pharmacology platform allows to explore different hypotheses about the translational disconnect and to learn about the underlying neurobiology.

Background: Histamine H3 antagonists have shown great promise in preclinical models of Alzheimer's disease (AD) and in scopolamine induced deficits in healthy volunteers. However, different H3 antagonists did not produce substantial improvement in clinical trials of AD. Learning from trial failure is important for a better understanding of the relevant human neuropathology.

Methods: We developed a complex computer-based cortical network model with implementation of the physiology of 12 different membrane CNS targets based on a biophysically realistic multi-compartment model of 80 pyramidal cells and 40 interneurons simulating the stability of a memory trace. The model was humanized by calibrating with multiple clinical trials for working memory tasks in healthy humans and schizophrenia patients and 28 different drug-dose-time combinations in Alzheimer patients. H3 antagonism was introduced using presynaptic increases in dopamine (DA), acetylcholine (ACh) and norepinephrine (NE), the size of which was determined using clinical scopolamine data in human volunteers. The clinical trials with H3 antagonism in Alzheimer's disease were simulated for different treatment durations, APOE genotypes and comedication with acetylcholinesterase inhibitors (AChE-I).

Results: Assuming an 80% target engagement in the human scopolamine trial, relative values for maximal cortical release of ACh, DA and NE were 25%, 15% and 10% respectively. Further simulations in the AD case showed an acute dose-dependent improvement of H3 antagonism at 4 weeks, however the effect significantly disappeared at longer durations, casting doubt on the value of this clinical development project. The model identified the dynamics of ACh release and clearance in the cholinergic cleft as the major difference between H3 antagonism and AChE inhibition. Progressing AD pathology changes the balance between excitation and inhibition as relatively more pyramidal cells become dysfunctional. Chronic AChE-I treatment drives the □□□□ nAChR more into desensitization and these receptors regulating GABA tone in cortical networks improve the balance between excitation and inhibition. In contrast by only increasing the level Ach and DA release, H3 antagonists are ineffective at progressively more severe pathology.

Discussion: This case-study illustrates the capability of a quantitative systems pharmacology approach to identify the pharmacodynamic difference of H3 antagonists in AD and scopolamine-induced deficits in healthy volunteers and preclinical studies on the other hand. A key role is attributed to the GABA tone in the cortical network, suggesting that the observed downregulation of □□□□ nAChR is a compensatory mechanism of the AD brain to reduce the GABA inhibitory tone that becomes dominant at later stages of the pathology when more excitatory pyramidal cells become affected.

Disclosure: All authors were employed by In Silico Biosciences

4 High Variability and Lack of Change on the ADAS-Cog: Placebo Analyses of the CODR Database

Popp D[1], Garzio LM[1], Boehm P[1], Randolph C[1,2]

[1]MedAvante, Inc., [2]Department of Neurology, Loyola University Medical Center

Methodological Question Being Addressed: How much variability is seen in ADAS-Cog placebo scores in the CODR database? How well do inclusion MMSE scores predict changes in ADAS-Cog score over time? What is the average visit-to-visit placebo change on the ADAS-Cog across 20 trials in CODR database?

Introduction (Aims): The Alzheimer's Disease Assessment Scale-Cognitive Section (ADAS-Cog)1 is the most commonly used primary neurocognitive endpoint in clinical trials of mild-moderate Alzheimer's disease. However, concerns have been raised regarding the lack of sensitivity of this instrument in mild disease and with respect to the magnitude of error variance2. In an effort to drive advances in Alzheimer's disease therapeutic development, the Critical Path (C-Path) initiative created the C-Path Online Data Repository (CODR) that contains data from the placebo arms of over 20 clinical trials. We analyzed the CODR database to explore change in the ADAS-Cog as a function of time and baseline MMSE score, to further inform clinical trial design for this population.

Methods: There are a total of 24 studies in the CODR database. We eliminated trials of less than 6 months duration (n = 2), open label extension studies (n = 2) and trials that did not include the ADAS-Cog as an endpoint (n = 6). There were 14 studies remaining, with a total of 3,939 subjects across 26,123 visits. We analyzed overall ADAS-Cog visit-to-visit change scores for visits <= 90 days apart to quantify variability. Change scores over time were analyzed as baseline to 6-month change, baseline to 12-month change, and baseline to 18-month change. Change scores over time were also analyzed as a function of baseline MMSE score.

Results: Mean ADAS-Cog scores at baseline ranged from 9.3 (± 5.2 SD) to 27.8 (± 9.6 SD). The visit-to-visit test-retest correlation for the ADAS-Cog was .91. Average visit-to-visit change on the ADAS-Cog was .5 (± 5.2) ranging from -35 to 40; 40% of subjects improved during the 90-day visit window. Average change from baseline was 1.0 (± 5.7), 2.8 (± 7.1), and 4.0 (± 8.1) for 6-, 12- and 18 months, respectively; 42%, 34% and 32% of subjects improved, respectively. MMSE was a significant predictor of change in ADAS-Cog score from baseline to 6-, 12-, and 18 months (β = -.18, -.22, and -.27, all p's < .05) with higher baseline MMSE scores associated with less decline over time.

Conclusions: Large variability and extreme outliers in visit-to-visit ADAS-Cog change scores, including those of a biologically implausible magnitude, were fairly common, suggesting significant error variance. Overall change on the ADAS-Cog over time was somewhat less than expected based upon published data. This may therefore reflect a publication bias whereby failed trials are less likely to be reported in the peer-reviewed literature. Change on the ADAS-Cog was strongly dependent upon baseline MMSE, and visit-to-visit variability on the ADAS-Cog was fairly high. These findings suggest that accurate subject selection is critical for obtaining placebo decline on the ADAS-Cog, and that in-study quality control methodologies should be further explored for efficacy in reducing error variance.

Disclosures: Dr. Popp and Mrs. Garzio are full-time employees of MedAvante, Inc. Dr. Randolph and Mr. Boehm are consultants to MedAvante, Inc.

1Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. Am J Psychiatry. 1984;141:1356-1364.
2Schafer K, DeSanti S, Schneider LS. Errors in ADAS-cog administration and scoring may undermine clinical trial results. Curr Alzheimer Res. 2011; 8(4): 373-376.

5 Electronic Administration of a Clinical Trials Version of the MMSE (eMMSE-CT)

Samuelson P, DeBonis D, Miller DS

Bracket, Wayne, PA

Methodological Question Being Addressed: Equivalence of electronic administration of a clinical trials version of the MMSE (eMMSE-CT) compared with paper MMSE.

Introduction: The MMSE is a brief assessment that assesses Orientation, Registration, Attention and Calculation, Recall, Naming, Repetition, Comprehension, Reading, Writing and Drawing. The assessment is widely used in clinical practice as well as clinical trials. Many clinicians have adopted their "own" way of administering this relatively simple test. In clinical trials, paper MMSE administration by trained raters is subject to procedural inconsistencies, including scoring, administration and transcription errors. Independent clinical reviews of MMSE source documents have been part of data quality control in trials. Our published research has shown that one-third of MMSEs completed at inclusion visits in Alzheimer's trials contain procedural errors.

An electronic version of the scale specifically targeted at clinical trials could eliminate common scoring and administrative errors on the MMSE, improving data quality and rater reliability.

Methods: We sought to create an electronic version of the MMSE for clinical trials (eMMSE-CT) by first obtaining a software license and agreement with PAR, the copyright holder of the scale. Bracket, With PAR's agreement, we sought to develop an electronic clinical trials version that went beyond an electronic CRF, one that integrated instructions to raters, training reminders, enforced scoring rules, and had the flexibility to incorporate alternative questions and cultural adaptations of questions. The eMMSE-CT requires raters to ask each question required and record a response, automatically calculating scores and enforcing administration conventions.

The eMMSE-CT underwent extensive internal user acceptance and regulatory compliance testing.

To support internal UAT, Bracket we conducted a brief test/retest comparison of the eMMSE-CT with traditional paper MMSE administration. Five patients ranged from age 57 to 79 and were interviewed in an outpatient clinic. Two patients had been diagnosed with MCI, another with post-concussion after MVI, and the other two had no complaints of memory loss.

Both versions of the MMSE were administered by clinical experts that have conducted rater training on the MMSE in dozens of trials throughout the world. The different versions were administered approximately 30 minutes apart in counterbalanced order. Raters and patients were asked to provide feedback about their experiences with the two administration methods.

Results: The total scores on the paired administrations were the same when the eMMSE-CT was administered first. The scores were slightly higher when the paper MMSE was administered first. The ICC between the paired eMMSE and

paper MMSE was 88, with no significant differences using a paired t-test.

Upon review of the paper administrations, an addition error by the rater was detected. This error was not repeated on the paired eMMSE.

Patients found the computer administration experience comparable to the paper administration.

Conclusions: The eMMSE-CT was equivalent to the traditional, paper-based MMSE in this small sample. In one instance, the eMMSE-CT did not repeat an error in the paired paper MMSE. The eMMSE-CT standardizes the administration of the scale and could be a useful tool to enhance data quality across sites and raters in large clinical trials.

Disclosures: Priscilla Samuelson, Dan DeBonis and Dr. David Miller are employees of Bracket

6 Modeling Short Term Data to Predict Long Term Results

Zhang RY[1], Nguyen H[1], Chuang-Stein C[2]

[1]Statistics, Primary Care, Pfizer Inc, New York, New York, USA, [2]Statistics Research Consulting Center, Pfizer Inc, Kalamazoo, Michigan, USA

Methodological Question Being Addressed: Can data from completed short-term clinical trials in a particular indication be used to predict the "treatment effect" for future long-term studies in the same indication? This question is relevant to the situations where a go/no go decision to start a required long-term phase III study is being considered but the available phase II data is limited to short-term treatment exposure.

Introduction: The increasing prevalence of Alzheimer's Disease (AD) and lack of effective agents to attenuate progression have accelerated research and development of disease modifying (DM) therapies. Almost all the trials investigating disease modification by necessity have long duration and require large sample sizes. Recent failures and terminations of several late-stage AD development programs have motivated us to investigate quantitative models that could be used to predict the outcome of future long-term trials from completed trials of shorter duration. Such models, if successful, can be used to build criteria and confidence for advancing a drug candidate from the exploratory stage to the confirmatory stage of drug development.

Methods: Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-cog) data from two completed, 18-month AD trials were used for this investigation. Two models, a liner model and a quadratic model, were proposed. MMRM (mixed-model for repeated measures) methodology was used to fit the observations up to 12 months first, then predicted values at 18 months were derived and compared with the observed values at the same time point. The error margin ε (i.e. the difference between the predicted and the observed) and associated variability σ (standard deviation) were calculated and examined to assess the reliability of the models.

Results: The error margin ε estimated from the linear model ranges from -0.197 to -0.455 and the standard deviation σ ranges from 5.55 to 7.06, which translates to 95% CI upper bound (underestimation) of 1.45-1.68 when the sample size is 175. The error margin ε for the quadratic model ranges from -0.739 to 1.33, standard deviation ranges from 5.72 to 7.34, and the corresponding upper bound of 95% CI is 1.44-3.02.

Conclusion: In AD trials, a linear model seems to offer a reasonable prediction of long-term treatment effect on ADAS-cog based on short-term data. The robustness of the model should be investigated further using data from other completed long-term studies. The approach described here could be used to develop models for other outcome measures and disease areas.

Disclosure: The methodology described in this manuscript resulted from discussions among all participating authors and was not funded by any corporate sponsors.

7 What are the Optimal Cognitive Outcomes for Trials in Preclinical Alzheimer's Disease?

Wesnes KA [1,2,3]

[1]Practice Leader, Bracket, Gatehampton Road, Goring-on-Thames, RG8 0EN, UK, [2]Visiting Professor, Department of Psychology, Faculty of Health and Life Sciences, Northumbria University, Newcastle, UK, [3]Adjunct Professor, Centre for Human Psychopharmacology, Swinburne University, Melbourne, Australia

Methodological Question Being Addressed: What are the essential properties of cognitive tests which make them effective diagnostic and outcome measures for trials in preclinical Alzheimer's disease.

Background: A 2011 joint National Institute on Aging - Alzheimer's Association workgroup published operational research criteria for preclinical Alzheimer's disease (AD). Preclinical AD patients are cognitively normal, and thus

diagnosis cannot be made by comparing individuals to norms on one-off testing; instead it will be necessary to employ repeated testing to detect greater than normal rates of decline over years. The 2011 workgroup commented that sensitive test measures needed to be developed in multiple cognitive domains to reveal evidence of early synaptic dysfunction in neural networks vulnerable to AD pathology. In March this year, FDA officers authored a paper in the New England Journal of Medicine acknowledging that as functional impairment would be difficult to assess in such patients; it could be feasible for the Agency to approve a treatment for preclinical AD via the accelerated approval pathway solely on the basis of cognitive test data. Long term clinical trials in preclinical AD are underway, or about to start (e.g. API, DIAN, A4), including the five-year worldwide TOMMORROW study in 5,800 cognitively normal individuals aged 65-83. This paper will consider the necessary properties of cognitive tests to enable them to serve as effective outcome measures in this field.

Methods: The workgroup recommendation that cognitive tests needed to be developed ignored the progress already made by numerous groups who have long been using such procedures to detect change in cognitive function in unimpaired individuals; both deterioration and improvement. While virtually any of the thousands of available mental ability tests can stratify individuals when administered once; measuring change over time imposes far more stringent demands on test design. In this poster, examples of tests which are and are not suitable for such applications will be presented.

Results: The majority of the procedures used to measure cognitive function in clinical trials have been adopted from the field of neuropsychology. An important limitation of such tests is that most were not designed for repeated use; consequently the risk of training effects was not considered in test design. However, such training effects can mask subtle cognition declines over time and/or prevent treatment induced enhancements from being identified. This will be illustrated from a recent analysis of the cognitively normal sample in the ADNI study, which has so far been tested annually up to 6 years. Here a range of traditional neuropsychology tests showed training effects in volunteers aged 60 to 90 years; which lasted in several cases up to 6 years. Further none of the tests identified reliable cognitive decline in the population over the study, despite the detection of yearly declines both in cortical thickness and hippocampal volume, as well as over 10% developing Mild Cognitive Impairment. In contrast, data will be presented from other tests including automated procedures developed specifically for clinical trials which have been able to detect evidence of year on year cognitive impairment in healthy but elderly individuals. Further, data will be presented from SCOPE, a 5-year intervention study in elderly but cognitively intact hypertensive subjects, showing that automated tests could identify a treatment benefit by slowing decline over the study; in sharp contrast to the neuropsychological tests employed. Other examples will be presented. Finally, the critical requirements of test design will be identified, including: pre-study training; the availability of multiple parallel forms; the avoidance of ceiling and floor effects; and the precise assessment of speed together with accuracy.

