

ISCTM 6th Annual Scientific Meeting

22-24 February 2010

The Fairmont – Washington, D.C.

Poster Abstracts – Session Monday 22 February 2010

Abstracts are listed in order of presentation.

Classifications: Methods 1-25; Clinical 26-32

1 The Role and Interpretation of RCTs with Small Sample Sizes in PTSD Intervention Research

Lori L. Davis, M.D.¹ Andrew C. Leon, Ph.D.²

1University of Alabama School of Medicine, 2 Weill Cornell Medical College

Introduction: A randomized clinical trial (RCT) is designed to determine whether there is evidence of efficacy and if so, to estimate the magnitude of the treatment effect. Towards this aim, an essential element in the RCT design is sample size determination (Leon and Davis 2009). Many PTSD clinical trials have been conducted with very small sample sizes, yet their results have been cited as evidence for use of various classes of medications. For example, studies of anticonvulsants and antipsychotics for PTSD have had sample sizes as small as 15 subjects in total, yet they have served as the empirical basis for off-label use of these agents.

Methods: The role, interpretation, and limitations of studies with small sample sizes in PTSD intervention research will be reviewed. Several published and unpublished datasets from small RCTs in PTSD will be used as illustrations. Confidence intervals will be calculated for each study. Sample size calculations for designing a hypothetical efficacy trial based on these pilot study results will be compared to those derived from alternative methods.

Results: Pilot studies are important for evaluating feasibility of recruitment, randomization, retention, assessments, and implementation of the intervention (Kraemer et al., 2006). However, it has been shown that the confidence interval around the effect size is quite wide when based on the small sample size used in a pilot study. It is not unusual for a pilot study to enroll less than 10 subjects per treatment arm. However, the between-group effect size of $d=.50$ (sd units) from a sample of 8 subjects per group, for example, has a 95% confidence interval ranging from about $-.50$ (i.e. a moderate advantage of the control) to $+1.50$ (i.e. an enormous benefit for the investigational intervention; Leon 2008).

Conclusion: The use of pilot data from small studies for either determining effects size for large scale studies or guiding clinical decision-making should be avoided.

2 Reporting Adverse Event Data From Clinical Trials of Antipsychotic Agents

I Turkoz,* CA Bossie,** YW Ma,* L Alphs**

**Johnson & Johnson Pharmaceutical Research and Development, LLC, Titusville, NJ, USA, **Ortho-McNeil Janssen Scientific Affairs, LLC, Titusville, NJ, USA*

Introduction: Establishing safety profiles of new pharmaceuticals is a critical element of their development. However, most formal safety data presentations have significant limitations. This analysis evaluated relative risk (RR) of individual adverse events (AEs) for antipsychotic treatment compared with placebo from clinical trial databases to confirm, further explore, and identify safety signals.

Methods: Three databases from randomized double-blind placebo-controlled studies of antipsychotics agents were identified: one 13-week study of an injectable antipsychotic in schizophrenia (analysis 1); two 6-week studies of an oral antipsychotic in schizophrenia (analysis 2); and two 6-week studies of an oral antipsychotic in schizoaffective disorder (analysis 3). AE incidence rates were evaluated using graphical displays including RR and corresponding 95%CI. The primary goal was to highlight potential signals by providing estimates (and their precision) of treatment effect. RR with active treatment versus placebo was considered potentially significant when 95%CI did not cross 1 on the x-axis. No computations were adjusted for multiplicity.

Results: Graphical displays were generated for rates and RRs of the most common AEs for each database ($\geq 5\%$, either group). Data suggested a potential higher risk with active treatment versus placebo for extrapyramidal disorder (RR=2.53; 95%CI=1.21,5.27) and tachycardia (RR=2.14; 95%CI=1.11,4.14) in analysis 2; and tremor (RR=2.33; 95%CI=1.05,5.18) in analysis 3. When 95%CIs for RR included 1, risk was not considered significant between groups. However, borderline RRs greater with active treatment may

be of clinical interest. AEs with a higher, though nonsignificant, risk with active treatment were injection site pain (RR=2.07; 95%CI=0.891,4.821) in analysis 1; akathisia (RR=1.74; 95%CI=0.99,3.05) in analysis 2; and hypertonia (RR=2.77; 95%CI=0.97,7.89), somnolence (RR=2.65; 95%CI=0.92,7.58), and dyspepsia (RR=2.21; 95%CI=0.85,5.74) in analysis 3.

