ISCTM 2014 Autumn Conference

6-8 October 2014

Boston Park Plaza, Boston, MA

Poster Abstracts

Groupings:
Cognitive Assessment and Tools: 1-10
Assessment Tools Other Than Cognitive: 11-19
Clinical Trial Designs and Applications: 20-24
Workshop Reports: 25

1 A Report on MCCB:UPSA Change Score Correlations from Clinical Trials for Cognitive Impairment Associated with Schizophrenia

Silbert R, Haig G

AbbVie Inc.

Methodological Question Being Addressed: What are the MATRICS Consensus Cognitive Battery (MCCB) and University of California Performance-Based Skills Assessment-2 (UPSA-2) change score correlations?

Introduction: Cognitive impairment is an important and under-treated aspect of schizophrenia. Two assessments commonly used in clinical trials to assess the degree of cognition and functioning are the MCCB and UPSA-2. The MCCB contains ten tests covering seven cognitive domains that are known to be affected in patients with schizophrenia. The UPSA-2 is a functional measure that uses role-play tests to evaluate cognitive functional capacity in six selected domains of basic living skills. Baseline correlations between the MCCB and UPSA are~0.60, however MCCB and UPSA change score correlations have not been reported. This analysis evaluated the change score correlation between the MCCB and UPSA-2 over the course of a 12-week treatment in two Phase 2 clinical trials. The standard deviation of the change scores, practice effects and placebo response were also investigated.

Methods: Two multicenter trials in stable subjects with schizophrenia, Trial A with the histamine-3 antagonist ABT-288, and Trial B with the alpha-7 agonist ABT-126, were conducted to investigate procognitive effects as add-on therapy to antipsychotics. Approximately 45 US sites were used. Subjects were receiving stable doses of antipsychotics at the time of enrollment. Each trial employed a placebo and two active dose groups. Trial A enrolled 207 subjects, Trial B enrolled 214. Eligibility criteria were consistent with MATRICS guidelines. In both trials, the MCCB was a primary efficacy measure while the UPSA-2 was a secondary efficacy measure. The MCCB was administered at screening, baseline, and final visit (12-weeks) while the UPSA-2 was administered at baseline and Week 12 only. This analysis included 325 subjects who completed both MCCB and UPSA-2 assessments in both trials at baseline and final visit.

Results: The overall correlation (R) from both studies between the MCCB and UPSA-2 total scores at baseline was 0.621 (n=411). The change score correlations (baseline to Week 12) between the MCCB and the UPSA-2 were 0.154 (n=163) and 0.162 (n=162) in Trials A and B, respectively, and 0.156 (n=325) overall. The standard deviations for the MCCB change from baseline to final visit scores were 5.216 (n=165) and 5.028 (n=166) for Trials A and B, respectively. The standard deviations of the change on the UPSA-2 were 12.605 (n=174) and 9.509 (n=168) for Trials A and B, respectively. The overall screening to baseline practice effect for the MCCB was 1.793 (n=417); for Trials A and B, respectively it was 2.357 (n=204) and 1.253 (n=213). Placebo response data are also reported for both the MCCB and UPSA-2.

Conclusions: While baseline scores on the MCCB and UPSA-2 have been shown to demonstrate high correlation, the correlation between change scores on the MCCB and UPSA-2 over the course of these two 12-week interventional studies is weak.

Disclosures: This work was funded by AbbVie Inc. AbbVie participated in the study design, research, data collection, analysis and interpretation of data, writing, reviewing, and approving the publication. RS and GH are both employees of AbbVie.

2 Effects of Levomilnacipran ER on Measures of Attention in a Phase III Trial of Major Depressive Disorder

Wesnes K[1], Gommoll C[2], Chen C[2]

[1]Wesnes Cognition Ltd, [2]Forest Research Institute

Methodological Question Being Addressed: Are measures of attention improved during treatment of major depressive disorder with levomilnacipran? Also, are changes in measures of attention the result of, or are they independent of, changes in depressive symptoms during treatment of major depressive disorder?

Introduction (Aims): This analysis was conducted to evaluate changes in cognitive measures in patients with major depressive disorder (MDD) treated with levomilnacipran ER (LVM) versus placebo (PBO), and assess relationships between cognitive changes and depression symptoms.

Methods: In an 8-week study of flexible-dose LVM 40-120 mg versus PBO in adult patients with MDD (NCT00969709), cognitive assessments included changes from baseline on the Bond-Lader Visual Analog Scales (VAS) of Mood and Alertness and the Cognitive Drug Research (CDR) System. The CDR tests speed and accuracy on 3 computerized tasks (Simple Reaction Time, Digit Vigilance, and Choice Reaction Time); test results are used to derive 4 composite scores: Power of Attention (PoA), Continuity of Attention (CoA), Cognitive Reaction Time (CogRT), and Reaction Time Variability (RTV).

Post hoc analyses evaluated relationships between changes in cognitive measures and baseline depressive symptoms (as measured by MADRS Total and Item 6 [Difficulty Concentrating] scores). Relationships between MADRS response (≥50% improvement) and changes in cognitive measures were assessed.

Results: Of 429 ITT patients, 187 PBO and 182 LVM patients had valid Week 8 cognitive assessments and were included in the analyses. Greater improvement for LVM versus PBO in depressive symptoms was accompanied by greater improvement over PBO in all measures of attention (CoA, *P*=.0016; PoA, *P*=.0382; RTV, *P*=.0237;) except CogRT (*P*=.3409). There was no significant correlation between attention measures and MADRS scores at baseline or changes from baseline. However, LVM MADRS responders experienced significant improvements from baseline in PoA, CoA, and RTV; in contrast, LVM non-responders and PBO patients did not show significant improvements in cognitive measures (except for significant improvement in PoA for PBO responders).

Conclusions: LVM treatment was associated with significant improvement over PBO in depressive symptoms and certain cognitive measures. MADRS response in LVM-treated patients was associated with significant improvement on most cognitive measures that were assessed.

Disclosures: Supported by funding from Forest Laboratories, Inc. and Pierre Fabre Médicament. During the conduct of this study Keith Wesnes was an employee of Bracket which provided the computerised tests used in the study.

3 Reliability of MATRICS Consensus Cognitive Battery (MCCB) in Institutionalized Patients with Treatment Refractory Schizophrenia

Savitz A, English J, Galvin K, Siegel D

Weill Cornell Medical College

Methodological Question Being Addressed: Can the MATRICS Consensus Cognitive Battery (MCCB) be completed with and reliable in a treatment refractory schizophrenia population.

Introduction: In people with schizophrenia, cognitive functioning has a strong correlation to functional ability. MCCB was developed to assess cognition in outpatients with schizophrenia but its reliability and usability has not been established with highly symptomatic, treatment refractory patients. The goal of this analysis was to determine if the MCCB could be completed in this population without large amounts of missing data and if test-retest reliability can be established which would be needed for future analysis of any change in MCCB scores.

Methods: To assess the MCCB in this population, certified raters administered the MCCB to subjects who were patients in an inpatient, psychiatric rehabilitation program. All subjects had a diagnosis of schizophrenia or schizoaffective disorder and ranged in age from 18 to 65. Participation in the research study was voluntary, open to all patients on the unit whose primary language was English, and written informed consent obtained. The MCCB was one part of the assessment to determine predictors of outcome in the program as well as better define the correlation between outcome in the program and cognitive ability. The initial MCCB assessment occurred soon after admission to the inpatient program,

and a second assessment was initiated two to four weeks after the first MCCB.

Results: Forty-seven subjects completed the initial MCCB. Twenty-two subjects completed the test battery twice within four weeks. Only one subject out of 47 had missing data from the MCCB. In these treatment refractory subjects, the MCCB was found to be reliable; Intraclass Correlation Coefficients for test-retest reliability of individual tests ranged from .728 to .975. Subjects who refused to be retested or could not be retested due to exacerbation of symptoms did not differ significantly from the retested cohort in terms of demographics or ratings of positive or negative symptoms at the first testing period. For both groups, the initial MCCB composite total scores correlated with scores of daily functioning on the unit, indicating the validity of the instrument within this population to predict functional ability.

Conclusions: The MCCB is a suitable instrument to test cognition in people who have been institutionalized with treatment refractory schizophrenia.

Disclosures: The authors report no conflict of interest with this work. Of note, the first author is primarily employed by Janssen R&D which had no input into this work. This work was performed at Weill Cornell Medical College and was supported by institutional resources.

4 Is Identical Scoring of the PANSS Across Consecutive Visits a Marker of Poor Data Quality?

Daniel D, Kott A

Bracket Global, LLC

Methodological Question Being Addressed: In clinical trials, is identical scoring of complex subjective rating scales across visits a marker of poor quality interviews and scoring?

Background: Risk based monitoring of subjective rating scales attempts to identify signals of aberrant patient assessment that warrant further investigation. We have recently become interested in high rates of consecutive visits rated identically as a potential data quality issue (Daniel and Kott, 2014). In the current dataset there are 76,062 visits out of which there are 43,584 that have a potential to be identical. Out of these there are 2067 identical visits making it 4.74% of all possible visits.

The purpose of the current analysis was to provide further context for interpretation of the finding of identical PANSS ratings across visits in clinical trials. Specifically, we asked how likely is it that a skilled rater could rate a clinical trials patient that was unchanged across visits the same on all thirty PANSS items? To address this, we examined the rate of identical scoring by investigators of the same videotaped PANSS interview.