Conclusion: Over the last 25 years considerable evidence has accumulated that many widely used scales and neuropsychological tests have very limited utility for repeated testing in normal individuals. This however has not deterred the field from widely employing such procedures (e.g. in the TOMMORROW study). In contrast automated cognitive tests have been available since 1868; and computerised systems exist which have already proven sensitive outcome measures in trials similar to those being conducted in preclinical AD.

Disclosure: The author is an employee and stockholder in Bracket, which provides services to the clinical trial industry. One of the services offered is the CDR System, a set of computerised cognitive tests, which was used to gather some of the data presented in this poster.

8 A 12 Week Proof of Concept Efficacy and Safety Study of ORM-12741 In Patients with Moderate Alzheimer's Disease

Rouru J[1], Wesnes K[2], Hänninen J[1], Murphy M[3], Riordan H[3], Rinne JO[4,5]

[1]Orion Pharma, Turku, Finland, [2]Bracket Global, Goring-on-Thames, UK, [3]World Wide Clinical Trials, King of Prussia, PA, USA, [4]Turku PET Centre, [5]Turku University Central Hospital, Turku, Finland

Methodological Question Being Addressed: The initial evaluation of the cognitive and behavioral benefits of new chemical entities (NCE) in patients with Alzheimer's disease (AD) in an optimal and efficient manner remains challenging. The current study was designed to evaluate an NCE in a reasonably sized safety and efficacy proof of concept setting that optimized selection of study endpoints and patient population.

Introduction: ORM-12741 is a highly potent and selective alpha-2C adrenoceptor (AR) antagonist that has demonstrated efficacy in rodent models suggesting beneficial effects on both cognition and behavioral symptoms in AD, as well as

good tolerability across seven Phase I studies. This is the first Phase IIa study of a selective alpha-2C AR antagonist in AD patients. The primary objectives of the study were to evaluate safety, tolerability and efficacy of ORM-12741 as add-on therapy in patients with AD.

Methods: This was a randomized, double-blind, placebo-controlled study of 100 moderate AD patients (MMSE scores 12- 21) with behavioral symptoms [Neuropsychiatric Inventory (NPI) score of ≥15]. Patients were allocated to two flexible dose levels of either 30 to 60 mg or 100 to 200 mg of ORM-12741 or matching placebo twice daily for 12 weeks as add-on to their stable cholinesterase inhibitor therapy. Use of stable memantine and SSRI medication was allowed. Efficacy was assessed primarily with computerized tests from the CDR System, from which standard composite scores were derived including: Quality of Episodic Memory, Quality of Working Memory, Quality of Memory, Speed of Memory and Power of Attention. NPI was assessed to quantify the effects on behavioral and psychological symptoms.

Results: Statistically significant treatment effects were noted for ORM-12741 on the composite CDR System measures Quality of Episodic Memory (p=0.03) and Quality of Memory (p=0.013) as well as on NPI Caregiver Distress (p=0.034). These favored ORM-12741 over the 12 week treatment period with no significant differences between dose groups. In addition, a positive trend was noted for both Quality of Working Memory and NPI total score primarily for the 30 to 60 mg group. No significant differences were identified on the other scores. ORM-12741 was generally well tolerated in the study.

Conclusion: Significant positive effects of ORM-12741 on composite measures of memory in moderate AD patients as add-on therapy over 12 weeks were observed. These findings are encouraging and warrant further exploration in longer term trials.

Disclosures: The study was sponsored by Orion Pharma. Juha Rouru and Jutta Hänninen are employees of Orion Pharma. Keith Wesnes is an employee of Bracket which supplied the CDR System for the study. Mike Murphy and Henry Riordan are employees of Worldwide Clinical Trials, the CRO that supported trial conduct.

9 Prefrontal GABA Abnormalities Are Associated With Reduced Hippocampal Volume In Major Depressive Disorder

Abdallah CG[1,2], Jackowski A[3], Sato JR[3,4], Mao X[5], Kang G[5], Cheema R[1,2], Coplan JD[6], Mathew SM[7,8], Shungu DC[5]

[1] Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA, [2] Clinical Neuroscience Division, National Center for PTSD, Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA, [3] LiNC, Departamento de Psiquiatria, Universidade Federal de São Paulo, SP, Brazil, [4] Center of Mathematics, Computation and Cognition. Universidade Federal do ABC, Santo Andre, Brazil, [5] Departments of Radiology, Psychiatry and Biophysics, Weill Medical College of Cornell University, New York, NY, USA, [6]Division of Neuropsychopharmacology, Department of Psychiatry, SUNY Downstate Medical Center, Brooklyn, NY, USA, [7] Michael E. Debakey VA Medical Center, Houston, TX, [8] Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA.

Methodological Question Being Addressed: Does decreased Hippocampal volume in Major Depressive Disorder reflect impaired amino-acid (specifically GABA) neurotransmission?

Background: Hippocampal volume reduction has been reported in major depressive disorder (MDD) and is believed to reflect impaired amino-acid neurotransmission leading to increased extracellular glutamate and excitotoxicity. In line with this preclinical-based hypothesis, small hippocampal volumes in human have been repeatedly associated with poor treatment outcome in depressed patients treated with monoaminergic antidepressants. To better understand the role of amino-acid neurotransmission in hippocampal deficits and subsequent resistance to monoaminergic drugs, the current study investigated the relationship between hippocampal volumes and abnormally low GABA levels in the anterior cingulate cortex (ACC) of MDD subjects, a neurochemical indicator previously associated with treatment-resistant depression.

Methods: Twenty-six medication-free MDD (mean age ±SEM, 39.0 ±2.0) and 26 healthy (mean age ±SEM, 37.4 ±2.6) subjects were studied. Participants underwent high-resolution magnetic resonance imaging (MRI) to estimate hippocampal volume using the fully automated FreeSurfer processing. Proton MR spectroscopy (1H MRS) was acquired to measure GABA levels in a single voxel placed in the ACC region. Using the median split cutoff point of the ACC GABA level, MDD patients were divided into two groups: Low GABA MDD (n = 13) and High GABA MDD (n = 13).

Results: Conducting general linear model (GLM) with repeated measures to examine right and left hippocampal volumetric differences among the three groups (healthy, Low GABA MDD, and High GABA MDD), while controlling for intracranial volume (ICV) and handedness, showed a significant group effect ($F(2,45) = 9.0$, $n = 52$, $p < 0.001$). Post hoc analysis with Bonferroni correction revealed a reduction in hippocampal volume in Low GABA MDD group

compared to High GABA MDD group ($p < 0.001$) and Healthy controls ($p = 0.01$). No hemispheric effect ($p = 0.50$) or hemispheric-by-group interactions ($p = 0.80$) were present. Age, gender, weight, education, IQ, age of onset, duration of illness, treatment resistance and psychotropics-naïve status, PSQI, PSWQ, HAM-A, HDRS24, and ACC voxel tissue composition were individually considered as covariates in the model. However, they had no significant ($p > 0.05$) effect on the model and they did not impact the significance of the group effect. In the total MDD group, ACC GABA was positively correlated with standardized hippocampal volume (hippocampal/intracranial x 10000) [$rs = 0.42$, $n = 26$, $p = 0.03$]. There was no correlation between these two measures in the healthy controls group ($p = 0.24$).

Conclusion: Consistent with preclinical suggestive evidence, we found a robust association between abnormally low ACC GABA and small hippocampus in MDD individuals; independent of the treatment resistance status or other clinical and demographic characteristics. The data support the presence of a subpopulation of MDD patients with more profound amino-acid neurotransmission abnormalities leading to neuronal remodeling deficits and resistance to Aminergic-based antidepressants.

Disclosures: This work was supported by National Institutes of Health Grants R01-MH07895 (DCS), K23-MH-069656, and MO1-RR-00071 (SJM), and in part by an Investigator-initiated research grant from the CFIDS Association of America, Inc (DCS) and by Weill Cornell Medical College New Faculty Development Funds (DCS). Preparation of this report was supported in part by NIDA T32-DA022975 (CGA) and the Clinical Neurosciences Division of the National Center for Posttraumatic Stress Disorder (CGA). Funding sources had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.
SJM: received research funding or salary support over the last three years from the Banner Family Fund, Brain and Behavior Fund (NARSAD), The Brown Foundation, Inc., Bristol-Myers Squibb, Department of Veterans Affairs, Evotec, Johnson & Johnson, and the National Institute of Mental Health. He has received consulting fees or honoraria from Allergan, AstraZeneca, Cephalon, Corcept, Noven, Roche, and Takeda. He has received medication (Rilutek) from Sanofi-Aventis for a NIMH sponsored study. Dr. Mathew has been named as an inventor on a use-patent of ketamine for the treatment of depression. Dr. Mathew has relinquished his claim to any royalties and will not benefit financially if ketamine were approved for this use. JDC received grant support from NIMH, NYSTEM, GlaxoSmithKline, Pfizer, and Alexza Pharmaceuticals. He is on the Pfizer advisory board and gives talks for BMS, AstraZeneca, GSK, and Pfizer. No biomedical financial interests or potential conflicts of interest are reported for CGA, AJ, JRS, XM, and DCS.

10 An International Study of the GRID-HAMD: Has it Fulfilled its Promise?

Williams JBW[1,2], Ondrus M[1], Kitzinger M[1], Persson J[1], Popescu M[1], Valjakka R[1]

[1]MedAvante, Inc., [2]Dept. of Psychiatry, Columbia University College of Physicians and Surgeons

Methodological Question Being Addressed: The new GRID-HAMD has been the major outcome measure for several large clinical trials. What is its acceptability to users, and how does this compare to the SIGH-D, a widely used version of the HAMD?

Introduction (Aims): The Hamilton Depression Rating Scale (HAMD) has been the target of many critiques (Bagby et al, 2004). An international group representing academia, clinical practice, the pharmaceutical industry, and government developed the GRID-HAMD in an attempt to address these critiques and improve administration of the HAMD (Williams, 1988; Williams et al, 2008). The GRID-HAMD provides a novel grid scoring structure that separates frequency and intensity to allow clinicians to rate these as independent axes. The newly-formulated instrument also provides a structured interview guide and scoring conventions on the same page as each item. Finally, the GRID-HAMD presents revised anchor points for items that were problematic or inconsistently rated.

Methods: A survey was distributed to 74 central raters that included 20 statements about the GRID-HAMD that were rated on a 7 point scale from strongly disagree (1) to strongly agree (7). Questions covered usability and ease of use as well as the new page layout, the revised item wordings, and the grid format. Finally, the questionnaire listed four statements asking raters to compare the GRID-HAMD with the SIGH-D, with a response from 1=GRID-HAMD to 7=SIGH-D.

Results: Fifty-seven questionnaires were completed (77%). Mean age of respondents was 40. All were mental health professionals; 37 (65%) were psychiatrists or doctorate-level psychologists. Half (51%) of the respondents live in Europe, 42% in the US, and the rest in Russia (5%) and South Africa (2%). All reported at least three years' experience assessing depression, and 81% reported more than seven years.

Most raters agreed that the wording of the questions in the GRID-HAMD made it easy to administer (77%), the conventions were clear (82%) and helpful (86%), and the guidelines for rating symptom intensity were clear (79%).

Fewer rated it "easy to decide on a frequency level" (61%). A large majority (89%) of raters thought that "having the scoring conventions integrated into the interview guide has made scoring easier." 75% agreed that "assessing symptom intensity and frequency separately makes it easier to score the items."

More than half (54%) of the 44 raters who had used both the GRID-HAMD and the SIGH-D preferred the graphical layout of the GRID to that of the SIGH-D. However, a slightly higher percentage preferred the SIGH-D for its "ease of use" (50% vs. 45%) and "efficiency" (39% vs. 36%). Finally, slightly more raters expressed "overall preference for the SIGH-D" (45% vs. 43%).

Conclusions: These central raters rated the clarity and ease of use of the GRID-HAMD positively. Surprisingly, however, they did not indicate an overall preference for the GRID-HAMD over the SIGH-D. Several areas of improvement were indicated for both scales, and the "most difficult" items were highlighted. In ongoing data analyses, demographic and other predictors of these views are being explored and will be presented.

References: Bagby RM, Ryder AG, Schuller DR, Marshall MB: The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? Am J Psychiatry 2005; 161:2163-2177
Williams JBW: A structured interview guide for the Hamilton Depression Rating Scale. Archives of General Psychiatry 1988; 45:742-747
Williams JBW, Kobak KA, Bech P, Engelhardt N, Evans K, Lipsitz J, Olin J, Pearson J, Kalali A. The GRID-HAMD: Standardization of the Hamilton Depression Rating Scale. International Clinical Psychopharmacology 2008; 23:120-129

11 Sensitivity of Self-report vs. Investigator Rated Depression Scales in Randomized Placebo Controlled Studies of Antidepressants in Major Depressive Disorder (MDD)

Rabinowitz J[1], Werbeloff N[1], Menard F[2], Mandel F[3], Caers I[4], Kapur S[5]

[1]Bar Ilan University, Ramat Gan, Israel, [2]Lundbeck SAS, Paris, France, [3]Pfizer, New York, [4] Janssen Research & Development, Beerse, Belgium, [5]Institute of Psychiatry, Kings College, London

Methodological Question Being Addressed: Sensitivity to treatment effects of self-report vs. investigator rated depression scales in randomized placebo controlled studies of antidepressants.

Introduction (Aims): Despite their apparent relevance, self-report measures are frequently not included in clinical trials of antidepressant drugs, and when they are included they are not used as primary endpoints. We compared treatment effects based on investigator rating scales and patient self-report ratings using data from 8 randomized placebo controlled trials.

Methods: Eight of the 39 studies in the NewMeds repository of placebo controlled RCT's of antidepressants used both the HAMD, an investigator administered rating scale, and a patient self-report measure (depression factor from Profile of Mood States, POMS; depression factor from SCL-56 Symptom Checklist). Change on active treatment vs. placebo was examined on both the self-report and investigator rating scales using ANCOVA controlling for baseline. Change scores per item were also calculated. Finally, we divided patients into two groups according to baseline severity of depression (median split by baseline HAMD score; 23), and examined investigator ratings and patient self-report ratings in these two groups.