Conclusion: RR (vs placebo) and AE rates were generated in graphical displays versus descriptive statistics for antipsychotic agents. The identified AEs may warrant greater attention by the clinician because the RRs provide additional information over aggregate summaries of patient data. This methodology may provide a clinically more interpretable way to report AE data from clinical trials. Supported by Ortho-McNeil Janssen Scientific Affairs, LLC

3 **Continuing Education Programs Can Assess Practitioner Knowledge of Drug Labelling**

Ingrid Kohlstadt¹ MD, MPH, FACN, FACPM, Dianne M Murphy¹ MD, FACP, FDA, Lon Osmond² Director

1FDA, Office of Pediatric Therapeutics; Rockville, 2Audio-Digest Foundation; Glendale, CA

Disclaimer: This scientific abstract is the work of its authors. No endorsement by the FDA is intended or should be inferred.

Introduction: How familiar are physicians with drug labeling? We hypothesized that continuing medical education (CME) pretest questions and evaluations can ascertain practitioner knowledge of drug indications and adverse effects. To test this hypothesis we developed a CME on atypical antipsychotics in children.

Methodology: The CME was a 60-minute moderated panel of four experts discussing the indications and adverse effects of atypical antipsychotics. It was presented in the biweekly audio CME program, Audio-Digest Pediatrics. The transcribed audio-program was then offered thru MedscapeCME. The concurrent offerings differed in their degree of self-selection and practitioner profiles. The pretest questions developed with the two programs were overlapping. The AudioDigest evaluation asked participants to write what they learned from the activity.

Results: 711 practitioners completed the AudioDigest program in 3 months and 2400 practitioners completed the MedscapeCME program in 6 weeks. Pretest scores are presented in Table 1. Mean scores were lower than 70%, the average pretest score for all AudioDigest Pediatrics programs. Of the 37 written responses ascertained from the AudioDigest evaluations, 29 pertained to adverse drug effects.

Table 1. Practitioner responses to the pretest of a continuing medical education program offered by AudioDigest and MedscapeCME

CME Pretest Question	Format	% Correct	
AudioDigest		% Correct	MedscapeCME
Indications in general	Multiple Choice	75%	83%
Autism as an indication	Multiple Choice	62%	48%
IM route in children	True/False	65%	N/A
Tardive dyskinesia	Multiple Choice	37%	N/A
Akathisia	Multiple Choice	N/A	78%
Life expectancy	Multiple Choice	48%	35%
Cardiometabolic effects	True/False	95%	N/A
Cardiometabolic effects	Multiple Choice	54%	N/A
Clinical applications for Cardiometabolic effects	Multiple Choice	N/A	77%
Prolactin levels	Multiple Choice	80%	73%
Drug interactions	Multiple Choice	67%	N/A
When to report to Medwatch	Multiple Choice	52%	N/A
Mean score		64%	66%

Conclusions: In this CME the 3,111 practitioners appeared to be knowledgeable about the drug indications

in general, but not the autism indication or the indicated route of administration. Most practitioners correctly identified akathisia as an adverse effect, but were not familiar with the risk of tardive dyskinesia. Practitioners demonstrated knowledge about cardiometabolic drug effects and monitoring prolactin levels; however, most were not aware of the extent to which chronic medical conditions shorten life expectancy among the mentally ill.

In addition to educating practitioners on drug indications and adverse effects, CME might assess practitioner knowledge. In some settings the pretest score among practitioners might assist in guiding regulators on post-marketing drug safety.