Method: Investigators participating in global schizophrenia clinical trials were required to rate videotaped PANSS interviews to establish and maintain accuracy and precision of PANSS ratings. One hundred and ninety four raters rated Tape One and re-rated the same tape after approximately six months. Sixty investigators rated Tape 2 and re-rated the same tape after approximately six months. Not all investigators participated in all four rating exercises, in part because some joined the study at different time points. The mean years of schizophrenia clinical trial experience of the raters was 7.4 years +/- 5.9. The median was 6.4 years. Two hundred and ninety eight (90%) of the raters held a doctoral degree (MD, PsyD, PhD).

Results: The intra-rater ICC for Tape 1 (n=194 pairs) was 0.54 (SE+/- 0.037). 25 (12.9%) of raters scored the video exactly the same way (all 30 PANSS items identical) on the second rating as on the initial attempt. The intra-rater ICC for Tape 2 (n=60 pairs) was 0.53 (SE+/- 0.070). Eleven (18.3%) of the raters scored the video exactly the same way (all 30 PANSS items identical) as on the initial attempt.

Discussion: The PANSS rates a wide variety of symptoms and behaviors based on patient and informant report over the rating period as well as direct observation during the interview. A valid rating of a patient as unchanged on all thirty PANSS items requires a very high degree of stability on the part of the patient and consistency in reporting by both the patient and informant. The results suggest that even in the case of re-rating a videotaped interview in which there would be no information variance most raters would not score the videotape identically across visits. Moreover, the test-retest correlation coefficients in the current analysis as well as in the literature suggest that the psychometric properties of the scale challenge achieving 30/30 identical ratings even in more stable patients. Published test-retest correlations of the PANSS subscales have been reported from .60-.80 in chronic patients (Kay et al, 1989) .37-.43 in subacute patients (Kay and Singh, 1989) and .13-.24 in acute patients (Lindenmayer et al, 1984).

In summary, in risk based monitoring of blinded clinical trials data, identical ratings across visits appear to constitute a signal of potential quality issues and warrants further investigation.

5 Cultural Adaptation of Translated Neurocognitive Assessments in Russia, Switzerland and Italy: Pilot Testing in Preparation for a Program to Delay the Onset of Mild Cognitive Impairment (MCI) due to Alzheimer's Disease (AD)

Atkins AS[1], Vaughan A[1], Turcotte N[1], Makeeva O[2], Monsch AU[3], Frisoni GB[4], Parapini M[4], Zhukova I[2], Melikyan Z[5], Brewster S[5], Oneil J[6], Schuster J[6], Lymboura M[7], Romero HR[8,9], Hayden KM[8,9], Plassman BL[8,9], Welsh–Bohmer KA[8,9], Keefe RSE[1,9]

[1]NeuroCog Trials Inc., [2]Neuropsychology Testing Center, Nebbiolo, LLC, [3]University Center for Medicine of Aging Basel, [4]IRCCS Centro San Giovanni di Dio Fatebenefratelli, [5]Zinfandel Pharmaceuticals Inc., [6]Takeda Development Center, Americas, [7]Takeda Development Centre, Europe, [8]Joseph and Kathleen Bryan ADRC, Duke University Medical Center, [9]Department of Psychiatry, Duke University Medical Center

Methodological Question being Addressed: Cross-cultural adaptation of neurocognitive assessments.

Introduction (Aims): Cultural adaptation of neurocognitive assessments can improve the quality of translated instruments by ensuring tasks, stimuli and instructions are understood and appropriate for use in populations of interest. We present data from pilot investigations collecting feedback on the cultural and linguistic accuracy of a translated neurocognitive battery adopted for use in an upcoming investigation of transition from normal aging to MCI due to AD. For all languages assessed, changes were incorporated to improve the quality of adapted instruments and account for cultural and linguistic differences.

Methods: Neurocognitive Battery measures included the Mini-Mental State Examination, Brief Visuospatial Memory Test, California Verbal Learning Test, Animal Fluency, Lexical Fluency, WAIS-II Digit Span, Trail Making Test (TMT), Clock Drawing Test and a relatively new measure, the Multilingual Naming Test (MiNT). Pilot studies were conducted in Russia, Switzerland (German) and Italy in accordance with both ICH guidelines for Good Clinical Practice and Pharmacoeconomics and Outcomes Research (ISPOR) guidelines for cultural adaptation. Each country sample included ten participants, ages 65-86. Each adapted measure received formal review by three independent neuropsychologists/psychologists in each country. Feedback was requested regarding construct validity, appropriateness for the target population, potential performance differences and recommended revisions.

Results: Pilot studies provided suggestions for improving the translation and adaptation of each measure and revealed cultural variations in participants' experience. Russian feedback indicated potential differences in speed of processing tasks due to a strong cultural emphasis on accuracy, reduced exposure to alphabetical sequencing (TMT B), and relative unfamiliarity with timed testing. Significant adaptations to the MiNT naming task were requested (25% of comments received) to account for regional differences in stimuli and culturally specific nomenclature. In all countries, changes to written language were requested (48% of comments) in order to clarify task demands to produce better understanding.

Conclusions: Adapting existing neurocognitive measures for use in other cultures requires careful balance in efforts to achieve cultural appropriateness while maintaining the integrity of the original instruments. In the present study, in-country pilot exercises provided important input regarding cultural and linguistic variations to improve translated assessments and forecast potential cultural influences on performance.

Disclosures: Funding was provided by Takeda Pharmaceutical Company Limited, Zinfandel Pharmaceuticals and NeuroCog Trials.

6 Cognitive Dysfunction in Major Depressive Disorder is Associated with Higher Levels of Self- rated Disability and Lower Work Productivity

Jaeger J, Maruff P

CogState Ltd

Methodological Question Being Addressed: The extent to which there is cognitive dysfunction in people with major depressive disorder

Background: Depressive disorders are associated with impairments in cognitive function but their impact on disability and productivity, is not well understood. The aim of this analysis was to assess the relationship between cognitive dysfunction in patients and patient-reported outcomes of productivity and disability in major depressive disorder.

Methods: 123 patients (72(58%) female, mean age = 53.2 years, SD age 12.1 years) who met DSM V criteria for major depressive disorder who were receiving treatment with antidepressants completed the CogState Brief Battery CBB), the Patient Health Questionnaire (PHQ-9), the Sheehan Disability Scale (SDS), and the Work Productivity and Activity Impairment (WPAI) in a single session conducted at a research laboratory at major metropolitan hospital. Cognitive

dysfunction was classified in an individual patient if performance on two of the four tasks from the CBB were impaired (i.e. z</=-1) relative to normative data. Ratings on the PHQ-9, SDS and WPAI were compared between the cognitive dysfunction and normal cognition MDD subgroups.

Results: Of the MDD group 69 patients meet criteria for cognitive dysfunction (56%). There was no difference between cognitive dysfunction and normal cognition groups in age, proportion of females, years education or depression severity. The cognitive dysfunction group showed significantly lower scores on the SDS total score and the WPAI productivity index, but no differences between cognition subgroups were observed for the PHQ-9.

Conclusions: Cognitive dysfunction was common in patients with MDD who were receiving treatment with antidepressants. In patients with MDD, cognitive dysfunction was associated with self-rated disability and lower work performance and productivity. These data suggest that cognitive dysfunction is an important consequence of MDD.

7 Development and Validation of a Concise Cognitive Assessment for use in Phase 1 Clinical Trials

Housden CR [1, 2], Hermans LF[1], Sambeth A[3], Verhaeg A[3], Rock PL[4], Blackwell AD[1, 2], Nathan PJ[1, 2]

[1]Cambridge Cognition, [2]Department of Psychiatry, University of Cambridge, UK, [3]Department of Psychology and Neuroscience, University of Maastricht, NL, [4]The Wharton School, University of Pennsylvania

Mehodological Question Being Addressed: Measuring the cognitive profile of drugs in Phase 1 trials is valuable as it can help to determine the cognitive safety of a drug early on, to define cognitively safe dose ranges and to manage risks in later stage development. However, this can be a challenge as the cognitive assessment needs to be brief, be able to detect changes in healthy individuals, and have good psychometric properties in a small sample sizes.

Introduction (Aims): We aimed to develop and validate a concise cognitive assessment, called Clinical Trial Information System-Profile (CTIS-Profile), to enable sensitive measurement of cognitive performance in Phase 1 clinical trials, with minimal burden to subjects and raters.
Abbreviated cognitive test variants were specified following in-depth analysis of clinical trial cognitive data, and the tests were optimized for healthy individuals. We investigated whether this brief, automated assessment would have the required psychometric properties to be used in a Phase 1 clinical trial.

Methods: This study compared cognitive performance measured using CTIS-Profile delivered on an iPad to established Cantab Solutions software delivered on a Motion touchscreen tablet.
Seventy-one healthy individuals aged 19 to 67 years (mean ± standard deviation: 40.4 ± 14.7) completed the Cantab Solutions tests and CTIS-Profile tests at Training, Baseline and Week 1. The test batteries assessed psychomotor processing with Cantab Reaction Time (RTI), episodic memory with Cantab Paired Associates Learning (PAL) and executive function with Cantab Spatial Working Memory (SWM). Compared to the Cantab Solutions assessment, the CTIS-Profile tests were shorter (delivered in under 15minutes), were more difficult, and the graphics had a more modern look and feel. Additionally, the test instructions in CTIS-Profile were delivered using an automated and standardised voiceover, whereas the test instructions in Cantab Solutions were delivered by a rater who read a standardised script.

