

ISCTM 5th Annual Scientific Meeting

3-5 March 2009

Crystal City Marriott ~ Arlington, VA

Poster Abstracts

1 “Knowing When a Drug is Not Going to Work in Bipolar Depression”: Absence of Early Improvement as a Predictor of Later Non-Response in 3,369 Patients From 10 Placebo-Controlled Acute Trials

Joseph R Calabrese,¹ David E Kemp,¹ Stephen J Ganocy,¹ Martin Brecher,⁷ Berit X Carlson,² Suzanne Edwards,⁹ James Eudicone,² Gary Evoniuk,⁹ Wim T Jansen,⁶ Robert McQuade,³ Brian Millen,⁸ Margaret Minkwitz,⁷ Randall Owen,⁵ Andrei Pikalov,⁴ Armin Szegegi,⁶ Mauricio Tohen,⁸ Estelle Vester-Blokland,² Arjen PP van Willigenburg,⁶

1 Case Western Reserve University, Cleveland, Ohio; 2 Bristol-Myers Squibb, Plainsboro, NJ; 3 Otsuka Pharmaceutical & Commercialization, Inc., Princeton, NJ; 4 Otsuka America Pharmaceutical, Rockville, MD; 5 Bristol-Myers Squibb, Wallingford, CT; 6 Schering-Plough, Roseland, NJ; 7 AstraZeneca, Wilmington, DE; 8 Eli Lilly, Indianapolis, IN; 9 GlaxoSmithKline, RTP, NC.

Background:

In testing for differences between antidepressants and placebo in major depressive disorder (MDD), significant differences are usually not observed until week 3, leading to the belief that trials of antidepressants require 4-6 weeks. Survival analytic techniques (Stassen et al 2007) have been applied to short-term randomized controlled trials in MDD and suggest the probability of achieving response or remission in subjects experiencing early improvement is high. Similar analyses are presented below from patients enrolled into acute bipolar depression trials.

Methods:

Ten similarly-designed, multicenter, randomized, double-blind, placebo-controlled trials in 3,369 patients with bipolar I or II depression were blinded and used to determine if early improvement predicts later response and remission [2 aripiprazole, 5 lamotrigine, 1 olanzapine, olanzapine-fluoxetine combination (OFC) study, and 2 quetiapine studies]. Early improvement was defined as $\geq 20\%$ reduction from baseline in MADRS total score at Week 2. Response was defined as $\geq 50\%$ reduction in MADRS total score at endpoint (LOCF). Remission was defined as MADRS total score ≤ 10 at LOCF. Sensitivity, specificity, and positive (PPV) and negative (NPV) predictive values were calculated. To manage heterogeneous study outcomes, predictive power analyses were blindly pooled using LOCF, including 4 positive studies separating from placebo and the corresponding segregated placebo data, 6 negative/failed studies and the corresponding segregated placebo data, and pooled placebo data from all 10 studies.

Results:

- 1,456 patients were randomized to placebo and 1,913 to active compounds in this 10 study analysis.
- Negative predictive values were high (68-90%), suggesting the clinician can have high levels of confidence in knowing when a drug is not going to work in the acute treatment of bipolar depression.
- Positive predictive values were low (38-73%), suggesting that the clinician never knows if a drug is going to work in the acute treatment of bipolar depression.
- For studies that separated from placebo, false negatives were low for response (12-23%) and remission (11-19%).
- For studies that separated from placebo, false positives were higher for response (31-58%) and remission (35-64%).

Discussion:

- If these trial data apply to clinical practice, medication change could be considered after 2 weeks of treatment in the absence of early improvement.
- Efficacy, safety, and tolerability are important in drug development, but clinicians also value ‘predictability’, ‘knowing when a drug is going to work and knowing when it isn’t’.
- Very little research is devoted to the study of predictability of psychotropic drugs. Sponsors, regulators, and clinicians should consider this factor, especially in formulating evidence-based treatment guidelines.

Reference: Stassen JJ, Angst J, Hell D, Scharfetter C, Szegegi A: Is there a common resilience mechanism underlying antidepressant drug response? Evidence from 2848 patients. *J Clin Psychiatry* 2007;68:1195-1205.

Use of a Relapse Monitoring Board for Independent Assessment of the Primary End Point in an International Clinical Trial of Bipolar Disorder

Norris Turner, Pharm.D., Ph.D.,¹ Wayne Macfadden, M.D.,¹ Ravi Anand, M.D.,² Sumant Khanna, M.D., Ph.D., M.A.M.S., M.R.C.Psych.,³ Mark Hyman Rapaport, M.D.,⁴ J. Thomas Haskins, Ph.D.,¹ Ibrahim Turkoz, M.S.,¹ Larry Alphas, M.D., Ph.D.¹

¹Ortho-McNeil Janssen Scientific Affairs, LLC, Titusville, NJ, USA; ²Anand Pharma Consulting, St. Moritz, Switzerland; ³CliniRx Research Pvt. Ltd., Gurgaon, Haryana, India; ⁴Department of Psychiatry and Behavioral Neurosciences at Cedars-Sinai Medical Center and Department of Psychiatry at the David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Background: Review and monitoring boards may provide an independent, objective method for reviewing clinical trial data. This study introduced the concept of an independent relapse monitoring board (RMB) that determined relapse to a mood episode in an international, randomized, double-blind, placebo-controlled study of adjunctive risperidone long-acting therapy (RLAT) in patients with bipolar disorder.