4 **Validation of the Extrapyramidal Symptoms Rating Scale—Abbreviated in Patients With Schizophrenia**

L Alphs,* CA Bossie,* I Turkoz**

**Ortho-McNeil Janssen Scientific Affairs, LLC, Titusville, NJ, USA, **Johnson & Johnson Pharmaceutical Research and Development, LLC, Titusville, NJ, USA*

Introduction: Movement disorders remain a concern when prescribing antipsychotics for patients with schizophrenia. The Extrapyramidal Symptoms Rating Scale—Abbreviated (ESRS-A), derived from the ESRS, was designed as a more concise instrument for assessing the full spectrum of movement disorders addressed in the original scale. In this analysis, the content and convergent/divergent validity of the ESRS-A were explored.

Methods: Baseline data were from a 6-week international study of patients with schizophrenia. Convergent/divergent and content validity were assessed for parkinsonism by examining correlations among ESRS-A global and total parkinsonian subscores and total scores on the SAS; for dyskinesia by examining correlations among ESRS-A global and total dyskinesia subscores and total scores on the AIMS; for akathisia by examining correlations among ESRS-A global and total akathisia subscores and total scores on the BAS. Divergent validity for each construct was assessed by correlations between ESRS-A global and total subscores and alternative movement constructs. Spearman correlations were used for all analyses. The following convention was used to express the correlation strength: <0.3=poor, 0.3-<0.5=moderate, 0.5-0.7=good, >0.7=very good.

Results: Data were available for 197 subjects. Correlations assessing convergent validity for ESRS-A global and total subscores for parkinsonism, akathisia, dystonia, and dyskinesia were all good to very good (0.60-0.98). As anticipated, correlations representing divergent validity were generally lower (0.09-0.51). However, numerous correlations among the divergent constructs were moderate, even among the widely used validated scales (0.35-0.48 for SAS, AIMS, and BAS scores).

Conclusion: The ESRS-A showed good to very good convergent validity with validated measures for movement disorders, suggesting the ESRS-A can be used to assess different movement constructs, compared with existing scales. Divergent validity results suggest individual movement disorder constructs are empirically distinguishable from other movement disorder constructs. However, the moderate correlations for divergent validity suggest incomplete discrimination among these measures. This failure to completely discriminate among these movement disorders may be related to low occurrence of movement disorders in the dataset, overlap among constructs, insufficient investigator training, and/or inadequacy of the scales to clearly distinguish among them. Nonetheless, taken together, these results support the construct validity of the ESRS-A.

5 **Exploring Factors Influencing Assay Sensitivity in a Clinical Trial in Schizophrenia**

Eric J. Watsky, Jeffrey H. Schwartz, Charlotte M. E. Kremer

Pfizer Inc

Objective: To generate a set of site-based exploratory factors relevant to clinical trial conduct to correlate with degree of placebo response and separation of active control from placebo in psychosis studies. This is an ongoing effort initiated at Pfizer to better understand factors influencing clinical trial precision.

Methods: This work was conducted in part utilizing placebo and active control data from a schizophrenia clinical trial. Seven site-based exploratory factors were identified for analysis based on their putative relevance to clinical trial outcome. These included: 1. number of patients screened per site; 2. the ratio of the number randomized to the number screened; 3. correlation of PANSS and CGI, at baseline, final visit, and the final visit-baseline difference; 4. the ratio of the number of subjects per site that met per protocol criteria relative to the intent-to-treat population; 5. the amount of non-missing data per site; 6. the mean number of days in the study; and 7. center size. A correlation estimate was then generated for each of these variables exploring their relationship with both placebo response and the separation of active treatment

from placebo. The study population defined as intent-to-treat (ITT) included 163 patients with acute symptoms of schizophrenia treated for 21 days, of which 68 were randomized to either aripiprazole or to placebo.

Results: There were 35 subjects randomized to placebo and 33 randomized to aripiprazole. The placebo least squares mean change from baseline to Day 21 and the aripiprazole-placebo least squares mean difference in change from baseline to Day 21 were correlated with each of the 7 variables of interest. Weak correlations were noted between: 1) increasing numbers screened ($r^2=16.68\%$) and decreasing placebo response, and 2) the decreasing ratio of the number randomized to the number screened with the increasing separation of active treatment from placebo ($r^2=18.55\%$).