Results: Overall, the active-placebo difference observed using the POMS was greater than that of HAMD (difference in ES=0.12), while SCL-56 showed slightly less active-placebo difference than HAMD (difference in ES=-0.02). In 4 of the 5 studies where both measures favored active treatment, self-report measures showed larger active treatment vs. placebo difference. In 2 studies placebo showed greater improvement than active treatment on both assessments, but the differences were smaller on the self-report measure. In one study the investigator rating scale favored active treatment over placebo whereas the self-report scale showed a minimal difference in favor of placebo. When comparing similar items on the HAMD and self-report scales, ES of POMS items were larger than ES of HAMD in 2 of 3 comparisons (depressed mood, feelings of guilt); ES of SCL items were larger than ES of HAMD items in 3 of 9 comparisons (feelings of guilt , psychomotor retardation, psychological anxiety). Self-report ratings using either POMS or SCL-56 showed greater active-placebo difference change from baseline than HAMD ratings for severe patients (difference in ES POMS=0.21; difference in ES SCL=0.05) but not mild-moderate patients (difference in ES POMS=-0.08; difference in ES SCL=-0.05).

Conclusions: The self-report measures, as used in these 8 studies, appear to be more sensitive to treatment effects than investigator rating scales among severe patients in antidepressant trials in MDD. Certain items show more change using self-report ratings than investigator ratings. Self-report measures appear to capture different aspects of the disease than

investigator administered rating scales, and should thus be included in trials.

Disclosure: This research was supported by the Innovative Medicine Initiative Joint Undertaking under grant agreement n° 115008 of which resources are composed of EFPIA in-kind contribution and financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013). The authors report no conflicts of interest for this work.

12 Patient Input toward the Final Development of the Rosenberg-Hassman Mood Scale (RHMS), a Patient Reported Outcome (PRO) for Major Depressive Episodes

George SV[1], Rosenberg LI[1], Hassman H[2], Wesnes K[3], Krefetz D[2]

[1]Center for Emotional Fitness, Cherry Hill, NJ, [2]CRI Lifetree, Mount Laurel, NJ [3]Bracket, Goring-on-Thames, United Kingdom

Methodological Question Being Addressed: Patient preferences are explored for a new depression PRO based solely on symptom frequency

Background: The Rosenberg-Hassman Mood Scale (RHMS) is being developed according to the FDA Guidance for Patient Reported Outcomes (PRO) to measure depression levels for patients in Major Depressive Episodes. Prior patient input has helped to generate new items among the 228 word/phrase synonyms in the 26 question RHMS and helped to determine the preferred response options for our scale. We are now seeking to obtain more patient input from our 2013.2 beta version of the computerized RHMS in preparation for finalizing the RHMS.

Methods: After signing informed consent, 50 patients diagnosed with either Bipolar Disorder most recent episode depressed or Major Depressive Disorder took our computerized RHMS scale one to three times over a 6 week interval. Subsequently we obtained feedback from these patients individually using a prepared questionnaire and in a group setting using an open-ended agenda.

Results: As per the FDA Guidance on PROs, feedback was both open-ended and in response to a questionnaire. The data is currently being accumulated and analyzed to prepare the final version of the RHMS.

Conclusion: Patient input toward the final development of the RHMS according to the FDA Guidance for Patient Reported Outcomes has been obtained and is being analyzed. A final version of the RHMS, a self-rated depression scale, will be presented.

13 The Full Range and Total Number of Suicidal Behaviors (All Types) Predict Future Suicidal Behavior

Posner, K[1], Greist, JH[2], Federico M[3], Goldstein EL[1], Pearlstein SL[1], Amira L1, Buchanan J[1], Yershova KV[1]

[1]Columbia University, College of Physicians and Surgeons/New York State Psychiatric Institute New York, NY, [2]Healthcare Technology Systems, Madison, WI, [3]ERT, Philadelphia, PA

Methodological Question Being Addressed: Are different types of lifetime suicidal behavior and ideation predictive of subsequent suicidal behavior, and does the number of behaviors increase risk?

Introduction (Aims): Suicide is a major public health crisis, and prevention hinges upon identification of high-risk individuals. A past history of a suicide attempt is one of the strongest predictors of subsequent suicidal behavior. However, the full range of suicidal behaviors has been ignored and not studied. In order to evaluate how individual types of ideation and behavior contribute to suicide risk and to obviate the issue of low suicidal behavior base rate in single studies, records from electronic Columbia-Suicide Severity Rating Scale (eC-SSRS) assessments of patients across multiple medication trials were pooled.

Methods: A total of 3,776 patients completed a baseline and at least one follow-up assessment across studies of major depression, insomnia, posttraumatic stress disorder, epilepsy, and fibromyalgia. Mean follow-up period was 64 days. Five types of suicidal ideation and 4 types of behavior were measured on the C-SSRS. Risk for subsequent suicidal behavior was calculated for each of the ideation types relative to no lifetime ideation.

The same method was used for the types of suicidal behavior relative to no lifetime behavior of the same type.

Results: Participants who reported any lifetime actual attempt, interrupted attempt, aborted attempt, or preparatory behavior were more likely to prospectively report suicidal behavior (any of the 4 types) compared to participants with no lifetime behaviors (OR: 4.56 (3.40 – 6.11 95%CI), 5.28 (3.88 – 7.18 95%CI), 4.75 (3.53 – 6.40 95%CI), and 4.92 (3.38 – 7.16 95%CI).Comparatively, the risk was similar across behavior types. Additionally, the total number of different types of lifetime suicidal behaviors monotonically increased the likelihood of future suicidal behavior (OR: 3.14 (2.22 – 5.23

95%CI), 6.86 (4.57 – 10.32 95%CI), 8.33 (5.50 – 12.62 95%CI), and 9.35 (4.98 – 17.54 95%CI). Participants who reported any type of lifetime suicidal ideation at baseline were more likely to report suicidal behavior at follow up than those who reported no lifetime ideation. For 'Wish to be dead', 'Active ideation, nonspecific' and 'Active ideation with method, but no intent or plan' odds ratios were 5.12 (2.90 - 9.27 95%CI), 7.90 (4.14 – 15.05 95%CI), and 10.20 (5.68 – 18.30 95%CI), respectively. The odds increased substantially for participants with more severe types of ideation – 'Active ideation with a method and intent, but no plan' (OR = 21.10; 11.96 – 37.24 95%CI) and 'Active ideation with a method, intent, and plan' (OR = 22.65; 12.55 – 40.86 95%CI).

Conclusions: Assessed prospectively, every type of lifetime suicidal behavior and total number of different behaviors, as well as every type of lifetime suicidal ideation predicted subsequent suicidal behavior. Investigators and clinicians do not routinely ask about important suicidal behaviors other than actual suicide attempts. Yet, the eC-SSRS data demonstrated that actual attempts constituted a relatively small proportion of total suicidal behaviors in this large sample; and each type and the total number of behavior types were independently equally predictive of future suicidal behavior. Thus, identification of the full range of behaviors, total number/types of suicidal thoughts are of paramount importance to suicide prevention.

Disclosures: Dr. Kelly Posner, along with several colleagues, developed the eC-SSRS tool that was used in this study. Dr. Posner is the director of the Center for Suicide Risk Assessment which, as part of an effort to help execute the FDA suicidality classification mandates, has received support through her employer, the Research Foundation for Mental Hygiene, from the following pharmaceutical companies: Abbott, Aerial Biopharma, Albany Molecular Research, Alder Biopharma, Alfresa, Alkermes, Amgen, Astellas Pharm, Astra Zeneca, Biogen, Biomarin Pharmaceutical, Biovail Technologies, Boehringer Ingelheim, Bracket, Bristol Myers Squibb, Cato Research, Celerion, Cephalon, Cetero Research, Chiesi Pharmaceuticals, Covance, CRI Worldwide, Daiichi Sankyo Company, Depomed, Douglas Pharmaceuticals/VersaPharm, EISAI, Elan, EnVivo, Epiomed, Forest, Gilead, GlaxoSmithKline, Grunenthal, GW Pharma Limited, Human Genome Sciences, i3 International, i3 Research, i3 Pharmaceutical Services, ICON Development Solutions, Impax Laboratories, INC Research, Ingenix, IntelGenx Corp, IntraCellular Therapies, Ironwood, IRIS, Isis, Ivax, Janssen, Jazz, Johnson & Johnson, Lilly USA, Lotus, Lundbeck, MedAvante, MedImmune, Merck, Mochida, Neurocrine Biosciences, Neuronex, Neurosearch, NextWave Pharma, Novartis, Noven, NovoNordisk, Omeros, Orexigen Therapeutics, Orion, Otsuka, Pamlab, Parexel, Pfizer, PGx Health, Pharmaceutical Research Associates, Pharmanet i3, Pierrel Research, PPD, Prana Biotechnology, ProPhase, Psyadon, QED Pharmaceuticals, Quintiles, Receptos, Reckitt Benckiser, Rho, Rhythm, Roche, Sanofi-Aventis, Schering-Plough, Schwarz Biosciences, SCOPE International, Sepracor, Shionogi, Shire, Siena Biotech, SK Life Science, Sunovion, Supernus Pharmaceuticals, Synosia Therapeutics, Takeda Global Research & Development Center, Takeda Pharmaceuticals, TauRx Therapeutics, Theravance, UCB Biosciences, UCB Korea, UCB Pharma, United BioSource Corp, Upsher-Smith Laboratories, Vaccinex, Valeant Pharmaceuticals, Vernalis, Vivus, WorldWide Clinical Trials, Wyeth Ayerst, Wyeth Pharmaceuticals, Wyeth Research, Xenoport and Zalicus. Dr. Posner receives royalty payments from the e-CSSRS, which are distributed to her by her employer, the Research Foundation for Mental Hygiene.

14 Comparing Clinician-rated CSSRS with Computer-administered eC-SSRS (Ver 2.0): Replication and Extension of Prior Research Findings

Mundt JC[1], Greist JH[2], Jefferson JW[2], Federico M[3], Paty J[3], Gwaltney C[3]

[1]Center for Telepsychology, Madison, WI, [2]Healthcare Technology Systems, Madison, WI, [3]ERT, Philadelphia, PA

Methodological Questions Being Addressed: Are patient-reported, fully-structured, electronic assessments of lifetime and recent suicidal ideation and behavior (eC-SSRS 2.0) comparable to semi-structured clinician-based assessments (C-SSRS)?

Introduction/Aims: A previously developed version of the eC-SSRS [1] was modified to include evaluation of the recency of reported lifetime suicidal ideation and behaviors (SIB). The repeatability of prior findings, and the reliability and validity of the additional assessment of recent SIB, were studied. Study aims included demonstration of technical feasibility, discriminant sensitivity to known patient differences, convergent scale validity, and methodological equivalence.

Methods: Twenty subjects, 10 hospital employee controls and 10 psychiatric inpatients, were recruited from Rogers Memorial Hospital (Milwaukee, WI). Subjects were administered MINI diagnostic interviews. Exclusion criteria included dementia, delirium, psychosis or deafness. Control subjects with lifetime MDD or dysthymia, or current disorder diagnoses other than specific phobias or nicotine/caffeine dependence were also excluded. SIB was assessed at baseline and one week later by two experienced C-SSRS clinical raters. Subjects also completed the eC-SSRS and the

paper-pencil Beck Suicide Scale (BSS).

Results: No between-group differences were found with respect to age, gender, marital status, education, or follow-up interval. Kendall's tau-b was used to compare agreement between assessments of suicidal ideation at lifetime, recent (6 months), and at follow-up. (Table 1). Paired t-tests of ideational intensity scores did not find differences between either rater or the eC-SSRS.

Table 1. Kendall's tau-b

	Suicidal Ideation Agreement		
	Most Severe Lifetime	**Most Severe Recent (6 months)**	**Most Severe Since Baseline**
C-SSRS Rater1 – Rater2	.92	.87	.97
eC-SSRS – Rater1	.91	.93	.96
eC-SSRS – Rater 2	.88	.90	.97
R1, R2, eC-SSRS (w/ BSS)	.75, .77, .72	.84, .87, .89	.64, .64, .66

Inter-rater and eC-SSRS agreement regarding the presence/absence of suicide attempts, preparatory behaviors, and non-suicidal self-injurious behaviors was evaluated using kappa statistics. Of 39 kappas computed, 15 indicated Excellent agreement (1 - .8), 7 were Very Good (.79 - .6), 6 were Good (.59 - .40), 2 were Questionable (.39 - .2) and 9 were Unacceptable (< .2). All of the unacceptable kappas were from follow-up assessment comparisons; three were comparisons between the human raters, two were comparisons between Rater 1 and eC-SSRS, and four were comparisons of Rater 2 and the eC-SSRS.

Conclusions: This study replicates and extends results of the previous eC-SSRS validation study, showing agreement between the patient-rated eC-SSRS and both clinical raters that is comparable to the agreement between the raters. Discrimination between healthy controls and psychiatric patients demonstrated implementation feasibility, as well as construct and predictive validity. Psychometric evaluation of suicide ideation severity was very good to excellent with strong convergent validity with the BSS. Assessment of lifetime and recent suicidal behaviors showed good to excellent agreement between the patient-rated eC-SSRS and clinician-rated C-SSRS, however kappas were reduced in the follow-up assessments. Subject feedback from study debriefings found the methods perceived similarly, with the majority having no preference for either assessment method. When preferences were expressed, most preferred the clinician interactions for the ability to clarify responses and for the therapeutic relationship.

Disclosures: This study was supported by ERT, a clinical services provider to the biopharmaceutical industry, including the eC-SSRS. Drs. Mundt, Greist, and Jefferson are stock shareholders in Healthcare Technology Systems, which receives royalty payments for the development of the eC-SSRS. Mr. Federico and Drs. Paty and Gwaltney are employees of ERT.

References: Mundt JC, Greist JH, Gelenberg AJ, Katzelnick DJ, Jefferson JW, Modell JG. Feasibility and validation of a computer-automated Columbia-Suicide severity rating scale using interactive voice response technology. *J Psychiatr Res.* May 27 2010.

15 Impact of Overzealous Patients and Raters on Drug-placebo Separation in an MDD Trial: Analysis of Tandem Ratings

DeBonis, D, Cummings, R, Sachs, GS

Bracket Global, LLC

Methodological Question Being Addressed: Impact of overzealous patients and raters on drug-placebo separation in an MDD Trial: Analysis of Tandem Ratings

Introduction: Baseline inflation is a recognized problem for signal detection in MDD studies. While this problem is often attributed to overzealous raters, little data is available to evaluate baseline inflation by patients.
Levomilnacipran ER (LVM; 1*S*, 2*R*-milnacipran) is a potent and selective serotonin and norepinephrine reuptake inhibitor approved in July 2013 for MDD. NCT00969709, an 8-week randomized, placebo-controlled, double-blind study of adult MDD, utilized an endpoint reliability program in which computer administered ratings ($MADRS_{COMP}$) were obtained shortly after site-based psychiatrist ratings on the primary outcome measure ($MADRS_{SBR}$). Metrics based on the tandem ratings were used to explore/compare impact of operationally defined overzealousness on the part of patients and/or site-based raters.