Methods: Principal investigators (PIs) conducted an initial blinded assessment of patients to determine if a relapse occurred based on predefined protocol criteria. The independent RMB included 3 psychiatrists from different regions (USA, India, EU) with expertise in diagnostic, clinical and therapeutic management of bipolar disorder. Following the PI determinations of relapse, blinded to study drug, the RMB reviewed relevant data in all patients from the 52-week double-blind phase (mood severity rating scales, medication increases or changes and clinical records), discussed the case with each other, then attained consensus for whether a relapse occurred. Date of relapse was determined by the RMB using predefined protocol criteria as a guide. Study ID: CR004693.

Results: The RMB met 6 times during the study. The PIs identified 42 relapses: 20.8% with RLAT, 40.3% with placebo; the RMB identified 48 relapses: 22.2% with RLAT, 47.8% with placebo. 8 patients were determined to have relapsed by the RMB (due to the severity of their symptoms as reflected by mood rating scales or treatment that was administered) but not by the PIs, whereas 2 patients were determined to have relapsed by the PIs (after reviewing both patients' course) but not by the RMB. Adjunctive RLAT significantly delayed the onset of a mood episode compared with placebo using RMB- (P=0.004, log-rank test) or PI- (P=0.023) identified relapses. The relative relapse risk was 2.4-fold (P=0.004 chi-square [Cox regression]) and 2.1-fold (P=0.022) higher with placebo than with adjunctive RLAT using RMB- or PI- identified relapses, respectively.

Conclusion: RMBs for clinical trials can provide a comprehensive, sensitive and standardized evaluation of relapse that supports validation and generalizability of site-based clinical decision making. This may be particularly important in complex, large international trials where clinically meaningful outcomes may be difficult to discern.

Supported by Ortho-McNeil Janssen Scientific Affairs, LLC

Learnings From the Ziprasidone Bipolar Depression Program

Gary Sachs¹, Ilise Lombardo², Ruoyong Yang², Charlotte Kremer², Elizabeth Pappadopoulos²

¹Massachusetts General Hospital, Boston, MA, USA; ²Pfizer Inc, New York, NY, USA

Introduction: Several recent clinical trial programs for bipolar depression have failed to yield positive results. To better understand these types of findings, we further examined HAM-D and MADRS scores from 2 trials of ziprasidone for the treatment of bipolar I depression that failed to show superiority on the primary outcome measure.

Methods: 2 similarly designed 6-week, randomized, double-blind, multicenter, flexible-dose, placebo-controlled trials evaluating the efficacy and safety of ziprasidone in adults with bipolar I depression were examined. HAM-D-17 scores ≥ 20 and YMRS scores ≤ 12 were required at screening and baseline. Change in MADRS total score from baseline to week 6 was the primary outcome measure.

Results: 1617 subjects were screened; 928 were assigned to treatment (study 1: 171, 165, and 168 subjects for higher-dose ziprasidone [120–160 mg/d], lower-dose ziprasidone [40–80 mg/d], and placebo, respectively; study 2: 185 and 196 subjects for flexibly dosed ziprasidone [40–160 mg/d] and placebo, respectively). Ziprasidone did not demonstrate superiority over placebo for the primary outcome in either study. Although all subjects met entry criteria of HAM-D-17 scores of ≥ 20 at baseline, 174 (36%, study 1) and 116 (31%, study 2) subjects had baseline MADRS scores that did not correspond to the moderate severity indicated by HAM-D. Furthermore, 12 (3%, study 1) and 19 (5%, study 2) subjects were considered in remission at baseline (MADRS scores ≤ 12). At LOCF, 98 (20%) and 54 (15%) subjects had MADRS scores ≤ 4 respectively, including 28 and 12 subjects with MADRS = 0.

Conclusion: To mitigate baseline inflation of the primary efficacy measure, we used different scales to determine eligibility and efficacy. This approach, however, allowed for the inclusion of subjects with baseline MADRS scores below those often considered indicative of moderate to severe bipolar depression. The occurrence of MADRS scores ≤ 4 during the study provides further evidence of the low severity of depressive symptoms in some of the subjects evaluated in these studies. Thus, the inclusion of these mildly depressed patients in these analyses may have resulted in a study sample with lower than expected sensitivity to treatment-induced change.

4 **Regional and Temporal Differences in Use of Doctorate Level MADRS Raters in Bipolar Depression Trials**

Authors: David Daniel, MD¹, John Bartko, PhD², Norman Sartorius, MD, PhD, FRCPSch³, Eduard Vieta, MD⁴, Adam Butler⁵

¹United BioSource Corporation, McLean, VA ²Independent Consultant, Newville, PA ³Geneva University, Geneva, Switzerland
⁴University of Barcelona, Barcelona, Spain ⁵United BioSource Corporation, Wayne, PA

INTRODUCTION: In multi-country clinical trials, agreement among raters may be complicated by variations in rating practices. The MADRS is a widely recognized and utilized measure in bipolar depression clinical trials. There is no universal agreement regarding the level of professional credentials required to rate the MADRS. As such, we hypothesized that there would be significant variation by region and over time in the use of doctorate vs. non-doctorate raters for the MADRS.

METHODS: This analysis was designed to compare the use of doctorate vs. non-doctorate raters to rate the MADRS across nine industry sponsored multi-center bipolar depression clinical trials over several years (2005, 2006, 2007 and 2008) and among regions [North America (NA), South and Central America (SCA), Western Europe (WE), Eastern Europe (EE) and Asia]. The proportion of doctorate level raters was compared by region and by year of certification to rate using the chi-square statistic.

RESULTS: Over the past four years combined, the differences among the five regions in the use of doctorate vs. non-doctorate MADRS raters was statistically significant (chi square =161, $p < 0.000001$; $df=4$). This difference was primarily due to lower use of doctorate level raters in NA (below 54%) compared to above 82% for all other regions in each of the years assessed.

For the five regions combined the differences among the years 2005, 2006, 2007 and 2008 in the use of doctorate vs. non-doctorate MADRS raters was statistically significant (chi square = 34, $p < 0.000001$; $df = 3$).