Results: There was a strong correspondence between performance assessed using CTIS-Profile and the Cantab Solutions for the Reaction Time (r_{71}= 0.82, p< 0.001), Paired Associates Learning (r_{71}= 0.68, p< 0.001) and Spatial Working Memory (r_{71}= 0.68, p< 0.001).
CTIS-Profile had high test-retest reliabilities between Baseline and Week 1 for Reaction Time (r_{71}= 0.81, p< 0.001), Paired Associates Learning (r_{71}= 0.85, p< 0.001) and Spatial Working Memory (r_{71}= 0.74, p< 0.001). The Cantab Solutions battery had high test-retest reliabilities between Baseline and Week 1 for CANTAB Reaction Time (r_{71}= 0.81, p< 0.001), CANTAB Paired Associates Learning (r_{71}= 0.79, p< 0.001) and CANTAB Spatial Working Memory (r_{71}= 0.88, p< 0.001).
In order to determine the test-retest reliability of CTIS-Profile in small sample sizes, 10 random sub-groups were generated with N=8. The test-re-test reliability was comparable to the results for the total group (N=71), with a median r value across the 10 randomly generated groups of 0.81 for RTI, 0.92 for PAL, and 0.83 for SWM.

Conclusions: Our results demonstrate that CTIS-Profile has good test-retest properties and that performance assessed using this battery corresponds well with performance assessed using the established Cantab Solutions technology. These findings provide support for the value of this technology in Phase 1 clinical trials. Future development and validation of test variants optimized for patient populations is required.

Disclosures: CH, LH, AB and PN are full-time employees at Cambridge Cognition. The authors report no conflicts of interest for this work.

8 Update: Are Current Touch Screen Devices Reliable and Valid for Timed Neurocognitive Testing: iPad mini vs. Standard Laptop

Abrahim S[1], Morgan D[2], Boyd A[2]

[1]UNC Eshelman School of Pharmacy, [2]CNS Vital Signs,LLC

Methodological Question Being Addressed: Reliability of touch screen devices for timed neurocognitive testing

Background: Tablets have become a common tool in today's society. They are also being used in doctors' offices and clinics as a light and simple substitute for computers, and even paper. It is important to understand that input lag is different for each touch screen devices and what effect the lag has on neurocognitive testing.

In a previous study poster presented at ISCTM "Are Touch Screen Devices Appropriate for Neurocognitive Testing" (Wei etal), the iPad 2 was tested against a standard laptop and was shown to register a statistically significant slower reaction time score as well as "miss" responses on the touch screen. For this study, to investigate whether the characteristics of the touchscreen had improved, we used the newer iPad mini.

Objective: The study was conducted to compare the response time performance and of a touch screen device to a keyboard-based device for the purposes of neurocognitive testing.

Methods: 14 young and healthy volunteers were randomized in this cross-over study to alternate taking the CNS Vital Signs (CNSVS) Continuous Performance Test (CPT) through the web browser interface on both a tablet and a laptop on 2 separate occasions split by an average of ~8 days.

Results: The difference of reaction time scores between tablet and laptop was significant. The paired t-test showed a significant difference between the reaction times of the iPad mini and the laptop. The iPad mini had a mean reaction time of 531.4 ms while the laptop mean reaction time was 462.3 ms ($p<0.0001$).

Conclusions: Despite improvements of the iPad mini over the iPad 2, use of touch screen-based devices for the use of neurocognitive testing reaction times is not as accurate as compared to keyboard-based devices.

9 Cognitive Function in Patients with Active Relapsing Multiple Sclerosis: Baseline Data from the SYNERGY Study of BIIB033, an Anti-LINGO-1 Antibody

Cadavid D[1], Phillips G[2], T Kaushik T[3], Xu L[1], Drulovic J[4]

[1]Biogen Idec, Inc., Cambridge, MA, [2]Biogen Idec, Inc., Weston, MA, [3]PanMedix, Inc., [4]University of Belgrade, Belgrade, Serbia

Methodological Question Being Addressed: is the reliability, validity, and usefulness of the MS-Cog as a measure of cognitive impairment in MS. MS has not had a measure of cognitive impairment that was designed for use in clinical trials. The MS-Cog is designed for use in clinical trials. This study demonstrates the MS-cog can differentiate various levels of cognitive impairment and it shows discriminate validity for known groups of MS patients.

Background: Cognitive deficits are frequently observed in patients with multiple sclerosis (MS). The SYNERGY Phase 2 study is comparing effects of BIIB033, an investigational anti-LINGO-1 antibody, versus placebo for maintaining cognitive function in participants with relapsing-remitting MS (RRMS) and secondary progressive MS (SPMS) when used with intramuscular interferon beta-1a (IM IFNß-1a).

Objectives: To report baseline results from the MS cognition assessment in SYNERGY.

Methods: SYNERGY is an ongoing, 84-week, international, multicenter, randomized, double-blind, dose-ranging study evaluating BIIB033 versus placebo when used concurrently with IM IFNß-1a in participants with active RRMS/SPMS. Participants receive once-weekly IM IFNß-1a and are randomized to intravenous infusions of BIIB033 (3, 10, 30, 100 mg/kg) or placebo for 72 weeks. Changes in cognition are measured using the MS-COG, a composite cognitive endpoint using Symbol Digit Modality Test (SDMT), Paced Auditory Serial Addition Test (PASAT) -3 and -2 seconds, Selective Reminding Test Total Learned (SRT-TL), SRT Delayed Recall (SRT-DR), Brief Visuospatial Memory Test-Revised-Total (BVMTR-T) and -Delayed Recall (BVMTR-D). Associations of MS-COG with baseline characteristics (MS type, sex, age, race, country of residence, degree of education, marital status, degree of independence, and work status) and with non-cognitive tests will be assessed. Baseline measurements were compared with published age matched healthy control data from a single US site.

Results: In an early data cut of 91 participants enrolled in 10 countries in Europe and North America: 57 (63%) were female, 57 (63%) had RRMS and 34 (37%) had SPMS. Compared with a reference MS population (Strober et al., *Mult*

Scler. 2009;15:1077-1084), mean (standard deviation) baseline z scores for RRMS vs SPMS, respectively, were: SDMT, -0.87 (1.32) vs -2.21 (1.11); PASAT-3, -0.09 (1.22) vs -0.81 (1.49); PASAT-2, -0.19 (1.30) vs -0.76 (1.46); SRT-TL, 0.03 (1.29) vs -1.26 (1.62); SRT-DR, -0.28 (1.30) vs -1.03 (1.38); BVMTR-T, -0.34 (1.44) vs -1.11 (1.57); BVMTR-D, -0.53 (1.70) vs -1.48 (2.06).

Conclusions: Mean baseline scores for all MS-COG components compared with a reference MS population revealed worse cognitive impairment in participants with SPMS than RRMS. Visual-based cognitive tests (SDMT and BVMTR) showed more impairment than auditory-based (PASAT and SRT) tests. Additional baseline results will be presented (N=396 participants).

10 Profiling Cognitive Dysfunction Across Therapeutic Areas using the Cambridge Neuropsychological Test Automated Battery (CANTAB)

Hermans L[1], Rock P[1], Housden C[1, 2], Riedel W[3], Blackwell A[1,2], Nathan PJ[1,2]

[1]Cambridge Cognition Ltd, [2]Department of Psychiatry, University of Cambridge, [3]Maastricht University

Methodological Question Being Addressed: The abstract/poster addresses the following methodological question. What is the effect sizes and profile of cognitive impairment in patients with Alzheimer's disease,, schizophrenia, ADHD and depression when assessed using the CANTAB battery

Introduction: CANTAB tests have been used extensively in cognition research in a variety of central nervous system (CNS) disorders including Alzheimer's disease (AD), mild cognitive impairment (MCI), schizophrenia, attention deficit hyperactivity disorder (ADHD), and depression. CANTAB tests measure performance in cognitive domains such as memory, executive function and attention. In addition, CANTAB tests are sensitive to detect cognitive impairments as well as effects of pharmacological and other treatments to achieve cognitive enhancement. Taking advantage of a large accumulated database across different studies over the past 10 years, the present study collated CANTAB data to demonstrate differential sensitivity relative to healthy controls in the cognitive profiles of patients with AD, schizophrenia, ADHD and depression.

Methods: Pooled data from our clinical trials database were analysed to investigate and compare the magnitude of cognitive deficits in patients with AD, schizophrenia and ADHD relative to healthy age-matched controls. Fourteen studies were included in the analyses, with a total of 985 AD patients; 554 schizophrenia patients; 783 ADHD patients; and 424 healthy control subjects. In addition, cognitive deficits in depression were examined by looking at pooled data from eight studies in 271 currently depressed patients and six studies in 168 remitted depressed patients and 178 controls.

Results: Distinct cognitive profiles were found across patient populations, with AD patients being particularly impaired on an episodic and visual memory test (Cohen's d = -1.55 and -1.64 respectively), while patients with schizophrenia were particularly impaired on tests of executive function (Cohen's d = -2.27). The ADHD patients mainly showed poorer performance on executive function and visual memory tasks relative to matched controls (Cohen's d = -1.79 and -1.17 respectively). Currently depressed and remitted patients showed significant moderate deficits relative to matched controls on attention, executive function and memory tasks (Cohen's d = −0.33 to −0.61).