Methods: The inclusion criteria for the study required a $MADRS_{COMP} \geq 26$ and a $MADRS_{SBR} \geq 30$ at screening and

baseline visits. We analyzed the MADRS change from baseline to endpoint using MMRM , exploring factors that were associated with drug-placebo separation in prior published analyses of tandem (paired computer and human) ratings:

- Concordant baseline $MADRS_{COMP}$ and $MADRS_{SBR}$, defined as being within 5 points of each other (representing one standard deviation of difference between the two scores).
- Discordant baseline MADRS scores, with a difference >5 between $MADRS_{COMP}$ and $MADRS_{SBR}$
    - Rater Baseline Inflation, defined as baseline $MADRS_{SBR}$ > $MADRS_{COMP}$ by >5 points.
    - Overzealous patients, defined as baseline $MADRS_{COMP}$ > $MADRS_{SBR}$ by >5 points.

Results: Significant improvement was seen in the LVM group (n=215) versus PBO (n=214) at the end of treatment on $MADRS_{COMP}$ (diff=-2.75, P=0.0142) and $MADRS_{SBR}$ (diff=-3.10, P=.0051).

Table 1: Concordant Baseline MADRS scores (difference $\leq$ 5)

	LVM (n =171) LSM (SE)	PBO (n =171) LSM (SE)	LSMD:LVM-Placebo (95% CI)
$MADRS_{SBR}$	-16.46 (0.88)	-11.24 (0.86)	-5.22 (-7.61, -2.83), p<0.0001
$MADRS_{COMP}$	-15.49 (0.90)	-10.39 (0.88)	-5.09 (-7.52, -2.67), p<0.0001

Table 2: Discordant Baseline MADRS scores (difference > 5)

	LVM (n =44) LSM (SE)	PBO (n =43) LSM (SE)	LSMD:LVM-Placebo (95% CI)
$MADRS_{SBR}$	-10.69 (1.69)	-15.93 (1.74)	5.23 (0.48, 9.99), p=0.0311
$MADRS_{COMP}$	-10.04 (1.72)	-16.38 (1.76)	6.34 (1.52, 11.16), p=0.0100

Table 3: Rater Baseline Inflation

	LVM (n =19) Mean	PBO (n =23) Mean	Mean Difference: LVM-Placebo
$MADRS_{SBR}$	-11.67	-16.31	4.65
$MADRS_{COMP}$	-9.13	-10.94	1.80

Table 4: "Overzealous patients"

	LVM (n =25) Mean	PBO (n =20) Mean	Mean Difference: LVM-Placebo
$MADRS_{SBR}$	-10.58	-14.82	4.24
$MADRS_{COMP}$	-13.89	-21.76	7.87

Conclusions: Both $MADRS_{SBR}$ and $MADRS_{COMP}$ achieved statistically significant drug-placebo separation in the study. The subset of subjects with baseline tandem rating concordance demonstrated significant drug-placebo separation on both the $MADRS_{SBR}$ and $MADRS_{COMP}$, while discordant baseline tandem ratings (22.5% of the population) significantly favored placebo on both outcomes. The discordant group was divided evenly between the rater and subject inflated scores. "Overzealous patients" demonstrated a disproportionate contribution to placebo response on $MADRS_{COMP}$. These results suggest that incorporation of tandem ratings into the inclusion criteria for studies of mood disorders might reduce the sample size required to detect an effect.

Disclosures: Dan DeBonis, Rachel Cummings and Gary Sachs are employees of Bracket. Forest Laboratories, the sponsor of the study, has reviewed and approved this abstract.

16 Psychomimetic Effects of Ethanol, Morphine, THC and Ketamine are Related to Functional Brain Connectivity

Kleinloog, D[1], Rombouts, S[2], van Gerven, JMA[1,2]

[1]Centre for Human Drug Research, Leiden, The Netherlands, [2]Leiden University Medical Centre, Leiden, The Netherlands

Methodological Question Being Addressed: Can resting state fMRI be used in a standardized way to detect drug-induced psychomimetic effects?

Introduction: Slow variations in fMRI signal within the brain are organized in spatial networks that are present in rest and under external stimulation and are consistent between subjects. These resting state networks (RSN) can provide valuable information about the effects of drugs acting on the central nervous system.

Methods: Data from previous studies in healthy male volunteers with pharmacological challenges using ethanol, morphine, THC and ketamine (12 subjects per drug), with repeated measures of resting state fMRI and subjective effects, were used. Imaging data were pre-processed and a set of ten 'template' resting state networks [1] was used to extract individual connectivity maps by dual regression. These connectivity maps were related to three clusters of subjective effects ('perception', 'relaxation' and 'dysphoria') as measured by visual analogue scales for each individual subject using linear regression. Group-analysis was performed on the individual slopes. The corrected significance level after threshold-free cluster enhancement was set at 0.005. All analyses were performed using FSL 5.0.4.

Results: The subjective effects of 'perception' correlated significantly across drugs with one cluster within the 'sensorimotor network' of 57 voxels on the border of the caudal anterior and posterior cingulate cortex (maximum at Talairach coordinates 6, -28, 46).

Conclusions: The analysis used a standardized approach with standard 'templates' for each resting state network. This approach improves the reproducibility of the analysis and is recommended for analysis of resting state fMRI data in clinical trials. A cluster was identified near the border of the anterior and posterior cingulate cortex that shows a consistent correlation between resting state activity and subjective effects of perception, across four different psychoactive drugs. This area was previously found to be related to positive symptoms in schizophrenia [2]. This suggests an underlying function of the area in psychomimetic effects. Resting state fMRI may be a useful tool to demonstrate pharmacological effects of newly developed CNS drugs.

Reference: [1] Smith et al., 2009, PNAS; 106(31): 13040-5. [2] Choi et al., 2005, Psychiatry Res; 139(3): 239-47.

17 Using the Cross-species 5C-CPT to Enhance the Vertical Translation of Drug Development from Rodents to the Clinic

Young JW[1],2, Light GA[1,2], Eyler LT[1,2], Perry W[1], Serences J[3], Zhou[1],X[2], Geyer MA[1,2]

[1]Department of Psychiatry, University of California San Diego, La Jolla, CA, [2]Research Service, VA San Diego Healthcare System, San Diego, CA, [3]Department of Psychology, University of California San Diego, La Jolla, CA

Methodological Question Being Addressed: Can attention/vigilance be readily examined across species from mice to humans with evidence for construct, predictive, and face validity? Questions for such a task: a) are the neural substrates underlying performance conserved across species; b) are task challenges exerting similar effects across species; c) is there task sensitivity to pharmacological perturbation and improvement; and d) is there sensitivity to deficits in clinical populations?

Introduction: Impaired attention/vigilance is present in numerous psychiatric disorders. Although pro-attention treatments are needed, the lack of FDA-approved medications for many disorders limits the generation of isomorphic treatments. Drug development programs need preclinical paradigms that can assess the targeted cognitive domain consistently across species. Thus, when model animals of each disorder are generated, the efficacy of novel treatments can be tested with the reasonable expectation that any positive effects observed in the model animal will be conserved in the relevant patients. We have assessed whether or not the 5-choice continuous performance test (5C-CPT) of attention/vigilance fulfills the criteria detailed above.

Methods: Over the past 5 years, we have generated evidence for the cross-species reliability of the 5C-CPT. While first developed for mice, we have subsequently adapted the task for rats and now humans. Human testing has examined: a) the neural substrates underlying performance using fMRI and EEG recording of subjects performing the task; b) the effects of task challenges using variations consistent with tests in rodents; c) whether improved performance can be seen in smokers vs. non-smokers as in other human attentional tasks; and d) whether clinical populations including schizophrenia and bipolar disorder exhibit impaired performance. Rodent testing has examined: a) lesion-induced impairment of performance, targeted at locations that were identified using fMRI studies; b) the effects of task challenges consistent with tests in humans; c) the effects of acute and chronic nicotine, as well as more selective nicotinic agonists and numerous other treatments; and d) model animals related to schizophrenia and bipolar disorder.

Results: Human studies indicate that 5C-CPT performance involves numerous neuroanatomical regions, including the

parietal cortex. Parietal cortical lesions of mice negatively impacted performance. Extending the time between trials (reduced event rate), inserting irrelevant flashing visual stimuli (distracters), and 36 hours of sleep deprivation deleteriously affected the same aspects of performance in both humans and mice. Human smokers and mice receiving nicotine via minipumps exhibited better 5C-CPT performance than non-smokers or mice receiving saline. Sp4 hypomorphic mice exhibit impaired 5C-CPT performance consistent with that of people with schizophrenia, while dopamine transporter knockdown mice exhibit impaired 5C-CPT deficits in a pattern consistent with people with bipolar disorder.

Conclusions: The 5C-CPT is a viable cross-species test for attention/vigilance. There is increasing evidence that the human and rodent 5C-CPTs exhibit construct, predictive, and face validities for one another. These data support the premise that the 5C-CPT is a tool that can enhance the vertical translation of a novel treatment from rodents to the clinic

Disclosures: The authors report no conflict of interest for this work.

18 Retrospective Review of Safety and Tolerability of Continuous Cerebrospinal Fluid (CSF) Collections During Phase 1 Pharmacokinetic/Pharmacodynamic Studies in Healthy Volunteers and Patients

English BA*[1,2], Gevorkyan H[3], Han D[3], Gertsik L[3], Avakian M[1], Pauly J[1], Jhee SS[1], Ereshefsky L[1,4]

[1]PAREXEL International, Glendale, CA, [2]Dept. of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, [3]California Clinical Trials Medical Group, Glendale, CA, [4]University of Texas College of Pharmacy, Austin, TX

Methodological Question Being Addressed: Evaluation of safety and tolerability during continuous CSF sampling procedures in healthy volunteers and patients.

Introduction: Sampling of cerebrospinal fluid (CSF) has become an increasingly common procedure as part of a Phase 1 (Ph1) clinical development program in order to characterize pharmacokinetic (PK) and pharmacodynamic (PD) relationships related to dosing and CNS penetration. Continuous CSF sampling via an indwelling catheter has been used to measure drug concentration and changes in PD biomarker targets in healthy volunteers (HNVs), healthy elderly (HNEs) and in patients (ie. Alzheimer's disease). The risks involved in conducting continuous CSF studies have been reported to be similar to those of lumbar puncture which include postural headache, backache, infection and rarely nerve root damage, epidural or subdural bleeding.1-4 In this study, we evaluated the adverse event (AE) profile of our continuous CSF procedure in both healthy volunteers and in patients diagnosed with either a psychiatric or neurologic disorder.

Methods: Retrospective chart review of all subjects who underwent continuous CSF collection procedures while participating in a Phase 1 clinical study program. Subjects consisted of young HNVs, elderly HNEs, and patients diagnosed with either Alzheimer's disease, Parkinson's disease, anxiety or major depression. Subjects underwent 1 to 2 continuous CSF sampling periods depending on study design. A continuous CSF sampling study consisted of either serial CSF collections of 6.0 cc of CSF, at a flow rate of 0.5 cc/min, collected over 26-36 hours, or the use of a fractional collection system. Typical CSF volumes are approximately 100 cc/24 hours. Variables such as subject demographics, health status, number of continuous CSF procedures and time interval between continuous CSF collections were compared to AE type and severity.

Results: In our interim data analysis, a total of 800 subjects underwent a continuous CSF procedure across 6 studies sampled. Subjects were predominantly male (65%) with ages ranging from 20-89 years. Catheter failure rate was 6% with 8% of subjects requiring repositioning of the spinal catheter to improve flow rate. The most common AEs were postural headache (45% HNVs vs 54% Patients), neck/back pain (34% HNVs vs 80% Patients) and dizziness (9% HNVs vs 8% Patients). Blood patch rates were 30% in HNVs vs 13.3% in Patients. The higher rate of blood patches in HNVs vs Patients was most likely attributed to differences in technique and changes in catheter size. There were no serious AEs (ie. Infection) associated with the continuous CSF procedure. Our interim results will be updated to include reviewing the AE profiles in about 1100 subjects across 8 separate studies.

Conclusions: Continuous CSF sampling is a highly useful technique employed during Ph1 studies to characterize PK/PD relationships of a CNS active compound. Our interim results show that there are AE differences between healthy volunteers and patients, but that no serious AEs occur as a result of the procedure.

Disclosures: BAE, MA, SSJ and LE are employees of PAREXEL International that receives financial compensation from numerous pharmaceutical companies for the conduction of clinical trials. HG, DH and LG are fulltime employees of the California Clinical Trials Medical Group and serve as Principle Investigators on industry-sponsored trials at

PAREXEL.

19 A PK/PD Study of Xen2174 Administered Intrathecally in Healthy Subjects

Okkerse, P[1], Hay, JL[1], Sitsen, E[2], Dahan, A[2], Klaassen, ES[1], Houghton, W[3], Groeneveld, GJ[1]

[1]Centre for Human Drug Research (CHDR), Leiden, the Netherlands, [2]Leiden University Medical Center (LUMC), Leiden, the Netherlands, 3Xenome Ltd, Indooroopilly, Australia

Methodological Question Being Addressed: Can highly frequent pharmacodynamic testing be performed in conjunction to long term (32 hour) CSF PK sampling, following intrathecal administration of a novel analgesic drug?

Introduction: Xen2174 is a synthetic 13-amino acid peptide that specifically binds to the norepinephrine transporter, which results in inhibition of norepinephrine uptake. Norepinephrine is involved in pain transmission; norepinephrine binding to α2-adrenoreceptors leads to anti-nociception through the activation of inhibitory γ-aminobutyric acid (GABA)-ergic interneurons. Xen2174 showed a strong anti-nociceptive and anti-allodynic effect in non-clinical studies when administered intrathecally. However, at higher dose levels, Xen2174 was epileptogenic. Based on CSF PK studies in dogs and previous clinical studies, intrathecal 2.5 mg Xen2174 was expected to lead to safe CSF concentrations. The current study was performed to assess the CSF pharmacokinetics and the pharmacodynamic profile of Xen2174 using evoked pain tests and EEG in healthy subjects.