CONCLUSION: The results are consistent with significant regional variation in the educational level of MADRS raters. NA utilized a significantly lower percentage of doctorate level raters than the rest of the world. Although the ratio of doctorate to non-doctorate MADRS raters appears to have declined over the last four years in the combined sample of all five regions, interpretation must be caveated by the disproportionate number of raters from NA in the sample in 2008. A longer observation period and more balanced sample geographically might reveal clearer temporal trends. In addition, all raters were trained by the same company which may be another source of sample bias.

5 **Memory Measures As Biomarkers In Psychiatric Drug Research**

Franco Di Cesare*, Dragana Kirola-Bugarski*, Cristiana Di Carlo§, Mario Fioravanti+

(*) i3 Research, CNS Psychiatry; (§) Leoben Concept Development, UK; (+) Department of Psychiatric Sciences and Psychological Medicine, University of Rome, "Sapienza"

INTRODUCTION:

Memory measures with increased sensitivity contribute to improve the knowledge on altered functional neurobiology underlying cognitive and affective disorder in psychiatric disorders. Robust evidence suggests memory and emotional-affective processing share common neural correlates (limbic and paralimbic, pre-frontal and sub-cortical structures). Memory measures derived from free recall paradigms (traditionally defined as episodic memory tasks) have been extensively used to elucidate the neurobiological and functional components of affective-cognitive processing. Their application as biomarkers may impact drug development by: more accurately identifying clinical target population, demonstrating proof-of-mechanism, providing surrogate endpoints for prediction of therapeutic intervention outcome.

METHODS:

We propose a tool for clinical drug research specifically designed to evaluate organizational processes

associated with memory retrieval. The method consists of the administration of a standardised sequentially-ordered verbal free recall tasks and allows for 4 different Acquisition and Recall modalities:

- Task 1: Acquisition/Immediate Recall of unstructured item list (List A)
- Task 2: Acquisition/Immediate Recall of structured item list (List B)
- Task 3: Acquisition/Immediate Retrieved Context Cued Recall of structured item list (List C)
- Task 4: Delayed Recall of unstructured item list (List A)

List A includes semantically unrelated items with randomly assigned position list. List B and C include semantically-related items, equally distributed across 4 different mutually-exclusive taxonomic categories, 5 items for each category. List items are sorted by taxonomic category and within category by alphabetical order. List B semantic organisation was kept latent to subjects; instructions were given to make the organisation of List C explicit to subjects before acquisition. It provides measures of recall accuracy (Correctly Recalled Items and Intrusions) and organization ((Adjusted Ratio of Clustering and Serial Clustering Index).

RESULTS:

Data from a sample of sixty healthy, both sexes (M=21, F=39), aged 33.4 y.rs ($SD\pm 10.0$), with 16.6 ($SD\pm 3.7$) years of education, suggests the method effectively describes organizational components of memory retrieval. Recall accuracy and organization measures are affected by the modality of acquisition/recall (ANOVA repeated measures, $p>0.001$).

CONCLUSION:

The results highlight the relevance of memory measures of retrieval and of semantic and serial associative processing as biomarkers with increased sensitivity to pharmacological effects and treatment response.

6 **Staggered Randomization in RCT's for Antipsychotics as a Method to Reduce Placebo Effect: Design and Issues with Implementation**

Glenn A. Phillips¹, Bruce Kinon¹, Brian Millen¹, Lu Zhang¹, Virginia Stauffer¹, David Adams¹

Eli Lilly and Company

Introduction: In clinical trials of antipsychotic treatments for schizophrenia there has historically been minimal placebo effect. Recently there has been an increase in the placebo effect in many schizophrenia clinical trials (Kemp et al, 2008). One possible way to reduce the placebo effect seen in clinical trials would be to employ a staggered randomization (SR) masking when patients actually begin treatment. Our group worked to develop a study using SR carrying out extensive discussions and significant investigative work by many groups involved in conducting a clinical trial. Here we review issues addressed in designing an RCT using staggering and discuss some potential approaches to overcoming some issues. We also provide issues that were not resolved in hopes of starting a discussion or hearing others' ideas on these issues.

Issues: The idea of SR is straight forward: patients are randomized not only to treatment but also to two or more time points for starting treatment. However there are several issues that arise from this straight forward concept. When implementing a SR design a decision has to be made about how many time points will be used and the spacing for these points. When implementing SR care must be taken to maintain the blind as to when treatment started. If the time randomization is maintained, study participants will necessarily receive treatment or placebo longer than the study period, which may raise ethical concerns. Can data from the staggered treatment groups be combined, or does the staggering create fundamentally different treatment groups that must be analyzed separately?

Discussion: The idea of randomizing patients to not only different treatments but to starting treatment at different time points may be a way to help address increasing placebo effects seen in neuroscience studies. Specifically SR may reduce placebo effect driven by rater/investigator/patient expectation as when treatment begins is an unknown. Implementation of SR will require careful consideration of several issues and may require new procedures for implementation.

Kemp, Schooler, et al (2008) What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? Schiz Bull published online on August 22, 2008.

7 **Psychiatric Illnesses in India**

Dr. Sumant Khanna, MD, PhD, MAMS, MRCPsych

Advisor CliniRx Research Pvt Ltd, Gurgaon, Haryana 122016; INDIA

Introduction:

The prevalence rate of Psychiatric illnesses in India is not significantly less than that in the western

countries. The mental health services in India consist of specialized mental hospitals, psychiatric units in general and teaching hospitals, private mental health clinics and nursing homes etc.

We aimed to determine the prevalence of mental illnesses in India, the prevalence of metabolic side effects and extra pyramidal symptoms in patients receiving anti psychotic treatment.