Conclusions: This pooled data analysis shows that CANTAB is an effective tool for measuring and symptomatic profiling of cognitive dysfunction in neurodevelopmental, neuropsychiatric and neurodegenerative disorders. Importantly, the study showed that, while individual cognitive tests are sensitive tools to detect differences in a variety of cognitive domains in patients with CNS disorders relative to healthy controls, a battery of CANTAB tests is an informative tool which can be used to reveal patterns of cognitive dysfunction across key therapeutic areas. This paves the way for a dimensional approach to measuring cognitive impairments across diagnostic categories.

LH, CH, AD and PN are full-time employees of Cambridge Cognition.

11 The Association Between Amyloid Burden and Sub-Clinical Depression in Healthy Older Adults

Maruff P, Harrington KD, Gould E, Lim YY, Ellis KA, Ames D, Martins RN, Savage G, Szoeke C, Rowe C, Villemagne VL, Masters CL, Maruff P

AIBL Research Group

Methodological Question Being Addressed: Determination of the incidence of depression in individuals at risk for Alzheimer's disease

Background: Depression is an important risk factor for Alzheimer's Disease (AD) and therefore when present in non-

demented adults may reflect increased amyloid-β (with Aβ) deposition. The aim of this study was to investigate prospectively the extent to which high Aβ levels were associated with depressive symptomatology in healthy older adults over 54 months.

Method: Healthy adults (n=359) from the Australian Imaging Biomarkers and Lifestyle (AIBL) study cohort, who had undergone positron emission tomography (PET) neuroimaging for Aβ and who did not have depression at enrolment, were included in the present study. Participants were classified as either Aβ+ (n = 81) or Aβ- (n = 278) on the basis of PET neuroimaging. Level of depressive symptoms was assessed at baseline, 18, 36 and 54 months with the Geriatric Depression Scale (GDS) and the depression scale of the Hospital Anxiety and Depression Scale (HADS-D). Relationships between depressive symptoms and performance on the CogState battery were also assessed.

Results: No association between Aβ status and classification of depression was observed at the baseline, 18, and 36 month assessments. Aβ status was associated with decline in performance on the measures of working memory and visual memory over the 54 months. At the 54 month assessment the Aβ+ group was 4.5 times more likely to meet criteria for depression according to their GDS score. In the Aβ+ group, depressive symptoms were increased significantly at the 54 month assessment compared to baseline. No increase was observed in the Aβ- group. Change in depressive symptoms was modelled over all assessments using a linear mixed model and a composite depression score derived from the GDS and HADS-D. This model showed a small but statistically significant increase in depressive symptoms over time in the AB+ group (Cohen's d = 0.11). This increase was not related to decline in memory.

Conclusion: In older individuals, who were not depressed when recruited, high Aβ was associated with small but significant increases in depressive symptomatology across 54 months. While high Aβ was also associated with decline in working memory and visual memory, cognitive impairments were not related depressive symptoms. Consequently, Aβ+ individuals also showed an increased rate of development of depressive disorder. The results indicate that subtle increases in depressive symptoms are associated with increased Aβ deposition and suggest that these symptoms may in fact form a part of the pathological sequelae of AD.

12 Can Classification and Regression Trees Predict Treatment and Placebo Response and Non-Response?

Khan A[1,2,3], Opler M[1,4], Byrum J[1]

[1]ProPhase LLC, [2]Nathan S. Kline Institute for Psychiatric Research, [3]Manhattan Psychiatric Center, [4]New York University School of Medicine

Methodological Question Being Addressed: Classification and regression trees (CART) have been successfully utilized in a number of therapeutic areas to categorize disease state and to determine the likelihood of future events. The ability to predict improvement of outcomes in patients with schizophrenia at the end of a clinical trial would enable researchers to tailor treatment early during the course of the clinical trial. Can the prognostic significance of rating scale data measured at baseline and during the first 2 weeks of participation in an antipsychotic clinical trial to predict response at the end of the trial?

Methods: Subjects were patients with schizophrenia who were in one of two, double-blind, randomized treatment trials conducted over an 8 to 12 week period: for both studies, patients received an atypical antipsychotic or placebo. The following patient groups were defined: 1. Early Improvers (N = 183; Treatment = 94; Placebo = 89): reduction in PANSS total score of ≥ 20% compared with baseline within the first 2 weeks of treatment. 2. Stable Improvers (N = 286; Treatment = 119; Placebo = 167): reduction in PANSS total score of ≥ 20% compared with baseline within the first 3 weeks of treatment. 3. Treatment Responders at endpoint (n = 254): assigned to the treatment group and having a reduction in PANSS total score of ≥ 20% at endpoint. 4. Placebo Responders at endpoint (n = 195): assigned to the placebo group and having a reduction in PANSS total score of ≥ 20% at endpoint. 5. Placebo Non-Responders at endpoint (n = 157): assigned to the placebo group and did not have a reduction in PANSS total score of ≥ 20%. 6. Treatment Non-Responders at endpoint (n = 222): assigned to the treatment group and did not have a reduction in PANSS total score of ≥ 20%.

Results: Of the 476 patients randomly assigned to treatment, 254 patients (53.36%) responded to medication. Response rates to medication between the two studies were not statistically different. The placebo response rate was 51.05% (195 of 382 patients). As expected, placebo responders were significantly different from non-responders on final PANSS score ($F_{(2, 350)}$=32.342, p<0.001; SE 1.825). Both responder groups had comparable clinical outcomes (final PANSS total score 62.32 ± 15.68 for placebo and 63.01 ± 14.25 and treatment); both non-responder groups also showed equivalent clinical endpoints (final PANSS total score 87.67± 14.53 for placebo and 84.37 ± 14.27 for treatment). A score of 3 to 7 on Item P03 (Hallucinatory Behavior) at Visit 2 can significantly predict a non-response to treatment at endpoint, whereas a score of 1 (Absent) can significantly predict treatment response at endpoint. A score of 4 to 7 on Item G14 (Poor Impulse

Control) at Visit 2 can significantly predict a response to treatment at endpoint, whereas a score of 1 to 3 can predict treatment non-response at endpoint.

Conclusions: These findings represent patterns of early detection of response in trials, and can lead to the development of algorithms that may allow investigators to develop cost effective study designs.

13 Not All Comedications are the Same. Careful Consideration of the Impact of Different Comedications can Substantially Improve Success in CNS Clinical Trials

Geerts H, Spiros A, Carr R

In Silico Biosciences

Methodological Question Being Addressed: How can we better simulate the effects of comedication on the dose-response of the investigative drug in virtual patient trials to optimize clinical CNS trial design?

Introduction: Success rates in CNS Clinical Trials have not improved despite increased understanding of the disease neurobiology. While this failure rate might be due to the wrong target being selected, the wrong patient population being included, insufficient target engagement or mismatch between rodent and human pharmacology, we will focus on the interference of comedications in clinical trials. Many approved CNS active drugs do have a rich pharmacology and can affect neuronal circuits in non-linear ways.

Methods: We use an advanced version of a computer-based Quantitative Systems Pharmacology (QSP) platform, a mechanism-based computer model of the relevant humanized cortical networks that has been developed for clinical readouts in psychiatry and neurology and has been calibrated with group average clinical data and has been able to blindly and correctly predict an unexpected clinical outcome in schizophrenia and Alzheimer's disease (AD). By implementing the human pharmacology of known comedications, we can simulate the impact of comedications on dose-responses for a number of single targets.

Results: In a blinded prospective prediction, histamine H3 antagonism improves modestly the outcome in a QSP model for cognitive impairment for schizophrenia (CIAS) when added to Zyprexa and Abilify, but much less when added to Haldol or Seroquel and worsens cognition when added to Risperdal and Clozaril. In a QSP model for AD H3 antagonism will modestly improve cognition in mild-to-moderate AD patients (with smaller effect than donepezil) and the effect will completely disappear at 26 weeks. These results will be compared to clinical publications on ABT-266 and MK3134. Similarly, blinded prediction suggest that the dose-response of an alpha-7 nicotinic receptor agonist in a QSP model of CIAS is substantially affected by the type of antipsychotic as comedication; Risperdal, Zyprexa, Seroquel and Clozaril increase the level of free ACh to different degrees, while Haldol and Abilify do not have any effect. The effect of an alpha4beta2 nAChR modulator on cognition is also modulated by smoking state. These outcomes will be compared to clinical data on EVP6124, ABT-126, TC5619 and ispronicline. A QSP model of negative symptoms associated with schizophrenia suggests that the non-monotonic dose-response of a glycine transporter inhibitor is differentially affected by antipsychotics, anticholinergics and benzodiazepines, possibly explaining the reduced effect size of bitopertin in large Phase III trials.

Conclusions: These results suggest that each CNS active comedication has a complex pharmacology and can affect the dose-response of single target drugs differentially. The study also suggests that the traditional concept of chlorpromazine equivalents for antipsychotic drug treatment has severe limitations, as the particular drug pharmacology fingerprint for each drug against human receptors has a very unique impact. Simulating virtual patients using a QSP-based approach is a way to identify possible comedications that need to be excluded or balanced across different treatment arms in order to significantly increase the probability of success.