Conclusions: Although this initial exploratory effort to identify variables relevant to clinical trial outcome in a schizophrenia clinical trial did not demonstrate robust correlations, the limited findings are consistent with the value of screening in lowering placebo response and increasing active-placebo separation. The low values noted may be due in part to the small study sample size. This post-hoc exploratory strategy may be applied to the results of other clinical trials and may provide a useful quantitative approach to identifying factors that are material to the successful conduct of clinical trials in psychosis.

6 **Post-HOC Endpoint Readjudication of the Secondary Sudden Death Endpoint Per ICD 10 Coding in the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) Trial**

Onur Karayal, MD MPH, Jamie Geier, PhD, Charlotte Kremer, MD, Sheela Kolluri, PhD, Robert Reynolds, ScD, Sybil Eng, PhD, John Camm, MD, Michael Lewis, MD, Martin Keane, MD, Brian Strom, MD MPH

Pfizer Inc, New York, New York [OK, JG, CK, SK, RR, SE], St George's Hospital, London, London [JC], EPES Epidemiology GmbH, Berlin [ML], Berlin, University of Pennsylvania Hospital, Philadelphia, PA [MK], University of Pennsylvania School of Medicine, Philadelphia, PA [BS]

Background: The ZODIAC trial compared the risk of non-suicide death associated with ziprasidone versus olanzapine in real-world use. Following the initial submission of the results, the FDA requested an additional analysis with a readjudication using ICD 10 coding guidelines of the secondary outcome of sudden death.

Methods: A post-hoc readjudication of all sudden death events was conducted using ICD10 coding criteria for sudden cardiac death (I46.1) and sudden death not otherwise specified (NOS: R96.0 and R96.1). A sensitivity analysis of this post hoc analysis was also conducted incorporating unattended or ill-defined causes of death (R98 and R99). An additional supplemental sensitivity analysis was conducted to capture cases which might fall outside of the restrictive ICD-10 coding schema. Analyses included treatment comparisons for the 1-year incidence of mortality by estimating relative risks (ziprasidone incidence/olanzapine incidence) and the corresponding 95% confidence intervals (CIs). Mortality rates were also compared during the time when subjects were using their assigned study medication by estimating rate ratios (ziprasidone rate/olanzapine rate) and the corresponding 95% CIs.

Results: Data from the post-hoc readjudication of sudden death were consistent with the study's initial findings. There was no statistically significant difference in 1-year incidence of mortality between ziprasidone and olanzapine for sudden death NOS and sudden cardiac death (ICD 10 Codes R96.0 or R96.1 or I46.1) (Relative Risk = 1.11, 95% CI: 0.45, 2.77). The sensitivity analysis incorporating unattended or ill-defined causes of death (R98 and R99) yielded a similar result (Relative Risk=0.73, 95% CI: 0.44, 1.22). Further, the supplemental sensitivity analysis resulted in a risk ratio of 0.99 (95% CI: 0.65, 1.50). The results from the analysis of mortality rates based on person-time on assigned treatment were consistent with the results from the analysis of 1-year incidence of mortality.

Conclusions: ZODIAC is the largest randomized study of patients with schizophrenia conducted to date but did not detect an increased risk of non-suicide death associated with the use of ziprasidone vs. olanzapine. There was no statistically significant difference in the risk of sudden death comparing persons randomized to ziprasidone vs. olanzapine across all readjudicated endpoints.

7 **Evaluation of an Interactive Voice Response Version of the Electronic Columbia-Suicide Severity Rating Scale (eC-SSRS)**

Mundt, JC¹ Greist, JH¹ Gelenberg, AJ¹ Katzelnick, DJ¹ Jefferson, JW¹ Modell, JG²

¹Healthcare Technology Systems, Inc., Madison, WI, ²GlaxoSmithKline, Inc., ResearchTriangle Park, NC

Aim: To evaluate a computer-automated version of the Columbia-Suicide Severity Rating Scale using interactive voice response technology. The eC-SSRS assesses Lifetime suicidal ideations and behaviors at

baseline and prospectively monitors ideations and behaviors thereafter.