Method:

Extensive literature search was done using Pubmed and other relevant Indian medical journals.

Result:

The seriousness of the problem in India is indicated by the fact that the estimated overall prevalence rates of mental illness vary from 10/1000 to 370/1000. Recent studies in India in previously drug-naive patients with schizophrenia revealed an increased incidence of metabolic syndrome, of over 30% cases, after 6 weeks of therapy with a single antipsychotic drug. Indian studies have shown that patients are prone to suffer EPS when taking conventional antipsychotic drugs and therefore initial prophylaxis with anticholinergic is very important. Studies in India indicate that low dose of anti psychotic medication is required in Indian population. This could be due to the genetic/environmental factors leading to pharmacokinetic and pharmacodynamic variations. Regulatory scenario in India is also becoming more and more permissible. Now Phase II and phase III clinical trials can be conducted concurrently in India and other countries. The variations in the efficacy and safety data due to ethnicity will thus now be available relatively earlier

Conclusion:

The picture that emerges from these findings is that the prevalence of psychiatric illnesses in India is very significant. The incidence of metabolic and extra pyramidal side effects is also significantly higher in the Indian population.

8

Rater Training on SANS and Monitoring of Rater Performance During Clinical Trials

Susan De Santi, PhD1, 2, Richa Gaur, PhD2 and Luis Ramirez, MD3

1NYU School of Medicine, Department of Psychiatry, New York, NY ; 2The Cognition Group, Newark, DE ; 3Quality Outcomes Training, Brecksville, OH

Introduction:

Training a heterogeneous group of raters on the Scale for the Assessment of Negative Symptoms (SANS) is challenging as negative symptoms of schizophrenia are notoriously difficult to measure. This study evaluated ratings on the SANS and presents a novel methodology, Signal Enhancement System (SES), used in monitoring rater performance during the trial.

Methods:

104 raters from 8 countries participated in a SANS rater training program via combining online and face to face training at the Investigators' Meeting (IM). Rater's performance was analyzed and inter-rater reliability and concordance with expert raters were measured. During the trial SANS patient interviews and ratings were monitored with the help of the SES methodology, which involved collaboration between the Site Rater (SR) and an Independent Rater (IR). Patient interviews were recorded at the site via a laptop and transmitted to a secure server and rated by an IR (blind to the site ratings) via the website. These independent ratings were shared with the SR to enable SANS score reconciliation before submitting to the sponsor.

Results:

Inter-rater agreement for the total scores between expert and study raters ratings was substantial (kappa = .69, $p < .001$), moderate (kappa = .40 $p < .001$), in certification session I & II and poor (Kappa = .18, $p = .06$) in certification III, suggesting some raters experience difficulties administering the SANS. 95% raters achieved the passing grade (80% or higher). For the individual items the McNemar Test for paired binomial proportions showed raters agreement on only 8/25, 5/25 and 4/25 items with the expert ratings.

Conclusions:

95% of raters certified on the SANS, but many raters were not able to achieve sufficient accuracy in ratings. Training alone does not ensure reliable ratings and statistical analysis of ratings from the site cannot establish whether the patient interviews were done appropriately in the study. To overcome the challenge of maintaining a high reliability of assessment during the trial, SES was implemented. This methodology helped in monitoring of recorded patient interviews which ensured compliance with interview techniques, provided an opportunity for SR performance feedback, and a forum for SR and IR calibration of scores.

The Symptoms of Trauma Scale (SOTS): A Pilot Study

Michaela Mendelsohn, PhD¹, Diya Kallivayallil, PhD¹, Jocelyn Levitan, BA¹, Michael Pratts, MD², Mark Opler, PhD², MPH, & Judith Herman, MD¹

¹ Department of Psychiatry, Cambridge Health Alliance/Harvard Medical School, Cambridge, MA ² The PANSS Institute, New York, NY

Introduction: There are an abundance of self-report and interview-based measures for assessing *DSM-IV* posttraumatic stress symptoms; however, relatively few simultaneously cover complex symptomatology frequently observed among survivors of prolonged and repeated interpersonal trauma, as described by Herman (1992), van der Kolk et al (2005), and others. There is a need for a relatively brief measure that covers the full spectrum of trauma-related symptoms, is sensitive to change, and is suitable for use in clinical and research settings. The Symptoms of Trauma Scale (SOTS; Opler, Muenzenmaier, Shelley, & Grennan, 2004) consists of a 30-40 minute semi-structured interview and a 7-point companion rating scale that is used to measure severity, ranging from absent to extreme, of twelve symptoms shown to be associated with trauma. The SOTS is not a diagnostic measure but is designed to complement existing systems that establish and trauma history and diagnosis.

Methods: Three research clinicians at the Victims of Violence program of the Cambridge Health Alliance received training in the use of the SOTS interview and rating scale. N = 30 patients seeking individual or group therapy for trauma-related symptoms at the Victims of Violence program. *Gender composition:* 77% women, 23% men *Ethnicity:* 73% Caucasian, 14% mixed ethnicity, 7% Latino, 3% African American, 3% Native American. All patients completed a packet of self-report questionnaires assessing trauma-related symptoms (see list below). Patients were invited to participate in a brief interview that was scheduled within two weeks of completing the self-report measures and for which they were compensated \$25. Two raters were present for each interview, one of whom conducted the interview. Immediately following the interview, raters independently scored each item on SOTS rating scale.