Methods: This was a randomized, blinded, placebo-controlled study in healthy subjects. The study was divided in three treatment arms (0.5 mg, 1.0 mg and 2.5 mg Xen2174 intrathecal). Each group consisted of 8 subjects on active treatment and 2 or 3 subjects on placebo. CSF was sampled for up to 32 hours using an intrathecal catheter. Pharmacodynamic assessments were performed using a battery of nociceptive tasks. This battery of pain tasks consisted of pressure pain, electrical pain, cold pressor test and a conditioned pain modulation paradigm. EEG measurements were performed up to 24 hours after dose administration.

Results: In total 25 subjects were administered Xen2174. CSF PK analysis showed a higher AUC of Xen2174 in the highest dose group than allowed by the predefined safety margin based on nonclinical data. Pressure pain measurements showed an increased tolerability in the highest dose group and for the electrical pain (repeated stimulus) an overall treatment effect on the AUC was observed ($p<0.05$). No overall treatment effect could be observed in the other pain tests. The most common adverse event was post lumbar puncture headache, with no increased incidence in the treatment groups. No treatment related abnormalities were observed on the 24 hour EEG evaluations.

Conclusions: In this study, Xen2174 was well tolerated. An increase in pressure pain tolerance was seen in the 2.5 mg group. To better determine the analgesic properties of Xen2174 it would be needed to test higher doses. However, at the highest dose level in this study, CSF concentrations already exceeded the pre-specified exposure limit based on the preclinical safety margin. Pharmacodynamic testing using a pain test battery could be performed in conjunction to continuous CSF sampling.

20 PK and PD of a Nicotinic Anti-cholinergic Challenge with Mecamylamine and Comparison to Scopolamine

Baakman, AC, Rissman, R, Klaassen, ES, van Gerven, JMA, Groeneveld, GJ

Centre for Human Drug Research, Leiden, the Netherlands

Methodological Question Being Adressed: Does administration of the nicotinic acetylcholine receptor antagonist mecamylamine result in a more selective anticholinergic pharmacological challenge than scopolamine and therefore be more suitable for proof of pharmacology studies of nicotinic agonists?

Introduction: Cholinesterase inhibitors are frequently prescribed to patients with Alzheimer's Disease. The increase in acetylcholine level has effect on both muscarinic and nicotinic acetylcholine receptors, causing cognitive enhancement, mostly via the nicotinic receptors, and peripheral side effects, mostly via the muscarinic receptors. Therefore, nicotinic receptor specific agonists are currently being developed. For proof of pharmacology studies of cholinergic compounds, an anti-cholinergic pharmacological challenge with scopolamine is often used. However, scopolamine is an antagonist of the muscarinic acetylcholine receptor, which does not seem appropriate for proof of pharmacology of nicotinic agonists. With this study we aimed to develop an anticholinergic pharmacological challenge with the nicotinic antagonist mecamylamine and to gain time-dependent information on pharmacodynamic effects of mecamylamine, compared to scopolamine and placebo.

Methods: This was a double-blind, double-dummy, placebo-controlled, randomized four-way crossover study with

mecamylamine 10mg or 20mg p.o. or scopolamine 0.5mg i.v. in 12 non-smoking healthy subjects, aged 18-45. In various studies, a maximum of 20mg mecamylamine orally produced few adverse effects, other than mild hypotension. Effective cognitive impairment was observed at a dose of 15mg. For the pharmacological challenge in this study, a lower (10mg) and higher (20mg) dose were chosen in order to determine dose-effect relationship, and because of adverse effect considerations. Scopolamine has been validated and frequently used in challenges with minimal adverse effects and demonstrable cognitive impairments at 0.5mg scopolamine*HBr intravenous solution.

Bloodsamples were taken for pharmacokinetic and neuro-endocrine measurements. Pharmacodynamic measurements consisted of computerised tests for memory, attention, psychomotor speed, eye movements, subjective scales for mood and alertness, stability (body sway) and pharmaco-EEG.

Results: Mecamylamine was well tolerated and had linear pharmacokinetics over the dose range tested. Mecamylamine appeared to more selectively affect memory than alertness compared to scopolamine 0.5mg iv. Mecamlyamine 10mg, mecamylamine 20mg and scopolamine all affected the visual verbal learning test (respectively -2.7 (CI-5.1 - -0.3), -3.6 (CI-5.9 - -1.4), -7.7 (CI-10.1 - -5.4)) and adaptive tracking (respectively -1.89 (CI-3,90 – 0.12), -2.06 (CI-3.97 - -0.16), -10.38 (CI 12.38 - -8.39)), a test for attention. However, mecamylamine did not have clear sedative effects, contrary to scopolamine. This is illustrated by the simple reaction time task (respectively 7.0 (CI-0.8 – 15.5), 3.8 (CI-3.5 - 11.7), 26.8 (CI 17.6 - 36.8)), a subjective visual analogue scale for alertness (respectively -1.3 (CI-3.7 – 1.2), -2.5 (CI-4.8 - -0.2), -5.3 (CI-7.7 - -2.9)) and sacadic peak velocity (respectively -14.3 (CI 33.5 – 4.8), -10.9 (CI-29.0 – 7.1), -25.4 (CI-44.2 - -6.6)), a marker for sedation.

Conclusions: In contrast to previously published studies, we were able to show that scopolamine has sedative effects, that are not present after administration of mecamylamine, while the 20mg dose of mecamylamine
leads to a reproducible pattern of cognitive disturbance that is nicotinic receptor specific. This model may be more suitable for proof of pharmacology and dose finding studies of nicotinic receptor agonists than the frequently used scopolamine model.

21 The Cognitive and Neurophysiological Effects of AZD7325, a Subtype Selective GABAAα2,3--Modulator in Comparison with Lorazepam in Healthy Male Volunteers

Chen X1,2, Jaeger J3*, Lappalainen J4, Maruff P3, Smith MA5*, Cross AJ4, van Gerven J2

*[1]Phase I Unit, Clinical Pharmacology Research Center (CPRC), Peking Union Medical College Hospital, Beijing,China [2]Centre for Human Drug Research ,Leiden, The Netherlands [3]Cogstate, Melbourne Australia and New Haven CT, USA [4]AstraZeneca, R&D, Wilmington Delaware and Cambridge MA [5]Shire Pharmaceuticals, PA, USA *Formerly employed by Astrazeneca (at the time the study was planned and conducted)*

Methodological Question Being Addressed: Does AZD7325, a novel α2,3-subtype selective GABAA modulator produce a benzodiazepine-like profile on cognition and neurophysiologic biomarkers at doses that result in high GABAA receptor occupancy?

Introduction (Aims): Benzodiazepines are nonselective modulators of GABAA receptors that contain α1, α2, α3 or α5 subunits. Animal studies suggest the α1subunit is associated with sedation; α2/α3 receptors are responsible for benzodiazepine's anxiolytic properties,; and the α5 subunit is related to modification of memory and cognition [1,2]. Hence AZD7325 was expected to show reduced effects on cognitive, visuo-motor and postural stability measures relative to lorazepam at doses that result in comparable or higher occupancy. Also we hypothesized that AZD7325 would replicate previous studies[3] in drugs from this new class by showing selective benzodiazepine-like suppression of saccadic peak velocity (SPV).

Method: Two and 10 mg doses of AZ7325 were compared with 2mg of lorazepam and placebo in a double-blind, double-dummy, placebo controlled, randomized, four-way crossover study in sixteen healthy males. Peak plasma concentrations were selected to exceed minimally efficacious concentrations in animal models for anxiety and provide GABAA occupancy levels of 40% - 50% and 80% - 90%, respectively. Cognition was measured using Cogstate tests of speed of processing and attention (Detection, Identification), short term spatial and verbal learning (One Card Learning, International Shopping List ISL), long term verbal memory consolidation and retrieval (ISL delayed recall), executive functioning (Groton Maze) and visuo-motor coordination (Chase Test). ISL was conducted at 1.75 hours post-dose with immediate and delayed (21 hour) recall in order to obtain a measure of long term retention for material presented while on drug. Apart from ISL, cognitive tests were performed three times pre-dose and post-dose at 1.25, 2.25, 3.25, 4.25 and 8 hours post-dose. Measures of neurophysiologic function (saccadic and smooth pursuit eye movements), postural balance (body sway), eye-hand coordination (adaptive tracking) and subjective alertness (VASalertness) were also studied frequently. For each treatment, the selectivity of SPV effects relative to other neurophysiologic markers was

tested by plotting the change from baseline (ΔSPV) against ΔSway and ΔVASalertness.

Results: Sixteen subjects completed the study. Cognitive performance was not different from placebo on either dose of AZD7325, whereas lorazepam was associated with significant impairments on all cognitive measures including ISL delayed recall (all $p<0.05$). No practice effects were detected. The time course of cognitive deficit tracked the observed plasma concentration curve of Lorazepam with maximum effect seen at 2 - 4 hours. Neither dose of AZD7325 showed significant effects on VASalertness, SPV, sway, smooth pursuit or tracking, which were all robustly affected by lorazepam 2 mg. Lorazepam-induced SPV reduction was linearly related to changes in other neurophysiologic biomarkers indicating a general suppression of CNS-activity. In contrast, the slopes of the regression lines were flatter for AZD7325, particularly for the ΔLog(Sway)-ΔSPV relation and the ΔVASalertness-ΔSPV relationship.

Conclusions: Lorazepam 2mg produced significant cognitive decrement on attention, speed of processing, visual and verbal learning, verbal long term retention and retrieval, visuomotor coordination and executive functioning. This effect was maximal near the time of maximal plasma concentration of lorazepam. Similar patterns were observed on the neurophysiologic biomarker measures. No significant cognitive decrement or neurophysiologic effect was observed at either dose of AZD7325. The ΔSPV-relative effect profiles of AZD7325 were consistent with the characteristic ΔSPV-ΔLog(way) and ΔSPV-ΔVASalertness relationship of other α2,3-selective GABA-A agonists. These results suggest that the compound has an anxio-selective profile or the dose selection may have been too low for clinical efficacy. The lack of effects on most CNS-PD parameters also suggests a mitigated side-effect pattern. This finding suggests that GABAAα2,3 subtype selectivity may be associated with lower cognitive and neurophysiological side effects burden than non-selective benzodiazepines.

References: 1.Rowlett, J. K., Platt, D. M., Lelas, S., Atack, J. R., & Dawson, G. R. Different GABAA receptor subtypes mediate the anxiolytic, abuse-related, and motor effects of benzodiazepine-like drugs in primates. Proceedings of the National Academy of Sciences of the United States of America, 2005;102, 915–920. 2.Atack, J. R., Bayley, P. J., Seabrook, G. R., Wafford, K. A., McKernan, R. M., & Dawson, G. R.. L-655,708 enhances cognition in rats but is not proconvulsant at a dose selective for α5-containing GABAA receptors. Neuropharmacology, 2006;51, 1023–1029. 3.Chen Xia, De Haas S, De Kam M, Van Gerven J. An overview of the CNS-pharmacodynamic profiles of non-selective and selective GABA agonists. Adv Pharmacol Sci. 2012:134523

22 Do Rater Certification Procedures Identify Poor Raters?

Daniel DG, Kott A

Bracket Global, LLC

Methodological Question Being Addressed: Does performance rating a videotaped patient interview as part of the rater training and certification process predict performance rating patients at the site?

Introduction: Prior to rating in a CNS clinical trial, site investigators are required to demonstrate proficiency in rating key outcome measures. A common test utilized to evaluate rater competence is successful rating of a videotaped interview either at the investigators' meeting or on-line. The relationship between performance on this test procedure and performance rating patients at the site is poorly defined.

Methods: Prior to study initiation, raters were trained to rate the PANSS at investigators' meetings and/ or on line by highly interactive procedures. In order to be approved to rate in these global schizophrenia trials, they were required to successfully rate a videotaped patient interview by scoring at least 80% of the 30 PANSS items within the acceptable range set by an expert panel. Raters who failed the first attempt were given remediation and retested. If successful on either the first or second attempt they were approved to rate patients at their sites. To assess the quality of ratings at the sites, a proprietary video/audio recording system was utilized to record PANSS rating procedures. External reviewers provided feedback on ratings quality on an ongoing basis to the site.

Results: On their first rating at the site after the investigators' meeting, raters who successfully scored the videotaped test interview on the first attempt (N=189) had a higher rate of matches with consensus scores established with external reviewers (76.3%) than raters who required two attempts (n=43) to successfully score the videotaped test interview (67.3%) ($t=3.39$; $df=230$; $p < 0.05$).

The number of deviations from the expert panel in scoring the recorded interview was modestly but statistically significantly correlated with the number of deviations from the external expert in scoring their first patient at the site ($r=0.24$, $n=232$, $p<.001$).

Discussion: Performance rating a videotaped interview of a CNS clinical trial outcome measure prior to study

participation appears to be modestly, but statistically significantly correlated with subsequent performance rating patients at the site. In addition, raters who fail their first attempt to rate the videotape to pre-determined standards prior to study initiation make more errors rating patients than raters who were successful on their first attempt.

The results suggest that performance on pre-study screening procedures is a useful factor in identifying raters at risk and thus may be a useful component of risk based monitoring procedures.

Future analyses will address contributions of geography, educational level, clinical and scale experience to ratings performance .

Disclosure: David Daniel and Alan Kott are full time employees of Bracket Global, LLC

23 Self Assessing Key Symptoms of Health Trajectories to Improve Treatment Decisions and Outcomes in Bipolar Disorder

Bowden CL, Singh V, Tohen M, Priessmeyer R

UT Health Science Center at San Antonio, University of New Mexico, and Biomedical Development Corporation

Methodological Question Being Addressed: Can a portable tool that patients regularly complete regarding change or stability of key symptoms associated with primary clinical states in bipolar disorder improve patient outcomes and inform clinicians regarding illness/health trajectories?

Introduction: Bipolar disorder (BD) has been studied extensively using methodologies that aggregate data and apply traditional statistical methods to determine the efficacy of treatment interventions. These static endpoints do not account for the dynamics of changes that individual patients experience. We developed software that employs novel nonlinear methodology to track changes in multiple clinical measures in BD as they evolve over time.

Methods: We created KIOSä, referred to herein as K-TABSTM (KIOS - Tracking Changes in Bipolar Symptoms), a novel software platform based on nonlinear systems theory and applied its analytics to data from an NIMH-sponsored clinical study of treatments for BD. This analysis identified several key variables that drive changes in BD patients. A team of psychiatrists recognized as experts on BD treatment further refined those variables into eight key measures and then grouped them into appropriate pairs that collectively define the state of a BD patient.