Disclosures: All authors are employees of In Silico Biosciences, a company that provides Quantitative Systems Pharmacology services for Pharma R&D in CNS Indications.

14 Detecting Declining Financial Skills in Preclinical Alzheimer's Disease: the Financial Capacity Instrument - Short Form

Marson DC[1,2], Triebel KL[1,2], Martin RC[1,2], Edwards K[3], Pankratz VS[3], Petersen RC[4,5]

[1]Department of Neurology, University of Alabama, [2]Alzheimer's Disease Center, University of Alabama, [3]Division of Biomedical Statistics and Informatics, Mayo Clinic and Foundation, [4]Department of Neurology, Mayo Clinic and Foundation, [5]Mayo Clinic Alzheimer's Disease Research Center, Mayo Clinic and Foundation

Methodological Question Being Addressed: Can declining financial skills be detected in persons with preclinical

Alzheimer's disease (AD) using a brief performance measure of financial capacity?

Introduction: Very little is known about possible changes in functional abilities in persons in the preclinical stage of AD. As currently conceptualized, functional impairment is the final and somewhat remote outcome of a cascade of preceding pathophysiological and clinical events characterizing AD: amyloid deposition, neurodegenerative cellular and pathway change, structural atrophy, and cognitive decline. However, the intriguing possibility that detectable functional change may actually commence in the preclinical phase of AD has recently been raised by some investigators[1,2]. For example, prior research by our group[3,4] and others[5,6] has shown that higher-order functional skills are impaired in mild cognitive impairment (MCI) and decline over time. Financial capacity in particular is a functional skill particularly sensitive to both MCI and mild AD, which raises the possibility that measurable financial decline may also occur in preclinical AD. A critical factor here is the sensitivity of the functional measure employed. Informant report measures commonly used to characterize functional decline in AD type dementia lack sensitivity to detect presumably subtle functional decline in cognitively normal persons with preclinical AD. In contrast, performance based measures permit finely grained measurement of function in terms of performance and completion time.

Methods: We recruited 186 cognitively normal, community-dwelling older adults age 70+ who were participants in the Mayo Clinic Study of Aging (MCSA) in Olmsted County, Minnesota. All participants were classified as cognitively normal controls based on MCSA diagnostic workup. Participants underwent 11C PiB amyloid imaging and were administered a brief performance measure of financial skills (Financial Capacity Instrument-Short Form (FCI-SF)) that evaluated both task performance and time to completion. Amyloid imaging resulted in subsamples of amyloid positive (A+) (n=66) and amyloid negative (A-) (n=120) controls. We used the Rank Sum test to compare A+ and A- groups on FCI-SF performance and task completion time variables. Logistic regression was used to examine how well the FCI-SF performance and timing variables predicted participant amyloid status (positive/negative) after controlling for demographic and cognitive variables.

Results: The FCI-SF detected both timing and performance differences between the A+ and A- groups. A+ participants performed below A- participants on FCI-SF total score (p=0.0318), and on items tapping specific complex financial skills. A+ participants were also slower completing two checkbook tasks (composite time) (p=0.0046), and all FCI-SF timed tasks combined (composite time) (p=0.0012). Both FCI-SF composite time variables remained significant predictors of amyloid status in a model controlling for age, gender, and cognitive performance.

Conclusions: Financial decline is detectable in cognitively normal persons with presumed preclinical AD using a brief performance based measure. Performance declines on complex financial tasks, and slower financial task completion times, likely represent measurable very early functional changes in preclinical AD. Brief performance-based measures of complex everyday function should be considered for AD prevention trials.

Disclosures: Dr. Marson is funded by the NIA and the National Endowment for Financial Education.

References:

1. Marshall GA, Amariglio RE, Sperling RA, Rentz DM. Activities of daily living: Where do they fit in the diagnosis of Alzheimer's disease? *Neurodegenerative Disease Management*. 2012. 2(5): 483-491.
2. Farias ST, Chou E, Harvey DJ, et al. Longitudinal trajectories of everyday function by diagnostic status. Psychol Aging. Dec 2013;28(4):1070-1075.
3. Griffith HR, Belue K, Sicola A, et al. Impaired financial abilities in mild cognitive impairment: a direct assessment approach. *Neurology.* 2003;60(3):449-457.
4. Triebel K, Martin R, Griffith HR, et al. Declining financial capacity in mild cognitive impairment: A one-year longitudinal study. *Neurology.* 2009;73:928-934.
5. Jefferson AL, Byerly LK, Vanderhill S, et al. Characterization of activities of daily living in individuals with mild cognitive impairment. *Am J Geriatr Psychiatry.* May 2008;16(5):375-383.
6. Tomaszewski Farias S., Mungas D, Reed, BR, Harvey D, Cahn-Weiner D, DeCarli, C. (2006). MCI is associated with deficits in everyday functioning. *Alzheimer's Disease and Associated Disorders*, 20: 217-223.

15 Dosing In Lurasidone Bipolar Depression Studies

Frye M[1], Tsai J[2], PhD, Kroger H[2], Pikalov A[2], Loebel A[2]

[1] Mayo Clinic Department of Psychiatry, [2] Sunovion Pharmaceuticals Inc.

Methodological Question Being Addressed: What dosage-specific information can be gleaned from flexible-dose clinical studies?

Introduction: Flexible-dose studies are not specifically designed to provide dose-response information. Lurasidone was approved for the treatment of bipolar depression based on a fixed-flexible dose monotherapy study, and a flexible-dose adjunctive therapy study. This analysis examined patterns of dose utilization, tolerability, and patient factors potentially associated with dose escalation.

Methods: Data were from two 6-week placebo-controlled studies and one 24-week extension study. In the monotherapy study, patients received lurasidone 20-60mg/d or 80-120mg/d (initially titrated from 20mg to 80mg). In the adjunctive study, patients received lurasidone 20-120mg/d (initially titrated from 20mg/d to 60mg/d) with Li or VPA. In the extension, all patients received monotherapy or adjunctive lurasidone 20-120mg/d starting at 60mg/d.

Results: ITT analysis of dosing in the monotherapy study showed that 49% of patients in both the 20-60mg and the 80-120mg dose groups had modal doses of 20mg and 80mg, respectively. Smaller proportions of patients had modal doses of 40mg (26%) and 60mg (25%) in the 20-60mg group, and 100mg (34%) and 120mg (17%) in the 80-120mg group. Distribution of terminal doses was similar (20mg/40%, 40mg/28%, 60mg/32% in the 20-60mg group; 80mg/44%, 100mg/30%, and 120mg/26% in the 80-120mg group). Comparable proportions (33-38%) of patients had maximum doses of 20-100mg during study participation; somewhat fewer (27%) had a maximum dose of 120mg (ITT). Analysis of dosing in the adjunctive study showed that the largest proportion (38%) of patients had a modal dose of 60mg. Smaller proportions had higher modal doses (80mg/28%, 100mg/17%, 120mg/10%); few patients had lower modal doses (20mg/3%, 40mg/5%). Distribution of terminal and maximum doses showed similar patterns. No significant differences in demographics or baseline characteristics were seen between patients who stayed at initial target dose and those with dose escalation. In the open-label extension study, in which all patients received lurasidone 60mg without dose titration, treatment-emergent adverse events (TEAEs) reported in ≥5% of patients transitioning from acute placebo to lurasidone 60mg during the 1st 28 days of extension treatment were akathisia (6%, monotherapy), nausea (8%, adjunctive), and somnolence (5%, adjunctive). Incidences of TEAEs during the 1st 28 days of adjunctive lurasidone treatment during the extension study were comparable to those during the 1st 28 days of acute treatment.

Conclusions: Lurasidone dose utilization patterns did not reveal specific preferred doses. A previous exposure-response modeling analysis of these studies (Chapel et al, APA Poster, 2014) found that higher doses are associated with greater efficacy. This analysis found that patients treated in a fixed-flexible dose monotherapy study or a fully flexible dose adjunctive therapy study tended to under-utilize the higher permissible doses. The reason for this is uncertain, but warrants further analysis. The current results illustrate the potential methodological trade-offs that occur when choosing a flexible-dose study design, which has a high degree of external validity and generalizability to clinical settings, and a fixed-dose study design, which has less clinical generalizability, but provides dose-response data.

16 CONNECT: Cognitive Behavioral Therapy Focused on Social Engagement – Rationale and Approach

Wu L, Velthorst E, van der Gaag M, Reichenberg A, Frangou S

Department of Psychiatry, Icahn School of Medicine at Mount Sinai

Methodological Question Being Addressed: Schizophrenia, with a lifetime prevalence of 0.7%, is one of the most disabling and costly disorders worldwide. There is growing consensus that targeting negative symptoms [the diminution of normal behaviors and functions] is essential for preserving social participation, thereby reducing its impact on quality of life and yearly healthcare costs. Attempts to ameliorate negative symptoms with prevailing antipsychotic medications have had only limited effect. In addition, recent trials investigating the efficacy of Cognitive Behavioral Therapies (CBT) for the treatment of schizophrenia show that they are not as effective as previously thought for reducing negative symptoms. Is a targeted CBT intervention focused on social withdrawal effective in improving social engagement?