Results: The interview was time efficient, easy to administer and score, and well tolerated by participants. Excellent interrater reliability (ICC > .8) was established for all items except Hyperarousal (see Table 1). Using a composite rating based on the average of the two raters, internal consistency reliability was .73 for the SOTS total. Cronbach's alpha for the four PTSD items was .37 and for the eight C-PTSD items was .64. There was moderate agreement between the SOTS total score and total scores on the PDS, BDI, and RSES (see Table 2). There was moderate agreement between the SOTS Complex PTSD subscale and total scores on the PDS (ICC = .43) and the RSES (ICC = -.51). There was moderate agreement between the SOTS PTSD subscale score and Re-experiencing (ICC = .58), Avoidance (ICC = .51) and Hyperarousal (ICC = .48) Clusters on the PDS but not the total score or other measures.

Conclusion: The SOTS performed well as a user-friendly measure of trauma symptom severity in a clinical setting. The SOTS demonstrated overall excellent interrater reliability and good internal consistency for the total score. Reliability problems associated with the Hyperarousal item will be addressed in the training program and SOTS manual. There was good agreement between the SOTS total score and measures of posttraumatic stress, depression and self-esteem. There was less initial psychometric support for the PTSD and C-PTSD subscales; factor analysis in a planned larger study will reveal whether other groupings of items provide meaningful information.

Comparison of Cognition Batteries for Use in Clinical Trials of Schizophrenia

Lyn Harper Mozley, Ajay Verma, Renee Vogt, Cindy Gargano, William Potter, Michael F. Egan

Clinical Neuroscience, Merck&Co, Inc, North Wales, PA 19454-1099

Introduction:

Developing drugs that improve cognition for patients with schizophrenia requires endpoints that are stable, reliable, and sensitive to drug effects. Several cognitive batteries have been used in cognition drug trials, but data are limited regarding test-retest reliability, sensitivity to change, and placebo/practice effects. We examined these features for two popular cognition batteries: 1) the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) battery, and 2) the Brief Assessment of Cognition in Schizophrenia (BACS) battery.

Methods:

Following a practice session at visit one, fifty-three patients with stable, chronic schizophrenia received single blind placebo at three consecutive weekly visits. Patients were then randomized to receive a single dose of placebo, d-amphetamine 15 mg, or modafinil 200 mg at the next three consecutive weekly visits in a double-blind, three-period crossover design. Patients were administered a 3 hr battery of cognitive tests at

each visit, including the BACS and MATRICS batteries. We also administered a computerized battery (CogState First in Man), results of which will be reported at a later date. We examined the composite scores' ICCs, effects of drugs on composite scores, and practice/placebo effects.

Results:

ICC values (with 95% confidence intervals) from Weeks 3 and 4 comparisons were: 1) MATRICS: 0.74 (0.61, 0.87) and 2) BACS: 0.78 (0.67, 0.89). No significant effects of amphetamine or modafinil were seen on either of the composite scores. Moderately large practice/placebo effects were seen for both batteries. These appeared to plateau after 3 sessions for the BACS, while MATRICS scores did not reach a plateau.

Conclusions:

ICC values both batteries were acceptable for clinical trials in this study. Placebo/practice effects were moderately large, with some variability regarding when they reached a plateau. The lack of effect of stimulants suggests these drugs may not be suitable as active comparators in drug trials of cognition-enhancing agents in patients with schizophrenia. It remains unclear how sensitive these batteries may be for detecting effects of non-stimulant drugs impacting novel cognition targets.

11 **Influence of Practice Effects on the Interpretation of Cognitive Function Data in Clinical Trials**

P. Turk^{1, 2}, B. K. Saxby^{1, 2}, M. C. Bardwell^{1, 2}, K. Wesnes^{1, 2}, C. Edgar³, P. Jongen⁴

1Cognitive Drug Research Ltd, Chicago, IL, USA, 2Cognitive Drug Research Ltd, Goring-on-Thames, United Kingdom, 3Clinical Trials Consultant, Reading, UK, 4MS4 Research Institute, Nijmegen, The Netherlands

Introduction: It is acknowledged that cognitive assessments are subject to learning/practice effects with repeated testing. Less known is that these factors may also lead to underestimation of decline in progressive disorders [1]. Consideration of these effects is critical to the planning, conduct and evaluation of longitudinal studies of progressive disorders as well as open-label drug trials [2] and might also influence results in placebo controlled trials [3]. Techniques to minimize the influence of such effects include alternate forms and pre-baseline training/familiarization. Impaired cognition and function is relevant to assessment in multiple sclerosis (MS), both as a progressive disorder and in clinical trials. Data revealing the influence of learning/practice effects are important to clinical trials methodology.

Methods: In an open-label study of INFb-1a (Avonex ®) (ClinicalTrials.gov identifier NCT00534261) cognitive, functional and quality of life assessment was conducted in forty-three relapsing-remitting MS patients. Prior to the baseline assessment on Day -1, training/familiarization assessments for the MS functional composite index (MSFC), the Digit Symbol Substitution Test (DSST) and a computerized cognitive battery with multiple parallel forms (The CDR System) were conducted (Days -60 and -30). Pre-treatment data were evaluated for change, test-retest reliability and correlation with functional status, quality of life and age.

Results: All assessments (MSFC, DSST and The CDR System total score) showed a statistically significant improvement from Day -60 to Day -30. The MSFC and DSST also showed a smaller though still statistically significant improvement from Day -30 to Day -1, while The CDR System total score showed no difference between the second two assessments. MSFC, DSST and The CDR System total score showed an equivalent relationship with the Expanded Disability Status Scale (EDSS).

Conclusions: These data showed the improvements which may be seen through practice/learning effects on cognitive and functional assessments in MS. These types of effects may influence both the longitudinal assessment of progressive decline in MS, but also the assessment of treatment effects. The findings confirm that the use of established training/familiarization procedures and test systems like The CDR System with alternate forms are able to minimize the impact of practice/learning effects as a confounding factor in the interpretation of clinical trial data.