Results: The panelists selected variables from an initial set that included more than 50 items in commonly used rating scales. Variables had to have exploratory factor analysis loadings ≥ 0.40 on 1 of the 5 symptom domains of bipolar disorder (depression, irritability, mania, anxiety, psychosis) (1). We also required that the mean severity of each variable be $\geq$ the mean severity (i.e., ≥ 1.8) of the variable in patients who were syndromally manic/hypomanic, mixed state or depressed. Although selected independently by the experts, there was high concordance for the 8 symptoms and 7 pairs recommended. The experts favored variables inherently experienced by patients vs. those not (e. g, reported sadness over observed sadness) and variables with higher mean severity scores (energetic over hyperactive) in an independent sample of bipolar patients studied in the randomized, blinded NIH funded 6 month treatment study data set that contained all the individual state changes for 86 subjects consisting of 628 assessments (2). By analyzing these state changes, the following types of unique reports were produced: Data Profiles, Phase Planes, Behavioral Frequency, Behavioral Transition and Trajectory Analysis.

Conclusions: K-TABS provides a brief narrative interpretation of clinical changes in illness trajectory. Because K TABS focuses on dynamic, not static information, it comports with learning processes by which individuals understand complex but common every day matters in other realms of life. K-TABS ease of use, coupled with a frequently-recorded trajectory of critical symptom variables constitutes a major innovation. KIOS TABS will be evaluated for patient usage, satisfaction, and usability in a three-month field trial of 90 bipolar patients. The expense of treating BD, the insufficient number of health professionals in the field, and the requirements of the Affordable Healthcare Act, call for the implementation of effective self-management programs.

References: 1. Thompson PM et al, Principal domains of behavioral psychopathology identified by the Bipolar Inventory of Signs and Symptoms Scale (BISS). Psychiatry Res, 2010. 175: 221-6. 2. Bowden CL, et al. Lamotrigine vs.lamotrigine plus divalproex in randomized, placebo-controlled maintenance treatment for bipolar depression.Acta Psychiatr Scand 2012; 126:342–350

24 Increased Medication Adherence and Decreased Healthcare Resource Utilization After Pharmacogenetic Testing of Psychiatric Patients

Fagerness J[1], Fonseca E[2], Hess GP[2,3], Dicker R[1], Gardner K[1], Koffler M[1] , Fava M[4], Perlis RH[4], Brennan FX[1], Lombard J[1]

[1]Genomind, LLC, Chalfont, PA, [2]Health Economics and Outcomes Research, IMS Health, Plymouth Meeting, PA, [3]Leonard Davis Institute, University of Pennsylvania, Philadelphia, PA, [4]Massachusetts General Hospital, Boston, MA

Methodological Question Being Addressed: Does a Clinician having access to genetic data impact patient adherence and costs in a psychiatric population?

Introduction: Tools for the practice of personalized medicine, including the use of genetic information, are just beginning to enter wider clinical use. The utility of genetic information as it pertains to clinical decision-making, treatment effectiveness, cost-savings, as well as patient perception remains to be fully elucidated. The objective of this study was to assess the effect of Clinician access to patient genetic information on subsequent patient medication adherence and overall costs.

Method: In this retrospective study, we examined health claims data in order to assess both adherence rates and healthcare costs for patients suffering from psychiatric disorders. These patients were analyzed as Cases, who were patients whose clinicians ordered a genetic test that assists in treatment-based clinical decision-making, versus Controls, who were patients who did not have the genetic test. Cases and controls were propensity score matched in order to avoid covariates likely to confound the analysis. An initial study of 111 cases and 222 controls was performed for both adherence and healthcare costs, and a replication study of 116 cases and 232 controls was performed in which only adherence was assessed.

Results: Overall, Cases were significantly more medication adherent ($p=1.56 \times 10^{-3}$; Cohen's $d=0.511$) than Controls and also showed a relative cost savings of 9.5% over a four month trial period ($562 total cost savings).

Conclusions: These results suggest that pharmacogenetic testing in psychiatric populations improves adherence while leading to cost-effectiveness. Randomized, controlled trials will be necessary to better characterize the direct impact on clinical outcomes, to address potential sources of confounding, and to identify the populations in which this testing may be the most advantageous. Also, additional data about both clinician and patient attitudes and experiences with personalized medicine will further refine how pharmacogenomics is used in practice, and could further influence the effectiveness and cost savings of this type of testing in healthcare.

Disclosures: This study was funded by Genomind, LLC.

25 S(+)-ketamine as a Model for Psychosis

Kleinloog D[1], Uit den Boogaard A[1], Dahan A[2], Klaassen ES[1], van Gerven JMA[1,2]

[1]Centre for Human Drug Research, Leiden, the Netherlands, [2]Leiden University Medical Centre, Leiden, the Netherlands

Methodological Question Being Addressed: How can the ketamine model for psychosis be improved?

Introduction: Symptoms induced by the NMDA antagonist ketamine could constitute a model for psychosis that can be used to determine antipsychotic drug action. However, most studies did not measure pharmacokinetics and different outcome measures were used. The current study aimed to explore the optimal ketamine concentration and outcome measures for use within a psychosis model.

Methods: Healthy subjects received intravenous infusions of S(+)-ketamine to obtain pseudo-steady state concentrations for two hours in a three-way cross-over design (placebo, low-dose, high-dose). Infusion rates were individualized by sex and weight and based on a pharmacokinetic model described by Sigtermans et al. (2009, Anesthesiology 111: 892-903). Target concentrations were selected based on simulations of dosing regimens used in the literature. Psychomimetic effects were measured using the Positive and Negative Syndrome Scale (PANSS), prepulse inhibition (PPI) and different visual analogue scales (VAS). General CNS effects were measured using CHDR's NeuroCart, including eye movements, body sway, pupil size, adaptive tracking and serum cortisol and prolactin levels. Adverse drug effects were evaluated by describing adverse events and measuring vital signs and ECG. Results were analyzed using an ANCOVA with a significance level of 0.05.

Results: The study started with target concentrations of 180 and 360 ng/mL. After six subjects, the cumulative development of adverse events led to a temporary halt and subsequent restart with lower target concentrations of 120 and 240 ng/mL. With these concentrations, an improved safety profile was seen, although there were still many symptoms. During the 240 ng/mL target the infusion was terminated prematurely in 56% of the subjects. A robust psychomimetic effect was seen on all outcome measures, including PANSS (positive subscale: +43.7%, CI 34.4-53.7, $p<0.0001$ for 120

ng/mL and +70.5%, CI 59.0-82.8, p<0.0001 for 240 ng/mL) and PPI (+28.5, CI 10.6-46.4, p=0.0025 and +23.9, CI 4.6 - 43.3, p=0.0164). For most outcome measures a dose-dependent effect was seen, although the effect for PPI and smooth pursuit eye movements was similar for both targets, suggesting a maximum effect. For pupil size and adaptive tracking, only the higher target concentration showed a significant effect.

Conclusions: The ketamine-challenge was found to be a robust method to induce psychomimetic symptoms, that could be measured by the PANSS, prepulse inhibition and VAS. The response was larger and more consistent than found for a similar model using THC. The optimal target concentration for use in a psychosis model would be between 120 and 180 ng/mL, based on a combination of tolerability and level of effect.

26 Identification of PANSS Items of Particular Challenge to Raters in Adolescent Schizophrenia Clinical Trials

Busner J[1], Daniel DG[2], Findling RL[3]

[1]Penn State College of Medicine, Department of Psychiatry, and Bracket, [2]Bracket, [3]Johns Hopkins Medicine, Department of Psychiatry

Methodological Question Being Addressed: Can particularly challenging PANSS items in adolescent schizophrenia trials be identified as a first step toward improving ratings precision and reducing interrater variance?

Introduction (Aims): Global regulatory initiatives have resulted in an increasing number of psychopharmacology trials in the pediatric age range. Challenges in ensuring valid and reliable data in such trials are multiple, and include developmental limitations in symptom description, the need to integrate and weight information from varied sources including parents/caregivers and other informants, and the global shortage of child-trained clinical investigators (Busner, 2013; Farchione, 2013). Adding to the challenges, there are few validated efficacy measures, and many pediatric trials use measures designed for and validated in adults; this is the case for adolescent schizophrenia trials, which frequently use for primary efficacy assessment the (adult) Positive and Negative Syndrome Scale (PANSS), a complex 30-item measure that has been extensively studied and shown to pose ratings challenges even in the adult patients for whom it was designed (e.g., Daniel and Dries, 2013). To identify PANSS items about which raters in pediatric trials might have particular difficulty, we examined PANSS item scoring variability based on standardized qualification videos from several adolescent schizophrenia trials conducted throughout the world by several sponsors.

Method: Using data from multiple clinical trials on adolescent schizophrenia by multiple sponsors, standard deviations were calculated for each of the 30 PANSS items scored by 171 clinical trials investigators/raters in 12 countries who had viewed one of two standardized adolescent patient videos (with accompanying written informant narrative information) as part of the qualification process for their respective clinical trial. The clinical trials investigators/raters had been trained extensively in live sessions on adolescent-specific conventions immediately prior to viewing and scoring the video. PANSS item standard deviations from each video were calculated and rank ordered from lowest to highest variability separately for each video. Rank order of PANSS item standard deviations across videos was examined statistically via Kendall tau.

Results: Ranking of standard deviations across videos did not differ statistically. In the separate calculations by video, rankings of each video included among the respective 10 most variable (highest SD) items N4 (Passive/apathetic social withdrawal), P7 (Hostility), G15 (Preoccupation), G9 (Unusual thought content), and P4 (Excitement). Rankings of each video included among the respective 10 least variable (lowest SD) items N1 (Blunted affect), P3 (Hallucinatory Behavior), G14 (Preoccupation), and P5 (Grandiosity).

Conclusions: Identification of scoring challenges and scoring disagreement for pediatric trials investigators is a first step in targeting training and in-study interventions and ultimately in reducing error variance. Item variability rankings across two differing patient videos were statistically similar, suggesting that scoring ease or difficulty of individual items for adolescent patients occurred over and above the specifics of the patients rated. The high variability items are different from items identified by our group as problematic in adult trials, suggesting there may be unique scoring issues requiring focused attention when PANSS items are applied to the pediatric age range.

Disclosure: J Busner and DG Daniel: Full-time employees of Bracket. RF Findling: past 12 months: receives or has received research support, acted as a consultant and/or served on a speaker's bureau for American Psychiatric Press, AstraZeneca, Bristol-Myers Squibb, Cognition Group, Forest, GlaxoSmithKline, Guilford Press, Johnson & Johnson, KemPharm, Lilly, Lundbeck, Merck, NIH, Novartis, Otsuka, Pfizer, Physicians Postgraduate Press, Rhodes Pharmaceuticals, Roche, Shire, Stanley Medical Research Institute, Sunovion, Supernus Pharmaceuticals, Transcept Pharmaceuticals, Validus, and WebMD.

27 Does Placebo Response Differ Across PANSS Items in Studies of Symptomatic Patients with Schizophrenia?

Bossie C[1], Lee E[1], Fu DJ[1], Ma YW[2], Alphs L[1]

[1]Janssen Scientific Affairs, LLC, Titusville, NJ, USA, [2]Janssen Research & Development, LLC, Titusville, NJ, USA

Methodological Question Being Addressed: Does the response to placebo differ among the 30 individual PANSS items in studies of symptomatic patients with schizophrenia?

Introduction: As the placebo response in schizophrenia trials has increased over the years, it has become challenging to evaluate to what extent improvement in symptoms is driven by a primary treatment effect as opposed to a nonspecific "placebo effect." The 30-item Positive and Negative Syndrome Scale (PANSS) is commonly used to measure symptoms in trials of patients with psychosis. A recent analysis suggested a differential placebo response among the 5 PANSS factors (Marder SR, et al. J Clin Psychiatry. 1997;58:538-546) in a 13-week study of symptomatic patients with schizophrenia, such that the uncontrolled hostility/excitement factor appeared to be the least responsive to placebo, and the positive symptoms factor appeared to be the most responsive (Sliwa JK, et al. Presented at the 166th Annual Meeting of the American Psychiatric Association, May 18-22, 2013, San Francisco, California, USA). This follow-up analysis further examined placebo response by individual PANSS items in this database and in additional, similarly designed studies. The objective of this research is to better understand drivers of placebo response as measured by the PANSS in studies of patients with schizophrenia.

Methods: Included in this evaluation were all placebo-controlled studies from the manufacturer's US registration program for paliperidone palmitate in patients with schizophrenia who were assessed on the PANSS. Four studies were identified; 3 had the same duration (13 weeks) and were used for this analysis (NCT00210548, NCT00101634, and NCT00590577). Only placebo arm data were used (n = 132, 125, and 160, respectively). Mean change from baseline to end point was examined and ranked for each of the 30 PANSS items in each study. Results were examined across studies by visual inspection for this exploratory analysis.

Results: Visual inspection of the rank order of mean change scores suggested some common trends across studies. The items "uncooperativeness," "hostility," "poor impulse control" (all from the uncontrolled hostility/excitement factor), and "poor rapport" (from the negative symptoms factor) were among the PANSS items least responsive to placebo across studies. Items that were among the most responsive to placebo included "hallucinatory behavior," "suspiciousness," "delusions" (from the positive symptoms factor), and "emotional withdrawal" (from the negative symptoms factor). The next most placebo-responsive items included "anxiety," "depression," and "guilt" (all from the anxiety/depression factor).

Conclusion: Exploratory findings from three 13-week studies suggest a possible consistent differential response of certain PANSS items to placebo in studies of patients with symptomatic schizophrenia.

Disclosure: Supported by Janssen Scientific Affairs, LLC. CB, EL, D-JF, and LA are employees of Janssen Scientific Affairs, LLC, and Johnson & Johnson stockholders. Y-WM is an employee of Janssen Research & Development, LLC, and a Johnson & Johnson stockholder.

28 Resilience of a Shortened Version of the Positive and Negative Syndrome Scale Using Parametric Item Response Theory

Anderson A[1], Wang D[2]*, Alphs L[3], Bossie C[3], Romano G[2], Bilder RM[1] on behalf of Janssen-UCLA Neuropsychiatric Phenotype Working Group†

[1]Department of Psychiatry and Biobehavioural Science, University of California, Los Angeles, CA, [2]Janssen Research & Development, LLC, Raritan, NJ, [3]Janssen Scientific Affairs, LLC, Titusville, NJ

Methodological Question Being Addressed: To develop a shortened version of the Positive and Negative Syndrome Scale (PANSS) using parametric item response theory (IRT) and to assess its resilience in an independent dataset.