Introduction (aims): Recent evidence suggests that negative symptoms can at least be partly accounted for by dysfunctional and defeatist beliefs (e.g., 'if you cannot do something well, there is little point in doing it at all'). Based on this premise, Aaron T. Beck, founder of Cognitive Therapy, and colleagues have developed and investigated a new approach that targets negative symptoms by challenging those beliefs (Grant et al., 2012). Results of research investigating this form of Cognitive Behavioral Therapy (CBT) targeted at the treatment of negative symptoms can be highly successful in establishing clinically meaningful improvements. The number of sessions needed to yield significant effects however, is considerable (mean= 51), a number not feasible in most medical practices. We therefore aimed to establish a shortened version of this targeted CBT by specifically focusing on the negative symptom that is found to be most debilitating and likely most amenable to change — social withdrawal (Velthorst, Reichenberg et al., in prep).

Methods: CONNECT will be based on the model of Beck and colleagues. Pilot data from a shortened protocol in the Netherlands has shown promising results (decrease overall PANSS negative sum score d= .40 in 20 sessions).

The CONNECT protocol will include:

- 16 individual weekly sessions
- Involvement of caregivers
- Social engagement as the main focus

The core assumption of our therapy will be that modifying dysfunctional beliefs will lead to increased social engagement in individuals with prominent negative symptoms – primarily social withdrawal. Its primary focus is to help participants overcome core elements of social withdrawal, including social isolation and inactivity, thereby improving quality of life. Dysfunctional beliefs that have been found to result in social withdrawal and inactivity would be modified.

Conclusions: Based on epidemiological studies, social engagement should be a primary target for negative symptom interventions in schizophrenia. The focus of our proposed new psychological treatment approach, based on components found to be effective in pilot studies, may offer new insights in how to overcome this disabling symptom of schizophrenia and related disorders.

17 Identifying Intervention Targets for Remediation of Social Deficits in Schizophrenia

Velthorst E, Wu L, Kapara O, Goldberg S, Weiser M, Frangou S, Reichenberg A

Methodological Question Being Addressed: Schizophrenia is a severe mental disorder that affects approximately 1% of the population. While mostly known for its psychotic features and neuropsychological deficits, the disorder is also typically characterized by profound and persistent social impairments. Social functioning, however, is a complex trait that constitutes multiple dimensions. Although mostly considered together, it may be that not all of the dimensions of social functioning are equally impaired in schizophrenia.

Introduction (aims): Traditionally, (mostly retrospective) studies examining social impairment in schizophrenia utilize broad definitions, lumping together various social constructs into one social functioning score. We were interested in seeing whether certain social dimensions should be prioritized in treatment over others.

Methods: We examined the severity and course of impairment in three key dimensions of social functioning: — Individual Autonomy (the autonomy in taking action and making decisions), Social Engagement (the ability to make and maintain friends) and Functioning in Structured Environments (such as school and work). The study sample consisted of 3,929 individuals hospitalized for schizophrenia. Differences in social functioning between the cases and the general population were examined using univariate analyses of covariance (ANCOVA) models.

Results: As can be seen in **Figure 1**, developmental trajectories differed between the three components of social functioning. Whereas mild impairments in social engagement and functioning in structured environments were already recognizable up to 15 years prior to first admission, individual autonomy seemed relatively preserved until the years immediately prior to first admission. For social engagement ($F=115.33$, $p< .0001$) and individual autonomy ($F=20.358$, $p< .0001$), differences between cases and the general population were greater, the shorter the time between testing and first hospitalization. Trend analysis revealed no significant decline in the ability to function well in structured environments ($F=1.813$, $p=.171$). In addition, the social dimensions Social Engagement and Functioning in Structured Environments, but not Individual Autonomy, were found to be significantly related to higher risks of hospitalization for schizophrenia [Social Engagement (effect size= .47, $p< .0001$); Functioning in Structured Environments (effect size = .19; $p< .0001$); Individual Autonomy (effect size = .035; $p= .852$)].

Discussion: Our findings suggest that treatment targeting social engagement specifically is warranted in schizophrenia patients. Recent evidence suggests that social disengagement can at least be partly accounted for by dysfunctional and defeatist beliefs (e.g., 'if you cannot do something well, there is little point in doing it at all'). Based on promising pilot data from a Cognitive Behavioral Therapy intervention in the Netherlands targeting certain beliefs in relation to negative symptoms (Staring et al., 2013), we propose that this new psychological treatment approach might be useful for tackling the most debilitating social domain of schizophrenia.

18 Neurocognitive Predictors of Response in Treatment Resistant Depression (TRD) with Deep Brain Stimulation (DBS) Therapy

McInerney SJ[1,2], McNeely H[3], Geraci J[1,2], Giacobbe P[1,2], Rizvi SJ [1,2],Cyriac A[1], Mayberg H[4], Lozano AM[2,5] Kennedy,SH[1,2,6,7]

[1]Dept. of Psychiatry, University Health Network (UHN), Toronto, Canada, [2]Faculty of Medicine, University of Toronto, Canada, [3]Dept. of Neuroscience & Behavior, McMaster University, Canada, [4]Psychiatry and Behavioral Sciences, Emory University, [5]Toronto Western Research Institute, UHN, Toronto, Canada, [6]St Michael's Hospital, Toronto, Ontario, Canada, [7]Arthur Sommer-Rotenberg Chair in Suicide Studies, St Michael's Hospital, Toronto, Ontario, Canada

Methodological Question Being Addressed: Cognitive impairment is highly prevalent in TRD and is a source of burden to the patient population. Neurostimulatory treatments such as deep brain stimulation are under scrutiny to ensure they provide effective treatment without further cognitive compromise. This study of 20 patients who underwent DBS had longitudinal cognitive assessments to investigate how the treatment impacted on their cognitive status.

Introduction: Deep Brain Stimulation (DBS) is a neurosurgical intervention for TRD with a number of positive efficacy studies but thus far there are limited outcome studies assessing cognition. The objective of this study is to examine cognitive performance, pre-surgery and after 12 months of continuous DBS stimulation. The primary purpose of the study is to identify baseline cognitive predictors of treatment response to DBS.

Methods: 20 TRD patients underwent DBS surgery to the Subcallosal Cingulate Gyrus (SCG). All patients met criteria for TRD and were in a current major depressive episode (MDE) with a minimum score of 20 on the HDRS-17. Response was defined as 50% or greater reduction in the HDRS-17 twelve months after surgery. A standard neuropsychological battery assessing frontal lobe and general cognitive abilities was administered at baseline and 12 month follow up.

Results: There was a significant reduction in HDRS-17 scores from baseline to follow up and 55% (n=11) of patients were responders. At baseline there was impairment (Z= -1 SD below the mean) in information processing speed and executive functioning while psychomotor speed, motor dexterity, verbal memory, phonemic fluency, and visual speeded attention were within 1SD below the mean. There was no statistically significant deterioration in cognitive functioning on any of the tests at the follow up assessment period. There were significant improvements ($P \geq 0.05$) in executive functioning at follow up. The only neuropsychological variable associated with change in depression score over the 12 month period was phonemic fluency. Change in phonemic fluency was positively correlated with change in depression rating, $r(13) = -0.63$, $p < 0.01$). Processing speed as measured by the Finger Tap Test, bilaterally was predictive of treatment response.

Conclusions: SCG DBS had no deleterious effects on cognition and the trend towards cognitive enhancement was independent of mood. This study suggests that psychomotor speed may be a useful predictor of response to DBS treatment.

Disclosures: Neurostimulation devices were provided by Medtronic. No additional funding was provided.

19 Small Phase 1b Trial Demonstrates Utility of Signal Detection Strategies to Detect Clinical and QEEG Effects in Major Depressive Disorder Patients Receiving NSI-189, a Neurogenic Compound

Ereshefsky L[1], English BA. [1,2], Johnstone J[3], Johe K[4], Gertsik L[5], Sherman M[1], Fava M[6], Freeman M[6]

[1]PAREXEL International, Los Angeles Early Phase Unit, [2]Vanderbilt University, [3]QMetrx Inc., [4]Neuralstem Inc.
[5]California Clinical Trials Medical Group, [6]Massachusetts General Hospital, Harvard Medical School

Methodological Question Being Addressed: How can small, potentially underpowered studies for CNS compounds in Phase I studies reliably detect target engagement and proof of relevant effect? Specifically what methodologies and statistical approaches improve signal detection for relevant pharmacodynamic (QEEG) observations? Given the fast pace and PK/Safety considerations of a Phase I study, how can patients with the target illness and relevant behavioral readouts be incorporated into the trials design?

Introduction: Phase I clinical trials designed to add pharmacodynamic measurements that demonstrate proof of mechanism and target engagement increase the likelihood of a successful Phase II outcome. Incorporating well characterized patients with clinical symptomatology into Phase IB trials further enhances estimation of drug effect and useful dose ranges. We illustrate the application of this approach using the Phase IB trial results of NSI-189, a possible treatment of Major Depressive Disorder. In typically underpowered (for efficacy) small Phase I studies, the use of QEEG

measurements can provide a sensitive readout for antidepressant CNS activity supporting progression to POC.

Methods: This double-blind, RPCT with 3 ascending cohorts; patients (n=24) enrolled had confirmed major depressive disorder (MDD) using remote SAFER interview. Patients were randomized to receive NSI-189 40 mg QD, BID or TID or placebo (PBO) for 28 days. QEEG measurements for antidepressant response (high frequency alpha in the left posterior temporal region) were obtained 6 hrs post-dose on Day 14 and 28. Efficacy assessments included the Montgomery-Asberg Depression Rating Scale (MADRS), the Clinician Global Impression – Improvement (CGI-I), the Symptoms of Depression Questionnaire (SDQ) and the Cognitive and Physical Functioning Questionnaire (CPFQ).