References 1. Gold, M., Study design factors and patient demographics and their effect on the decline of placebo-treated subjects in randomized clinical trials in Alzheimer's disease. *J Clin Psychiatry*, 2007. 68(3): p. 430-8. 2. Goldberg, T.E., et al., Cognitive improvement after treatment with second-generation antipsychotic medications in first-episode schizophrenia: is it a practice effect? *Arch Gen Psychiatry*, 2007. 64(10): p. 1115-22. 3. Freedman, R., et al., Initial phase 2 trial of a nicotinic agonist in schizophrenia. *Am J Psychiatry*, 2008. 165(8): p. 1040-7.

12

A Comparison of the Matrics Consensus Cognitive Battery (MCCB) with the Cognitive Drug Research (CDR) System in Schizophrenia

M. C. Bardwell^{1, 2}, H. Hassman³, D. Krefetz³, M. Pinho³, S. Satek^{1, 2}, H. Brooker^{1, 2}, B. K. Saxby^{1, 2}, J. Ferguson⁴, K. Wesnes^{1,2}

1Cognitive Drug Research Ltd, Chicago, IL, USA, 2Cognitive Drug Research Ltd, Goring-on-Thames, United Kingdom, 3CRI Worldwide, Clementon, NJ, USA, 4University of Utah School of Medicine, Salt Lake City, UT, USA.

Background:

There is considerable interest in the attempt to treat the cognitive deficits which exist in schizophrenia. The MATRICS Consensus Clinical Battery (MCCB; www.matrics.ucla.edu) has been developed to provide an outcome measure for clinical trials of cognition-enhancing drugs for schizophrenia. The Cognitive Drug Research (CDR) computerized cognitive assessment system has been employed in worldwide clinical trials over the last 25 years and has been widely used to study the effects of cognition enhancers, in both normal volunteers and a variety of patient populations including schizophrenia (www.cdr.us.com). The present study was designed to compare the two batteries in schizophrenic patients.

Methods:

59 Schizophrenic patients on stable medication attended four study visits. The first two were for screening and training purposes, and on each of the third and fourth visits the patients performed the MCCB and The CDR System. An equal number of patients were administered the MCCB and The CDR System on each visit.

Results:

The various CDR scores revealed widespread impairments in the patients with large effect sizes (up to 3). Both the MCCB and The CDR System showed satisfactory test-retest reliability. The correlations between the batteries were also acceptable and in the appropriate direction demonstrating the tests are assessing the same domains, e.g. MCCB attention/vigilance correlated with CDR Power of Attention ($r=0.38$), Continuity of Attention ($r=0.36$), Fluctuations in Attention ($r=0.43$). Sizeable practice effects were seen on the MCCB verbal and visual memory scores at 20% and 37% respectively, whereas no practice effects were evident on the CDR equivalents.

Conclusions:

Both the MCCB and the CDR System showed acceptable test-retest reliability, and both systems correlated on comparable domains. The MCCB showed a greater likelihood of showing practice effects, which may limit its utility in clinical trials. Both systems are well validated and aside from the practice effects on the MCCB episodic memory tests, the choice between the two systems may rest on the available parallel forms, language versions available and the requirement for specialist test administration.

13

Is There a Better Way to Evaluate Cognition in Patients with Bipolar Disorder and Schizophrenia?

C Thomas Gualtieri & Sandeep Vaishnavi

North Carolina Neuropsychiatry Clinics, Chapel Hill, Charlotte & Raleigh

Introduction: A conventional approach to evaluating cognition in clinical trials of patients with bipolar disorder and schizophrenia is a test battery that generates correct responses and errors in tests of memory, attention, social acuity, working memory, reasoning and processing speed; a reasonable approach in its day, but one that unaccountably neglects the executive functions, reaction time (RT) and reaction time variability (RTV). It is likely, therefore, that important aspects of neurocognition in these conditions are not being captured.

Method: A computerized neurocognitive battery, including tests of memory, attention, social acuity, working memory, reasoning and processing speed; as well as measures of executive function, RT and RTV was administered to 3 groups: normal controls, bipolar and schizophrenic patients, matched for age, race, gender, education and computer familiarity.

Results: The strongest differences among the three groups were registered in tests of executive function, complex attention, RT and RTV in a measure of shifting attention. The correlation between RT and RTV was inversely related to the severity of the condition, a finding that the authors have also determined in studies of patients with brain injuries, MCI and dementia.

Conclusion: The evaluation of neurocognition in patients with severe mental illness should not be confined to a limited battery, but should incorporate measures that speak to regulatory control and neural timing mechanisms, that are obviously and severely impaired, especially in patients with schizophrenia.

Neurocognitive Evaluation of Patients with Traumatic Brain Injuries

C Thomas Gualtieri & Sandeep Vaishnavi

North Carolina Neuropsychiatry Clinics, Chapel Hill, Charlotte & Raleigh

Introduction: Every year, more than 2M Americans sustain brain injuries and at least 10% are moderate to severe, with lasting motor or sensory deficits, cognitive impairment and emotional instability. The judicious application of psychotropic drugs and other somatic treatments has the potential to improve cognition and neurobehavioral symptoms. Monitoring treatment response usually entails serial cognitive testing.

Method: A computerized neurocognitive test battery was administered to 777 patients age 15-70 who had sustained moderate or severe TBI, and compared to results from normal control subjects, matched for age, race, gender, education and computer familiarity. Comparison groups were patients with post-concussion syndrome (N=97), mild brain injury (recovered)(25) and post-traumatic stress disorder (203).

Results: Although severe TBI patients were impaired in all cognitive domains, relative to normals, the most sensitive and specific differentiators were measures of psychomotor speed, processing speed, executive function and reaction time variability. The same measures were also successful in distinguishing mild TBI patients from patients with PTSD, and in tracking recovery from concussion.