Introduction: The PANSS is a 30-item instrument widely used in clinical studies for assessing the symptom severity of patients with schizophrenia and the effects of antipsychotic compounds. Using a shorter, more efficient version of PANSS could reduce the noise associated with poor items and thus significantly reduce the sample size needed to show a significant drug effect. Recently, non-parametric IRT was applied to evaluate the usefulness of individual PANSS items and to develop shortened versions of PANSS (Mini-PANSS) by removing poor items (Santor et al 2007, Khan et al 2011). In this study, we applied a parametric IRT approach to develop a Mini-PANSS and demonstrated its resilience in

an independent dataset.

Methods: Baseline PANSS scores of schizophrenia and schizoaffective patients enrolled between 1995 to 2008 in 26 risperidone or paliperidone clinical trials were used as the sample. These patients were divided into two datasets: a training dataset of 7,281 patients from the 16 trials analyzed by Khan et al. (2011) and a test dataset of 2,540 patients from the remaining 10 trials. IRT analysis was performed in the training dataset using the graded response model (GRM) and the generalized partial credit model (GPCM). PANSS items were rated using a multivariate approach, where each item was rated not just by its flaws, but also by the unique information it contributed to the whole test. Bootstrap sampling was also applied to obtain more precise estimates of an item's underlying flaws. We proposed a Mini-PANSS which excluded the items rated as poor by the GRM model and compared its robustness with two previously published non-parametric IRT based Mini-PANSS in the test dataset using two criteria: the information/signal captured and the noise removed.

Results: Both the GRM model and the GPCM model rated 11 items as poor within the training dataset. Out of these poor items, 10 were consistent across models. However, the consistency between the poor items from the parametric IRT and the poor items from non-parametric IRT was modest. Only 5 out of the 11 poor items from the GRM model were also rated as poor by a previously published non-parametric IRT analysis performed using the same dataset. In the test dataset, the proposed Mini-PANSS that excluded 11 items rated as poor by GRM captured more information and removed more noise than the two Mini-PANSS developed using non-parametric IRT.

Conclusions: The rating of poor PANSS items is dependent on the analytical approach. The Mini-PANSS proposed using parametric IRT is more resilient in an independent sample than the Mini-PANSS proposed using non-parametric IRT.

Disclosures: This study is sponsored by Janssen R&D. Dai Wang and Gary Romano are full-time employees of Janssen Research & Development, LLC. Larry Alphs and Cynthia Bossie are full-time employees of Janssen Scientific Affairs, LLC.

References: Santor et al. Item response analysis of the Positive and Negative Syndrome Scale. BMC Psychiatry 2007, 7:66. Khan et al. Use of non-parametric item response theory to develop a shortened version of the Positive and Negative Syndrome Scale (PANSS). BMC Psychiatry 2011, 11:178.

†Members of Janssen-UCLA Neuropsychiatric Phenotype Working Group: Anderson A1, Bilder RM1, Wang D2, Larry A3, Bossie C3, Romano G2, Salvadore G2, Li Q2, Chung H3, Gopal S2, Savitz A2, Fu D3, Rassnick S2, Wilcox M2, Turkoz I2, Cohen N

29 A Comparison of the CANTAB Schizophrenia Battery and the MCCB in Two Phase 2 Clinical Trials of Subjects with Stable Schizophrenia

Baker JD[1], Bain EE[2], Barnett JH1, Haig GM[2]

[1]Cambridge Cognition, Cambridge, UK, [2]AbbVie Inc., North Chicago, IL, USA

Methodological Question Being Addressed: How does the CANTAB schizophrenia battery compare to the MCCB in two phase 2 clinical trials of stable subjects with schizophrenia?

Introduction: The MATRICS consensus cognitive battery (MCCB) was developed as a consensus approach to testing cognition in clinical trials of cognitive impairments associated with schizophrenia (CIAS). The selected tests were chosen to meet psychometric criteria, including adequate test-retest reliability, minimal practice effects, and evidence of sensitivity to drugs. The battery, presented in a mostly pencil-and-paper format, contains ten tests covering seven cognitive domains known to be affected by schizophrenia. The CANTAB schizophrenia battery includes eight computerized neuropsychological tests presented on a touchscreen computer and selected to assess the same seven domains covered by the MCCB. This study evaluated the relationship between the MATRICS and CANTAB schizophrenia batteries in the baseline evaluations of two phase 2 clinical trials.

Methods: Two multicenter trials in stable subjects with schizophrenia were conducted at approximately 45 sites in the United States. Both investigated the procognitive effects of novel investigational compounds as add-on therapy to antipsychotics. The same design was used in both trials, with one placebo and two active dose groups (planned N=70/group; total N=420 for both trials). Eligibility criteria were consistent with MATRICS guidelines. Subjects were currently taking one or two atypical antipsychotics at the time of study entry. The CANTAB battery (excluding Verbal Recognition Memory) and the MCCB were administered for practice on the initial screening visit. Baseline assessments for the two batteries were staggered in order to reduce subject burden such that the MCCB was administered on Day -1

and the CANTAB Battery administered approximately 1-2 weeks prior at Screening Visit 2. The baseline UPSA-2 was administered with the MCCB on Day -1, with no practice assessment during screening.

Results: The correlations between MCCB and CANTAB for the seven MATRICS cognition domain scores varied from r = 0.28 to r = 0.60. The relationship of MCCB with the UPSA-2 agreed with previous published results with r = 0.62. The relationship of the first principal component of the CANTAB schizophrenia battery also correlated with the UPSA-2 at r = 0.55. The MCCB composite correlated with the principal component of the CANTAB battery at r = 0.69

Conclusions: Using data from two large, US-based phase 2 trials of schizophrenia, we compared the cognitive scores obtained at baseline on several hundred patients on the MCCB and CANTAB cognitive batteries. We found that while domain score correlations varied considerably across the different domains, summary scores from the two batteries were highly correlated. The better correlation between composite scores than individual domain scores may reflect the heterogeneity of cognitive domains and that tests from the different batteries assess different aspects of those domains. Both batteries were moderately correlated with functional capacity as measured by the UPSA-2.

Disclosures: J Baker: Employee of Cambridge Cognition. Former employee of Abbvie; holds Abbvie stock. J Barnett: Employee and shareholder of Cambridge Cognition. E Bain: Employee of AbbVie; holds AbbVie stock and stock options. G Haig: Employee of AbbVie; holds AbbVie stock and stock options.

Role of Sponsor: These studies and statistical analyses were supported by AbbVie. AbbVie was involved in the original concepts, analysis, interpretation, writing, and approval of the report.

30 How to Measure Social Cognition in Schizophrenia? A Comparison of Measurements

Khan A[1,2,8], Lucic L[1,7], Rothman R[1,5], GaoL[1], Petersen T[1,3], Opler M[1,4]

[1]ProPhase LLC, [2]Nathan S. Kline Institute for Psychiatric Research, [3]Harvard University, [4]New York University School of Medicine, [5]Columbia University, [6]Pratt Institute, [7]Healthy Russia, [8]Manhattan Psychiatric Center

Methodological Question Being Addressed: Deficits in social cognition have started to show promise as significant factors that contribute to poor social functioning among patients with schizophrenia. Social cognition includes a large number of cognitive abilities that may be essential for adequate social functioning and interpersonal success, for example, recognizing and interpreting emotions, recognizing notable social cues, understanding the mental states of others and evaluating social context. Although some advances have been made in social cognition research in patients with schizophrenia, significance of outcomes has been impeded by a lack of scientifically sound measures. Developing empirically sound measures for social cognition is a key step in interpreting and improving the factors that contribute to deficits in social functioning in schizophrenia. Can investigating the psychometric properties and acceptability of social cognitive instruments contribute to identification of a reliable and valid instrument for patients with schizophrenia?

Introduction (Aims): The main aim of the study was to investigate the psychometric properties and acceptability of a dynamic social cognition scale compared to three commonly used instruments for assessing social cognition in patients with schizophrenia.

Methods: 41 patients with schizophrenia were evaluated in an inpatients and outpatient psychiatric facility to evaluate acceptability (completion time, rate of completion), internal consistency and validity (construct, convergent using confirmatory factor analyses (CFA), root mean square approximation and pearson's t-test) of five social cognition measures: Dynamic Social Cognition Battery (DSCB), Emotion Recognition-40 (ER-40), Facial Emotion Identification Task (FEIT), Tone Matching Task and the MSCEIT. Multiple linear regressions were conducted to identify variables which perform as social cognition determinants.

Results: 41 patients (mean age = 43.15, SD = 9.45; 74% males) were included. The DSCB and FEIT showed good acceptability as evidenced by shorter administration time, completion and patient preference favoring the DSCB. Good levels of internal consistency were found for the DSCB (α = 0.851), ER-40 (α = 0.803), and FEIT (α = 0.782). CFA indicated sufficient to good model fit. The DSCB (RMSEA = 0.08, CFI = 0.91, IFI = 0.89, Chi square = 78.25, p = 0.025) and the ER-40 (RMSEA = 0.06, CFI = 0.84, IFI = 0.84, Chi square = 61.23, p = 0.041) demonstrated a good model fit. The correlations for the DSCB was only significant for the ER-40 (r = 0.512) and FEIT (r = 0.500).

Conclusions: Study findings suggest that DSCB is the preferred instrument to evaluate social cognition, potentially due to it's dynamic nature and short administration time. Further research is needed to develop and improve these existing measurements. Additionally, the DSCB, FEIT and ER-40 show adequate to good reliability and validity.

Disclosures: Authors do not have any disclosures to report.

31 Use of Reliable Change Index to Evaluate Clinical Significance in the Positive and Negative Syndrome Scale (PANSS): a CATIE Analysis

Khan A[1,2,8], Linda Gao[1], Petersen T[1,3], Opler M[1,4], Rothman B[1,5], Lucic L[1,7], Ivanova E[1,7]

[1]ProPhase LLC, [2]Nathan S. Kline Institute for Psychiatric Research, [3]Harvard University, [4]New York University School of Medicine, [5]Columbia University, [6]Pratt Institute, [7]Healthy Russia, [8]Manhattan Psychiatric Center

Methodological Question Being Addressed: The PANSS is the most widely used measures of psychopathology in schizophrenia. It is commonly used in both randomized controlled trials (RCT) and non-controlled evaluations. RCTs assess clinical efficacy of an intervention relative to a placebo or control condition by making group comparisons and evaluating for statistically significant differences. However, statistical significance does not in itself provide concise information about a given intervention's clinically meaningful effects. The process of defining clinical significance remains a challenge. As an attempt to develop a standard method of estimating clinically significant change, we propose adoption of a two-part strategy: The first part of the strategy involves using the Reliable Change Index (RCI). The second part involves use of examination of clinical significance (CS). RCI is whether patients changed sufficiently that the change is unlikely to be due to measurement unreliability. CS change takes the patient from a score typical of schizophrenia to a score typical of the "normal" population. Studying RCI and CS has moved the outcomes paradigm from studying treatment groups to studying individual change within those groups. Assessments must move beyond symptom focus and evaluate individuals with respect to the complex broader domains of their functional, real-world, lives in which clinically significant change is operationalized.

Introduction (Aims): To provide a comparison of concepts and analysis of clinical significance (CS) and the reliable change index (RCI) using pre and post PANSS scores.

Methods: Data on symptomatology, PANSS, from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) were analyzed. Three methods of RCI (Jacobson-Truax, Edwards-Nunnally, and Hageman-Arindell methods) were compared to CS change (pre to post change of at least 2 SDs from the original mean, 20% improvement, and change in PANSS remission criteria).

Results: For the three RCI methods, 29.73%, 31.08% and 52.70% showed reliable improvements in PANSS scores. For CS, 22.97% showed greater than 20% improvement, 29.73% improved on the PANSS remission criteria, and only 8.11% showed CS improvement of 2 SDs from the mean. When comparing RCIs with CS, only 18.92% of CS improvement also resulted in RCI significant improvement. Regarding clinically meaningful improvement, the Hageman-Arindell method was most concordant with all three RCI measures and with the 20% improvement as this method differentially analyzes clinically meaningful change at the individual level and at the group level (i.e., obtaining proportions of patients who have reliably changed and passed the cutoff point).

Conclusions: Reliable and clinically significant change should be reported in articles to complement the more familiar group summary methods. Assessment of clinically meaningful change is useful for evaluating treatment response. Outcome studies often assess statistically significant change, which may not be clinically meaningful. Comparisons of the proposed methods of determining clinically significant PANSS outcomes to biomedical standards of clinical significance will help determine the validation of this procedure, and improve the precision and effectiveness of the PANSS in clinical trials.

Disclosures: Authors do not have any disclosures to report.

32 Assessing the Personal and Social Performance Scale as a Key End Point in a Relapse Prevention Trial for Mental Illness

Turkoz I[1], Ascher S[1], Mao L[1], Fu, DJ[2], Alphs L[2], Olson W[1]

[1]Janssen Research & Development, LLC, Titusville, NJ, USA, [2]Janssen Scientific Affairs, LLC, Titusville, NJ, USA

Methodological Question Being Addressed: How should one evaluate a treatment difference with respect to a functional measure in a relapse prevention trial in which a substantial number of patients have missing outcome data?

Introduction: In many relapse prevention trials, patients are scheduled to provide outcome data at a fixed number of follow-up visits; however, patients may be prematurely discontinued from follow-up after meeting predefined relapse criteria. In these trials, patients also may withdraw prematurely for other reasons, such as withdrawal of consent, loss to follow-up, and adverse events. As a result, a substantial number of patients may not provide outcome data at the last preplanned study visit. Missing data make it challenging for investigators to estimate the treatment difference that would

exist at this study visit under complete follow-up. Various statistical methods have been or are being developed to address this issue. The objective of this analysis is to illustrate and compare methods of estimating treatment differences with respect to the Personal and Social Performance (PSP) scale at the last study visit.

Methods: Analysis was based on the PSP total score from a randomized, double-blind, relapse prevention trial in patients with bipolar disorder. PSP total score was reported at study entry and at several follow-up visits until week 48.

To assess robustness and consistency of findings at week 48 end point, 4 methods will be used: (1) mixed model repeated measures (MMRM), which assumes data were missing at random; (2) a pattern mixture model (PMM), which assumes data were not missing at random; (3) last observation carried forward (LOCF) assessment with forced worsening of the PSP total score (up to 10%) in the treatment arm for patients who relapsed or discontinued; and (4) a G computation–based global sensitivity analysis method, which formally evaluates the robustness of results to increasing deviations from a missing-at-random–type assumption.

Results: 109 (40%) subjects completed 48 weeks of double-blind treatment, 115 (42%) discontinued because of relapse, and 51 (19%) discontinued for other reasons. Of 109 subjects who completed the study, 35 (26%) were in the placebo group and 74 (53%) were in the active treatment group. Results obtained by the aforementioned methods will be reported.