Results: QEEG antidepressant signal preceded changes in the patient's depressive and cognitive symptoms. A possibly meaningful reduction in depressive and cognitive symptoms estimated by effect size was observed: MADRS (Day 28 d=0.95), SDQ (Day 28 d=0.9), and CPFQ (Day 28 d=0.94) assessment scales was observed for patients on NSI-189 compared with placebo. The quantitative EEG analyses showed increased high frequency alpha with NSI-189 treatment (change from baseline to Day 14 and to Day 28) in contrast to decreased high frequency alpha or less change with placebo. In univariate tests in the Alpha 1 band sensitive dose effects are observed, suggestive of a curvilinear relationship. The NSI-189 effect is most prominent in the left posterior temporal (T5) (t=2.45, p=0.0247) and left parietal regions (P3) (t=3.31, p=0.004) and is similar when comparing QEEG at baseline to either Day 14 or Day 28 assessments.

Conclusions: These findings demonstrate the utility of QEEG in small Phase I trials using a priori hypothesis based on antidepressant research from larger clinical trials. NSI-189 QEEG effects at two weeks on brain electrical activity in patients with major depressive disorder suggests that short Phase I trials can provide a read out on possible efficacy and dose. Clinical outcomes require longer and larger trials to achieve significant results. Patient adjudication in Phase I studies is rare, but can enhance signal detection of relevant efficacy and surrogate markers. Methodologically, small Phase Ib patient studies can yield relevant data to contribute to go/no-go decisions. These preliminary findings demonstrate the benefit of incorporating electrophysiologic and clinical measures in a Ph1b study and support the further development of NSI-189 as a potential neurogenesis activator for the treatment of depression. Limitations of this study include the small sample size, and the sequential escalating dose structure of a Phase 1 trial.

Disclosures: All authors' institutions received funding from Neuralstem Inc. Each of the author's institutions engages in clinical trials work supported by numerous Pharmaceutical and biotechnology companies.

20 Retrospective Analysis of a Failed Depression Trial: What Can We Say About the Impact of Poor Rating Practice Using Data-monitoring Algorithms?

Yavorsky C[1], DiClemente G[1], Wolanski, K[1]

[1]*Cronos CCS*

Methodological Question Being Addressed: Can data-monitoring algorithms/rules be used retrospectively to determine the contribution of rater error to a failed completed trial?

Introduction: Increasingly academic and industry researchers have been using algorithms – pre-set rules of association – to evaluate study data for risks associated with rater error. The essential idea is that by looking at expected relationships between scale items as well as the clinical likelihood of change across visits for a given indication, mistaken or idiosyncratic use of a scale can be detected and ameliorated within the study. There have been efforts to systematize the approach in neuroscience clinical trials due to the high rate of failure and inefficient use of patients (Brosteanu et al, 2009). In this study we used this systematic technique to retrospectively analyze the results of a failed Phase III depression trial with a compound of known efficacy. We sought to determine to what extent rater error in the clinical application of the Montgomery-Åsberg Depression Rating Scale (MADRS) may have been responsible for the compound's failure to separate from placebo.

Methods: Unblinded data was obtained from the sponsor that included MADRS individual item and total scores for each of 14 total visits. Analysis was conducted using SPSS 17.0 for Windows. Data monitoring algorithms that compare expected relationships across items within and across visits were applied. Based on this, the data was partitioned as belonging to "probable" and "improbable" based on the categorical risk severity present in the data. Both partitions were analyzed to determine if there was any difference in separation from placebo.

Results: There were 336 patients included in the analysis with 178 randomized to placebo and 158 randomized to active compound. Pairwise comparison analysis (Tukey's test) indicated that data from poor performing (n=79; 41 placebo, 38 active compound) sites showed no separation from placebo and was not significant at $p = .384$. Data from well performing sites (n=257; 137 placebo, 120 active compound) showed clear separation from placebo and was significant

at $p = .044$.

Conclusions: Data-monitoring algorithms appear to be useful in the post-hoc analysis of data to determine what the contribution of rater error was to a failed trial. The fact that poorly performing sites (as given by the presence of significant assessment violations) had no ability of the raters to distinguish between placebo and active compound ($p = .384$) and the well performing sites were clearly able to distinguish with a p value of .044 seems to suggest that the contribution of error in the clinical application of the MADRS was a component in the failure of this trial. With the contribution of placebo response to failed depression trials well-known (e.g., Dago & Quitkin, 1995) it seems ever more essential to detect problematic sites during the conduct of a clinical trial.

21 The Positive and Negative Syndrome Scale Revisited: Seeking Dimensions that Span Schizoaffective, Bipolar and Schizophrenia Diagnoses

Anderson A[1], Chung H[2], Li Q[3], Wilcox M[3], Salvadore G[3], Reise S[4], Bilder R[5]

[1]UCLA, Dept of Psychiatry & Biobehavioral Sciences, [2]Janssen Scientific Affairs, [3]Janssen Research & Development, [4]UCLA, Psychology, [5]Dept of Psychiatry & Biobehavioral Sciences

Methodological Question Being Addressed: Does a bifactor model of the PANSS better explain the common disease structure of schizophrenia, schizoaffective and bipolar disorder than a traditional five factor model?

Aims: Although both the DSM-V and RDoC specify dimensional approaches to measuring mental illness, the shared dimensions (factors) behind common diseases such as schizophrenia (SZ), bipolar (BP), and schizoaffective (SA) disorder are unknown. Given the common genetic variants identified for those different mental illnesses (COMT, MTHFR) and the similarity of clinical presentations during acute episodes, the PANSS may be able to define a common dimension across those disorders. Here, we assess whether the five factors defined by the PANSS in Schizophrenia could also define the same dimensions in Bipolar disorder and Schizoaffective disorder, and evaluate whether a bifactor model, which includes a general mental illness dimension, could more effectively define their shared architecture by providing a shared dimension for all symptoms.

Methods: The PANSS dataset consisted of baseline ratings from 16 Janssen Research & Development, LLC sponsored clinical trials, involving 5,261 patients (SZ, n=3811; BP, n=858; SA, n=592) who had participated in the respective study for at least 21 days. First, we assessed whether established PANSS factor models fit well within Schizophrenia using confirmatory factor analysis. We then extracted the unique factor structures found within Schizophrenia, Schizoaffective and Bipolar disorder with exploratory factor analysis. Finally, we assessed whether a bifactor model, which proposes five specific dimensions as well as a common mental illness domain, would better fit all patients pooled across diagnoses than a five-factor model or a unidimensional model, using the omegaSem modeling approach as implemented in the psych R package.

Results: Published factor models of the PANSS replicated poorly within the Schizophrenia patient group in our confirmatory factor analysis. The exploratory factor analyses within diagnoses showed different factors: the original Excited and Anxiety factors were the most consistent across diagnoses, while the Positive and Disorganized factors differed the most. The bifactor model provided superior fit to all observations compared to either a unidimensional or five-factor model: the Omega Hierarchical ($\omega_{\square}$) statistic (one measure of reliability of the construct) from a confirmatory model using Structural Equation Modeling (sem) = 0.71, while the Omega Total (ω_t) from a confirmatory model using sem = 0.93. This indicates that the general factor explains much of the variance, but the five specific factors contribute as well to explaining the total variance in scores.

Conclusions: The original 5 specific factors defined by the PANSS may not optimally model a common symptom spectrum across Bipolar, Schizophrenia, and Schizoaffective disorder. We propose that the bifactor model provides a more general structure, which includes a common "mental illness" dimension across bipolar, schizophrenia, and schizoaffective disorder that is measured by the PANSS for all patients.

22 The Impact of Assumptions About the Missing Data Mechanism on the Statistical Assessment of Functional Outcome Data in a Randomized Relapse-Prevention Trial

Turkoz I[1], Scharfstein D[2,3], McDermott A[2,3], Olson W[4], Ascher S[1], Fu DJ[4], Alphs L[4]

[1]Janssen Research & Development, [2]Saphire Consulting Inc, [3]John Hopkins University, [4]Janssen Scientific Affairs LLC

Methodological Question Being Addressed: How should researchers evaluate the sensitivity to assumptions about the missing data mechanism for functional outcome data in a randomized relapse-prevention trial?

Introduction: SCA-3004 was a randomized, 15-month, double-blind, placebo-controlled study designed to evaluate the ability of a once-monthly long-acting injectable (LAI) atypical antipsychotic to delay relapse in patients with schizoaffective disorder. A key secondary objective was determining whether function as measured by the Personal and Social Performance (PSP) scale is better maintained with continued treatment vs placebo. Ethical requirements obliged investigators to look for first signs of relapse and intervene to prevent adverse outcomes, resulting in considerable missing PSP data. This analysis examined the impact of assumptions about the missing data mechanism when evaluating effects of treatment on function in the counterfactual setting, where all patients would continue their assigned therapy during the full 15-month relapse-prevention phase.

Methods: Stabilized patients were randomized 1:1 to receive LAI or placebo injections at baseline (visit 0) and every 28 days thereafter (visits 1–15). The final visit (visit 16) was scheduled 28 days after last scheduled injection. 164 patients were randomized to LAI and 170 to placebo; 60% and 38% of patients in LAI and placebo arms, respectively, completed all 16 visits.