Conclusion: Neurocognitive assessment of TBI patients requires a comprehensive battery, with tests of memory, psychomotor speed, reaction time and RT variability, executive function and attention. Specific tests in a comprehensive battery are expected to be particularly impaired, and may be useful for distinguishing patients with PTSD.

Tiagabine in Combat Related Post Traumatic Stress Disorder: Single Site Analysis from a Multicenter, Double-Blind, Randomized, Placebo Controlled Study

¹Kenneth N Sokolski, MD, ²Debra M Hoffmeyer M.A. ³Dimitrios Gripios, B.A., ⁴Aleasa A Sokolski, ²Tiffany L Castanon, B.A.

¹University of California, Irvine ²Clinical Innovations, Inc, ³VA Long Beach Healthcare System ⁴University of California, San Diego

BACKGROUND: Combat related Post Traumatic Stress Disorder (PTSD) is a serious, disabling condition which often fails to respond to pharmacotherapy. Although Tiagabine has demonstrated some benefit in civilian PTSD, there are no studies investigating its effectiveness in war related trauma.

OBJECTIVE: To investigate the effects of Tiagabine given as monotherapy in Vietnam combat veterans with PTSD.

METHODS: Seven Vietnam veterans diagnosed with PTSD were randomly assigned Tiagabine or matching placebo as part of a multicenter study on the effects of Tiagabine in Post Traumatic Stress Disorder. Subjects were treated for twelve weeks with flexible doses of Tiagabine or placebo up to 16 mg per day followed by 12 months of open label Tiagabine for those consenting to continue the study. Symptom severity was evaluated using the Clinician Administered Post Traumatic Stress Test (CAPSS), the Treatment Outcome PTSD Scale (TOP-8), the Davidson Trauma Scale (DTS), the Connor Davidson Resilience Scale (CD-RISC), as well as the Clinical Global Impression Severity (CGI-S) and Clinical Global Impression Improvement (CGI-I) ratings administered at weeks 2, 4, 8, and 12 during double-blind treatment and monthly during open-label therapy.

RESULTS: Subjects were markedly ill prior to randomization (mean CGI-S=5.0). Following 12 weeks double-blind treatment, mean total CAPSS scores decreased from 86.7 to 66.0 in the active treatment group and from 73.8 to 69.0 in placebo treated subjects. Two of the three subjects on active medication and one of the four on placebo were rated “much improved” or “very much improved” on the CGI-I. Open-label treatment with Tiagabine for months resulted in a decrease in mean total CAPSS scores from 77.0 at open label baseline compared to 65.5 at end of treatment. Of the four subjects involved in the open-label phase, three were judged “much improved” while the fourth subject showed no benefit. Less than a 10% difference between Tiagabine treated subjects and placebo was obtained on other rating scales. Nausea and sedation were the most common adverse events during up-titration however no subject withdrew from the study due to side effects.

CONCLUSIONS: These preliminary results from a small post hoc subgroup of Vietnam combat veterans suggest that monotherapy with Tiagabine may benefit symptoms of combat related PTSD. Larger sample sizes are needed to verify these findings.

Test-retest Characteristics of the MATRICS Consensus Cognitive Battery in a 20-site Schizophrenia Clinical Trial of R3487/MEM3454 Versus Placebo

Richard S.E. Keefe¹, Cynthia Siu², Kolleen Hurley Fox³, David A. Lowe⁴, George Garabaldi⁵, Luca Santarelli⁵, Stephen R. Murray⁵

1Duke University Medical Center, Durham NC, 2USA Data Power (DP), Inc. 3NeuroCog Trials, Inc. 4Memory Pharmaceuticals 5Roche

Background: The Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Project produced a battery of tests, the MATRICS Consensus Cognitive Battery (MCCB), designed to assess cognitive treatment effects in clinical trials of patients with schizophrenia. In validation studies, the MCCB demonstrated excellent reliability, minimal practice effects and large correlations with measures of functional capacity. It has been an empirical question whether the MCCB would demonstrate these favorable characteristics when administered in the context of the type of large multi-site industry trial for which it was designed.

Methods: 215 patients with schizophrenia maintained on a stable dose of a second generation antipsychotic therapy were enrolled into a randomized, double-blind, placebo-controlled trial of R3487/MEM3454. Testers from 20 sites were trained and certified, and all MCCB data were reviewed and re-scored centrally. The MCCB was administered at screening and 7-14 days later at baseline. A measure of functional capacity, the UCSD Performance-based Skills Assessment (UPSA) was also measured at baseline. The MCCB generates a composite score and cognitive domain scores standardized to a normative population with mean (T) = 50 and SD = 10.

Results: Baseline T-scores for the 7 MCCB cognitive domains and a composite score were determined for 154 male and 61 female subjects, mean age 39.6 years (SD=9.0), mean PANSS total score 57.0 (SD=10.3) and mean UPSA-2 total score 85.0 (SD=15.0). Only 14 test scores were missing out of a total of 4300 test assessments for the 10 MCCB tests performed in 215 subjects at 2 occasions (99.7% complete). All 215 (100%) patients had sufficient data for computing a composite score according the MCCB criteria. The mean (SD) MCCB composite score was 25.1 (11.6) at screening and 27.6 (12.1) at baseline. The test-retest reliability for the MCCB composite score was very high (ICC=0.88). Construct validity was also strong, as the MCCB composite score demonstrated a large correlation with the UPSA composite score ($r=.56$, $df=187$, $P<.001$). The practice effect on the composite score was small ($z=0.21$).

Discussion: In the context of a 20-site clinical trial in stable patients with schizophrenia, the MCCB is sensitive to cognitive deficits in all domains, demonstrates excellent test-retest reliability and construct validity, and small practice effects.