Conclusions: Appropriate management of missing data is a complex, yet common problem in CNS clinical trials. This work demonstrates the variation in results seen with 4 different statistical approaches. Reasons for the observed variation in results will be discussed.

Disclosure: Funding provided by Janssen Scientific Affairs, LLC. IT, SA, LM, and WO are employees of Janssen Research & Development, LLC, and Johnson & Johnson stockholders. D-JF and LA are employees of Janssen Scientific Affairs, LLC, and Johnson & Johnson stockholders.

33 Novel Assessment Tool for Patients with Schizophrenia: Development of the Daily Activity Report

Velligan DI1, Gregory A. Maglinte GA1, Sierra C1, Martin, ML3, Corey-Lisle P2

[1]Department of Psychiatry, University of Texas Health Science Center, San Antonio, [2]Amgen Pharmaceuticals, [3]Health Research Associates, Inc.

Methodological Question Being Addressed: Might a measure of detailed daily activity be a useful outcome measure in studies of medications targeting negative symptoms or functional outcome in schizophrenia?

Introduction: The assessment of real-world functional outcomes for patients with schizophrenia is complicated by a host of factors. Functional capacity measures can capture a patient's ability to perform a limited set of instrumental or social skills but are limited in their ability to, predict whether a patient actually performs these skills in the real world. Clinical interviews are limited by availability of reliable informants who can provide an accurate report of a patient's daily activities and behavior. Ecological momentary assessments evaluate what an individual is doing at a specific time via real-time data collection, but may not provide a holistic evaluation of activity and behavior. Our objective is to develop a comprehensive and patient-centric measure of daily functioning called the Daily Activity Report (DAR).

Methods: Development of the DAR will be an iterative process including inputs from multiple evaluations. One evaluation is focused on a critical review of available clinical outcome assessment tools to evaluate the feasibility of adaptation for use in assessing daily activities and functional outcomes in patients with schizophrenia. Another component obtains qualitative data through individual interviews with key informants (clinicians and non-clinical caregivers) about the most relevant concepts they observe that are related to productive activity engagement. A third evaluation assesses the importance and relevance of different productive activities through focus groups and structured interviews of patients with schizophrenia. When combined, this information will enable identification of relevant concepts for incorporation into a preliminary version of the DAR. This version of the DAR will undergo psychometric evaluation, including reliability and validity testing. The DAR is anticipated to collect data on instrumental activities, social activities and work/ academic activities performed by a patient over a 7-day period using structured interview reports.

Results: Thus far, 10 semi-structured clinician interviews and 12 semi-structured caregiver interviews have been conducted to broaden the information around importance of functional activity to assessing patient improvement. In addition, 12 patients have participated in focus groups yielding both qualitative and quantitative data on the importance and difficulty of various daily activities. Structured interviews have also been conducted among 50 patients, of whom 30 have completed a one-month follow-up.

Conclusions: When DAR development is complete, it is anticipated to provide in-depth assessments of a patient's daily activity, which may aid the evaluation of a patient's level of motivation, engagement with their environment, and provide greater sensitivity in assessing the impact of treatment interventions on functioning. Additional steps may allow for the DAR to be adapted to a patient-reported outcome (PRO) measure, pending further development and validation.

Disclosures: PK Corey-Lisle and G Maglinte are employees and shareholders at Amgen, Inc. Thousand Oaks, CA

34 A Recovery Outcome Model Illustrated by the Methodology of a VA Cooperative Study Program #589

Davis LL[1], Yang S[2], Kyriakides T[2], Guarino P[2], Toscano R[1], Resnick S[3], Mueller L[4], Lisa Ottomaneli L[5], Drake R[6]

[1]Tuscaloosa VA Medical Center (VAMC), [2]West Haven Cooperative Studies Coordinating Center, [3]VISN 1 New England Mental Illness Research and Clinical Center, West Haven VAMC and Yale School of Medicine, [4]Bedford VAMC, [5]Tampa VAMC, [6]Dartmouth Psychiatric Research Center

Methodological Question Being Addressed: Can an occupational measure serve as the primary outcome of a mental health intervention study in keeping with a recovery model?

Introduction: Often the primary outcome of a randomized clinical trial (RCT) is focused on symptom improvement for the mental disorder being studied, which is in keeping with a medical model. Relevant social and occupational outcomes, based on a recovery model, are often entirely ignored or relegated to an exploratory aim of an RCT. Adults with mental illnesses, such as posttraumatic stress disorder (PTSD), have a broad range of symptoms and severity that do not always correlate with the individual's ability to work in a competitive job. For many individuals, the primary treatment goal is improved occupational and social functioning rather than symptom remission. Many mental health providers wait until symptom improvement is maximized before referring the patient to vocational rehabilitation services, based on inaccurate beliefs that employment exacerbates mental health symptoms. However, the chronic symptoms of PTSD often improve as the individual regains work and a meaningful activity.

Methods: The primary objective of "CSP#589 Veterans Individual Placement and Support Towards Advancing Recovery (VIP-STAR)" is to evaluate the effectiveness of Individual Placement & Support (IPS), a model of supported employment, in unemployed Veterans with PTSD. The primary hypothesis is that, compared to those treated with a traditional Department of Veterans Affairs transitional work program (TWP), unemployed Veterans with PTSD treated with IPS will be significantly more likely to become a steady worker. The primary outcome is the proportion of study participants who achieve the status of a steady worker, defined as holding a competitive job for greater than or equal to 50% of the 18-month study follow-up period (i.e., greater than or equal to 39 of the 78 weeks). All participants will be followed for 18 months post-randomization. Other secondary and exploratory outcomes include PTSD symptoms, self-esteem, PTSD functional outcomes, quality of life, and negative health outcomes/events.

Conclusion: There are compelling reasons for utilizing functional outcomes as a primary aim of a psychotherapy or pharmacotherapy trial. A full description of the methods and rationale of CSP #589 VIP-STAR will be presented, as an illustration of an RCT with a primary outcome that is recovery oriented.

Disclosure: The authors report no conflicts of interest for this work.

Source of Support: Veterans Administration Cooperative Studies Program; ClinicalTrials.gov Identifier: NCT01817712

35 The Significance of Social Cognition in Patients with Schizophrenia

Lindenmayer JP[2,3,4], Khan A[1,2,3], Sebastien L[3], Jit R[3], Michaels T[3], Mortensen D[3,6], Thanju A[3], McGurk S[5]

[1]ProPhase LLC, New York, NY, [2]Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, [3]Manhattan Psychiatric Center, Wards Island, NY, [4]New York University School of Medicine, New York, NY, [5]Boston University, Boston, MA. [6]Columbia University, New York, NY

Methodological Question Being Addressed: Deficits in social functioning, including communication, work, social skills, and community functioning, are a defining feature of schizophrenia. Functional outcomes of schizophrenia are affected by several factors such as social cognition, neurocognition, psychopathology, and clinical outcomes. The multifaceted association among these factors and functional outcome continues to be unclear. Given the significant role of functional outcomes in schizophrenia, there has been increasing importance in factors that may underlie these outcomes. If the characteristics of these factors can be defined, interventions may be developed to improve them, which, in turn, will have a parallel impact on long term functioning and outcome. Can determining the relationship between social cognition, neurocognition, and psychopathology on functional outcome, indicate which of these factors act as a

potential treatment target for improving functional outcomes in schizophrenia?

Introduction (Aims): The current study utilized structural equation modeling (SEM) and path analysis to examine whether social cognition outcomes, neurocognition and clinical symptoms have a relationship on functional outcomes in patients with schizophrenia.

Methods: 41 patients with chronic schizophrenia received evaluation of cognitive function (using the Measurement and Treatment Research to Improve Cognition in Schizophrenia [MATRICS] Consensus Cognitive Battery), clinical symptoms (as measured by the Positive and Negative Syndrome Scale), social cognition as measured by the Dynamic Social Cognition Battery (DSCB), emotion identification (FEIT, ER-40) and functional outcome as assessed by University of San Diego – UPSA-Brief and the Personal and Social Performance Scale.

Results: SEM identified clinical symptoms, attributional styles and emotion recognition as significantly related to functional outcome and working memory, processing speed as a mediator between cognition and functional outcome. The relationship between social cognition as measured by the DSCB and functional outcome was significant in the basic model. In the mediation model, the link between neurocognition and functional outcome was mediated by negative symptoms. The results of the analysis support a causal model that indicates that social cognition underlies and is causally primary to functional outcomes.

Conclusions: This study suggests that social cognition, negative symptoms, mediate the influence of neurocognition on functional outcomes of schizophrenia. Our findings suggest that social cognition may be essential to improve functional outcomes. These findings provide evidence that may help develop novel interventions.

Disclosures: Authors do not have any disclosures to report.

36 Development of the Readiness for Work Questionnaire in Schizophrenia – Further Assessment of Construct Validity

Edgar C[1], Potkin SE[2], Bugarski-Kirola D[3], Le Scouiller S[1], Kunovac J[4], Velasco EM[5], Berardo CG[3], Garibaldi GM[3]

[1]Roche Products LTD, Welwyn Garden City, UK, [2]University of California, Irvine, Department of Psychiatry and Human Behavior, 5251 California, Suite 240, CA, USA 92617, [3]F. Hoffmann-La Roche LTD, Basel, Switzerland, [4]Excell Research Inc., Oceanside, CA, USA, [5]Resolution Psychopharmacology Research Institute, Mendoza, Argentina

Methodological Question Being Addressed: Unemployment rates are high among patients with schizophrenia and unemployment has a negative impact upon quality of life. There is also a significant societal and economic burden associated with unemployment. Ability to work should be considered as a therapeutic goal with potentially wide-ranging benefits. However, employment status is dependent on the availability of work, stigma of mental illness and cultural influences, and patients' work history and educational background. A clinician-rated scale evaluating readiness for work (independently from work status) would be a potentially valuable assessment tool.

Introduction: The objective of this study was to evaluate the association between negative symptoms of schizophrenia and readiness for work, as well as the validity of a work readiness questionnaire (WoRQ) that would allow clinicians to assess patient function with respect to ability to engage in socially useful activity. One of the prespecified validation steps was to evaluate the WoRQ determinations with presence of negative symptoms.

Methods: Construct validity was evaluated in a global, cross-sectional, observational, stand-alone validation study. Two hundred, male and female, adult outpatients with schizophrenia (DSM-IV) were included in the study. Recruitment was oversampled such that 25% of patients were working independently at the time of assessment. The association between readiness for work status and negative symptoms was evaluated based on scores of the 4-item Negative Symptom Assessment (NSA-4).

Results: Readiness for work showed an overall association with statistically significantly lower levels of negative symptoms on the NSA-4. Consideration of the individual NSA-4 items showed the strongest relationships to be between readiness for work and both 'reduced interest' and 'reduced social drive', as opposed to the 'restricted speech' and 'reduced emotion' items.

Conclusions: The WoRQ was strongly associated with negative symptom severity in the predicted direction. Those symptoms more related to apathy-avolition on the NSA-4, as opposed to deficit of expression, showed a stronger relationship to readiness for work.

Disclosures: This research was supported by F. Hoffmann-La Roche LTD.

37 Reliability and validity of the PANSS negative symptom factor score in outpatients with schizophrenia prescribed select antipsychotics and with prominent negative or disorganized thought symptoms

Edgar C[1], Blaettler T[2], Bugarski-Kirola D[2], Le Scouiller S[1], Garibaldi GM[2]

[1]Roche Products Ltd, Welwyn Garden City, UK, [2]F. Hoffmann-La Roche Ltd, Basel, Switzerland

Methodological Question Being Addressed: Negative symptoms of schizophrenia lead to poorer functional outcomes and greater reductions in quality of life (QOL) than positive symptoms, and targeting negative symptoms may have significant functional benefits. No consensus exists regarding the best instrument for measuring negative symptoms in clinical trials. The Positive and Negative Syndrome Scale (PANSS) 'Marder' factor scores (Marder et al., 1997) have been widely used as secondary/exploratory clinical trial outcomes. Use of the negative symptom factor score (NSFS) from the PANSS could reduce burden on sites and patients, versus the use of new/separate negative symptom scales. However, there are limited published data supporting the validity of 'Marder' factor scores.

Introduction: The PANSS is a valid instrument for the assessment of severity of schizophrenia symptoms. Good test-retest reliability for the total score and subscales has been reported and adequate inter-rater reliability at the item level. Analyses have consistently identified a five-factor solution for the PANSS items, typically in domains of negative, positive, disorganized thought, hostility/excitement and anxiety/depression symptoms. The NSFS is perhaps the most robust, having been replicated in multiple separate analyses. The NSFS may be considered to have several aspects of improved content validity compared with the original negative subscale and is strongly correlated with 16-item Negative Symptom Assessment and weakly with the Montgomery–Åsberg Depression Rating Scale. Despite its wide use as a clinical trial outcome and the converging evidence for reliability and validity, some aspects of validation remain to be confirmed in patients with predominant negative symptoms.

Methods: The present data are from a Phase II proof-of-concept study in schizophrenia outpatients treated with select antipsychotics and with prominent negative or disorganized thought symptoms (sum of PANSS negative and disorganized thought item scores ≥40). Analyses were conducted to evaluate test-retest reliability (intra-class correlation coefficient [ICC]), internal consistency (Cronbach's alpha) and construct validity (association to Clinical Global Impressions severity [CGI-S] and Personal and Social Performance [PSP]).

Results: Test-retest reliability from the screening to the baseline assessment around 2 weeks later was high (ICC=0.93). Internal consistency at the baseline visit was good (Cronbach's alpha=0.71) and increased with subsequent assessments, indicating a strong degree of association between the seven NSFS items. This association was maintained when ICCs were calculated between successive visits in the study for those patients with no change in CGI-S ratings. Correlations at baseline (raw scores) showed a good association between NSFS and the CGI of negative symptom severity (0.63), but a weaker association to overall CGI-S (0.31). Association with PSP score at baseline was moderate (-0.39), though higher than for the other PANSS factor scores and the PANSS total score.

Conclusions: The analyses have demonstrated both the reliability (test-retest) and validity (internal consistency and relationship to CGI-S) of the NSFS, in a population of patients with schizophrenia and with prominent negative or disorganized thought symptoms.

Disclosures: This research was supported by F. Hoffmann-La Roche Ltd.

References: Marder SR, Davis JM, Chouinard G. The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: combined results of the North American trials. J Clin Psychiatry 1997;58:538–546

38 Working Group Reports