To estimate the effect of interest, a class of untestable assumptions about the nature of the discontinuation process was introduced. These assumptions are indexed by a treatment-specific sensitivity analysis parameter (α), where $\alpha=0$ corresponds to the missing at random (MAR) assumption.

MAR states that, among study patients at visit k, the risk of discontinuing before visit $k+1$ depends only on the observable PSP history through visit k. In general, α quantifies the influence (above and beyond the observable PSP history) of the potentially unobserved outcome at visit $k+1$ on the risk of discontinuation. Positing $\alpha>0$ (<0) is equivalent to assuming that patients with better (worse) PSP scores at visit $k+1$ are at higher risk for discontinuation. Higher absolute values of α represent greater deviations from MAR. In this setting, it is most plausible to assume that $\alpha<0$ because discontinuations are more likely to occur with worse PSP scores; α was varied from –10 to 10.

Results: Under MAR, the estimated mean PSP treatment difference at visit 16 is –4.8 (95% CI: –10.9, 0.1); the estimated mean is 69.6 (95% CI: 62.1, 73.9) for placebo and 74.4 (95% CI: 72.3, 76.2) for LAI. Although the treatment difference favors LAI, the difference was not statistically significant. The estimated treatment effects for various combinations of the treatment-specific sensitivity analysis parameters will be presented using a contour plot. As the treatment specific alphas vary over negative values, the estimated treatment effect becomes statistically significant in regions that are consistent with observed PSP data and the plausible effects of lower values of PSP on the discontinuation rate.

Conclusions: Although this analysis does not conclusively demonstrate that LAI is superior to placebo with regard to maintaining function, statistically significant effects are observed when it is plausibly assumed that patients with worse PSP scores are more likely to discontinue treatment $\alpha<0$.

Disclosures: Janssen Scientific Affairs, LLC. One or more authors report potential conflicts, which are described in the program.

References: Morosini PL, Magliano L, Brambilla L, et al. Development, reliability and acceptability of a new version of the DSM-IV social and occupational functioning assessment scale (SOFAS) to assess routine social functioning. *Acta Psychiatr Scand*. 2000;101:323–329.

Nasrallah H, Morosini P, Gagnon DD. Reliability, validity and ability to detect change of the Personal and Social Performance scale in patients with stable schizophrenia. *Psychiatry Res*. 2008;161:213–224.

23 Identification of CDRS-R Items of Particular Challenge to Raters in Child and Adolescent Depression Clinical Trials

Busner J[1,2], Findling RL[3], Robb A[4]

[1]Penn State College of Medicine, Department of Psychiatry, [2]Bracket, [3]Johns Hopkins Medicine, Department of Psychiatry, [4]Children's National Medical Center, Department of Psychology and Behavioral Health

Methodological Question Being Addressed: Can particularly challenging Child Depression Rating Scale, Revised (CDRS-R) items in child/adolescent depression trials be identified as a first step toward improving ratings precision and reducing interrater variance?

Introduction (Aims): 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). In a previous study presented at ISCTM, we examined ratings variability on the Positive and Negative Syndrome Scale (PANSS) to identify items that posed particular difficulty for raters in pediatric trials (Busner, Daniel, Findling, 2013). This information has been used subsequently to guide training efforts. In the present investigation, we applied the same methodology to the most often used efficacy scale in pediatric depression clinical trials, the Children's Depression Rating Scale, Revised (CDRS-R) (Poznanski and Mokros, 1996). Our goal was to identify items of particular ratings challenge by detecting items of highest inter-rater variability.

Method: Using data from 798 score submissions from multiple pediatric depression clinical trials by multiple sponsors, standard deviations were calculated for each of the 17 CDRS-R items of ratings scored by US clinical trials investigators/raters who had viewed at least one of 9 standardized adolescent patient/parent videos as part of the qualification process for their respective clinical trials. The clinical trials investigators/raters had been trained extensively in live sessions on scale administration conventions immediately prior to viewing and scoring the videos. CDRS-R individual item standard deviations from each of the 9 videos were calculated and rank ordered from lowest to highest variability separately for each video. Rank ordering of SDs for each of the 9 videos was measured for cross-video agreement using Kendall W. **Results:** The cross-video agreement of item standard deviation rankings was statistically significant (Kendall $W=0.31$, $p < .0001$). Within the 5 most variable items for at least 4 of the 9 videos were Depressed Facial Affect (7 videos), Hypoactivity (6 videos), and Excessive Weeping, Low Self Esteem, and Morbid Ideation (each 4 videos).

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 nine completely different patient and parent videos were statistically similar, suggesting that scoring ease or difficulty of individual items occurred over and above the specifics of the patients rated. The high variability items provide targets for training and surveillance efforts.

Disclosures: J. Busner: Full-time employee of Bracket. R. L. Findling: In the last 12 months receives or has received research support, acted as a consultant and/or served on a speaker's bureau for American Academy of Child & Adolescent Psychiatry, American Physician Institute, American Psychiatric Press, Bracket, Clinsys, CogCubed, Cognition Group, Coronado Biosciences, Dana Foundation, Forest, Guilford Press, KemPharm, Lundbeck, Merck, NIH, Novartis, Otsuka, Oxford University Press, Pfizer, Physicians Postgraduate Press, Roche, Sage, Shire, Sunovion, Supernus Pharmaceuticals, Transcept Pharmaceuticals, Validus, and WebMD. A. Robb: receives or has received research or travel support, acted as a consultant, served on a speaker's bureau, or owns stock or equity in American Academy of Child and Adolescent Psychiatry, American Academy of Peditarics, Bracket, Bristol-Myers Squibb, Eli Lilly, Bristol-Myers Squibb, Forest, GlaxoSmithKline, Janssen Pharmaceuticals, Johnson & Johnson Services, Lundbeck, Merck, NICHHD, Otsuka, Pfizer, Patient-Centered Outcomes Research Institute, SyneuRX, Sunovion, Takeda.

24 Comparing PROMIS and Legacy Depression Measures in a Community Sample of Older Adults with Varying Levels of Cognitive Impairment

Levin J[1,2], Aebi M[1], Smyth K[2], Tatsuoka C[2], Sams J[2], Scheidemantel T[1], Sajatovic, M[1,2]

[1] University Hospitals Case Medical Center, Department of Psychiatry, [2] Case Western Reserve University School Of Medicine

Methodological Question Being Addressed: This study evaluated the utility of the Patient Reported Outcomes Measure Information System (PROMIS) depression battery to capture the relationship between depression and cognitive status in 304 community-dwelling older adults (61 cognitively normal, 139 with mild cognitive impairment (MCI), 104 with mild/moderate dementia). The study also compared performance of the PROMIS-8a short form and selected "Legacy" depression scales, including the Geriatric Depression Scale (GDS) and Montgomery-Asberg Depression Rating Scale (MADRS), in a well-educated and largely euthymic sample of older adults with varying levels of cognitive functioning.

Introduction: Depression is widely prevalent among older people and may be both a risk factor for and complication of dementing illnesses such as Alzheimer's disease (AD). While there are a variety of standardized tools to evaluate depression in older adults, there is still a need for brief and practical measures that are useful in elders with mild cognitive impairment and AD.

Methods: In this cross-sectional baseline analysis, which was part of a larger prospective cohort study, we administered the PROMIS depression battery and Legacy measures of depression to 304 individuals, age 70 and over. Study participants were grouped into three categories of cognitive functioning: normal cognition, mild cognitive impairment (MCI) or mild/moderate dementia. Cognition was assessed with the Saint Louis University Mental Status Examination (SLUMS).

Results: Mean age of the sample at baseline was 78.3, with a majority being female (71%), white (78.6%), and with at least a high-school education (89%). Most individuals did not have clinically significant depression at baseline, with means (SD) for the PROMIS-8a, GDS, and MADRS of 50.6 (4.6), 5.3 (5.1), and 4.4 (5.1), respectively. The proportion of individuals meeting at least mild symptomatology ("threshold") for clinically relevant depression was 62/300 (20.7%) on the PROMIS-8a battery, 53/296 (17.9%) on the GDS, and 57/300 (19.0%) on the MADRS. Total score on the PROMIS-8a correlated moderately with the GDS (r = .68) and MADRS (r = .56). Results of Chi-square analyses suggest that while all three measures identified a similar proportion of depressed individuals, the classification of which individuals were identified as having depression was different between the measures. Also, both PROMIS-8a and GDS measures appeared to underestimate the severity level of depression relative to MADRS. There were no significant differences between normal, MCI and dementia groups on PROMIS or Legacy depression measure scores.

Conclusions: The brief 8-item PROMIS depression battery correlates moderately well with the legacy 30-item self-administered GDS and the rater-administered MADRS. Measures differed on sensitivity to identify individuals with clinically relevant “threshold” depression in community-dwelling, well-educated elderly with varying levels of cognitive functioning. Still, given the brevity and ease of administration, our data suggest that the short-form PROMIS depression measure may be useful as a depression screen, particularly in community based samples.

Disclosures: Acknowledgment: this study was supported by ARRA grant AG038825 from the National Institute on Aging (NIA) to Drs. Smyth and Sajatovic. The authors report no conflicts of interest for this work.

25 Workshop Reports