Methodological Considerations for Long-term Efficacy Evaluation of a Treatment for Breakthrough Pain

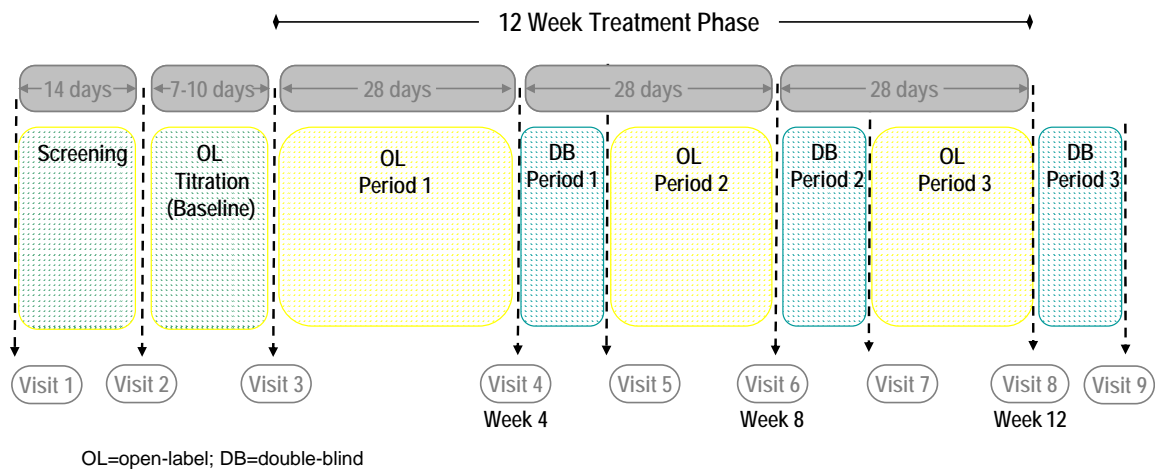
John Messina, Pharm D¹; Fang Xie, PhD¹

¹Cephalon, Inc., Frazer, Pennsylvania

Introduction: Methodological and practical challenges are inherent in designing a study to investigate a treatment for the breakthrough pain component of chronic pain. The design must consider the typical onset (sudden), duration (transient), and intensity (often severe/excruciating) of breakthrough pain. Thus, episodic pain evaluations and limited placebo exposure are important.

Objective: Assess the efficacy of fentanyl buccal tablet (FBT) following 12-week treatment in opioid-tolerant patients with chronic, persistent pain and breakthrough pain.

Methods: A modified, randomized withdrawal study design was utilized. Open-label FBT titration was used to determine each patient's successful dose (Portenoy RK, et al. *Curr Med Res Opin.* 2007;23(1):223-233). Open-label treatment with FBT was followed by within-patient randomization to a sequence of 9 double-blind treatments (6 FBT, 3 placebo). This open-label/double-blind cycle was repeated twice. Sum of pain intensity differences over 60 minutes (SPID₆₀) after week 12 was the primary outcome. Multiple in-patient comparisons between FBT and placebo were made during each double-blind period using efficacy variables, including pain intensity differences (5-120 minutes) and pain relief scores.



SPID₆₀ was analyzed using an analysis of variance (ANOVA) model for a 2x9 crossover design and included treatment, episode, and carryover as fixed effects and patients as a random effect. Treatment effect hypothesis was tested at the 2-sided 5% level. Point estimates and a 95% confidence interval for treatment effect were calculated from the ANOVA model.

Conclusion: Because this novel study design allowed for controlled efficacy assessments of active treatment over time while limiting patient exposure to placebo, it was appropriate for the study of breakthrough pain.

18

Age at Onset of Major Affective Disorders

Ross J. Baldessarini MD; Leonardo Tondo MD, MPH; Núria Cruz MD, PhD

International Consortium for Mood Disorder Research, McLean Hospital, Boston, MA

Background: Previous studies of onset age in major affective disorders have rarely considered large samples of various standard diagnoses made reliably by modern criteria in the same setting.

Methods: We considered age at first mental-health contact for major illness among 3038 adults with bipolar (BP) type I (BP-I; n=563), type II (BP-II; n=383), or unipolar major depressive (UP-MDD; n=2092) disorders, evaluated at intake and repeatedly at Lucio Bini Mood Disorders Center in Cagliari, Sardinia since the 1970s (by LT), and providing informed consent.

Results: Median ± IQR onset-age ranked: BP-I (24±12) < BP-II (29±19) < UP-MDD (31±24 yrs) ($p<0.0001$; Fig. 1). There was no evidence of secular trends in the consistent relationships of onset age versus diagnosis (between 1975 and 2009; Fig. 2). Men & women differed little, but the female:male *sex-ratio* of onset age ranked UP > BP-II > BP-I (Table 1), and onset tended to be youngest for psychosis or mania and oldest for depression and hypomania (not shown). Multivariate following bivariate (not shown) regression associated *younger* onset with: being unmarried, BP diagnosis (or psychotic-manic onset), substance-abuse, unemployment, and suicidality (Table 2).

Discussion: In a large patient cohort with consistent, modern diagnoses over four decades, the findings support impressions that BP-I illness begins younger, on average, than MDD, with intermediate onset for BP-II cases. Skewing toward older onsets (larger IQRs) follows the same order, and women have later onset only in major depressive disorder and perhaps bipolar II disorder. In bipolar I patients, psychosis and mania have somewhat earlier onset than depression, but all onsets tend to occur younger than in bipolar II or unipolar patients. Major mood disorders are not all alike psychobiologically. Onset-age may serve as a quantifiable phenotype for biological studies, and different treatments are required for each syndrome.