Methods: Ten control volunteers and ten psychiatric inpatients were administered the C-SSRS at baseline and 4-8 days later by two experienced clinical trial raters. Subjects used touch-tone telephones to complete the eC-SSRS. Beck Scale for Suicide Ideation forms and study feedback documents were also completed. Kappa measures compared inter-rater reliability between the human C-SSRS administrations, and compared the C-SSRS results with the eC-SSRS data. Convergent validity with the BSS was evaluated and patient feedback was obtained.

Results: Twenty baseline and nineteen follow-up assessments were completed. Suicidal ideation was identified by either C-SSRS or eC-SSRS in 12 of the 39 assessments. In six assessments, both raters and the eC-SSRS were in agreement regarding the most severe suicidal ideation. In the other six, all three rating methods differed once, both raters agreed but differed from the eC-SSRS once, and the eC-SSRS agreed with one rater but not the other in four. Suicidal behaviors were identified in 7 assessments by both raters and the eC-SSRS. In comparisons of four *types* of behaviors (actual, interrupted, or aborted attempts, or preparatory behaviors), both humans agreed 82% of the time, Rater 1 agreed with the eC-SSRS 64% of the time, and Rater 2 agreed with the eC-SSRS 82% of the time. Kendall's tau concordance with BBS scores supported convergent and construct validity of the C-SSRS and eC-SSRS; patient feedback was generally supportive of face validity between scale administration methods.

Conclusions: The reliability and validity of the C-SSRS and eC-SSRS were comparable in this study, supporting the feasibility and validity of the eC-SSRS for prospectively monitoring suicidality in clinical trials or clinical care.

Disclosures and Acknowledgements: Study support provided by GlaxoSmithKline and eResearch Technologies. eResearch Technologies provides ePRO services, including the eC-SSRS. Licensing fees are provided to the Research Foundation for Mental Hygiene and Healthcare Technology Systems for delivery of eC-SSRS assessments.

8 Placebo Response in Antipsychotic Trials

Potkin, SG,¹ Agid, O,² Remington, G,² Kapur, S,³ Watsky, E,⁴ Vanderburg, D,⁴ Siu, CO⁵

1University of California, Irvine, Irvine, CA; 2University of Toronto, Toronto, Ontario, Canada; 3Kings College London, IOP, London, London, United Kingdom; 4Pfizer, Inc., New York, NY; 5Datapower (DP), Inc., Ringoes, NJ.

Objective: Large placebo responses in antipsychotic trials present a major challenge for psychopharmacologic drug development. The analysis aimed to identify moderators of placebo response in antipsychotic trials, and to investigate the trajectory patterns of long-term placebo response.

Method: We searched the MEDLINE database for RCTs published in 1966 to 2009 supplemented by other electronic databases. In this analysis, placebo response in short-term treatment (2 to 12 weeks) was defined as mean change from baseline in BPRS total score (derived from PANSS in 11 studies). A meta-regression analysis was performed to identify influential moderators of placebo response. Patient-level analysis was conducted to identify classes of trajectory for the placebo response using Growth Mixture Model (GMM).

Results: A total of 1246 placebo-treated patients from 41 RCTs had valid BPRS total scores. Demographics included: weighted mean age 38 years, duration of illness 16 years, and 77% male. The weighted mean baseline, endpoint, and reduction in BPRS were, respectively, 48.58, 46.10, and -2.59 (95% CI -4.08, -1.09). The average effect size was 0.27 (-0.44, -0.11) and heterogeneous across studies ($p < 0.001$). Meta-regression analysis showed that greater placebo response was associated with shorter trials ($p < 0.001$), community hospital (or mixed) treatment settings ($p = 0.02$), more recently published studies (1990-2009) ($p < 0.01$), and higher baseline severity score ($p < 0.01$). Analysis of patient-level PANSS total score in the ziprasidone long-term study, however, showed no improvement over a 1-year period in the higher baseline PANSS subgroup using GMM. Analysis of the placebo arms in the 2 short-term ziprasidone trials showed placebo responses in SAD bipolar patients were significantly lower than in schizophrenia patients. In the long-term 1-year trial (1), GMM identified 4 classes of placebo response patterns for PANSS total score: 1) immediate worsening class (15%), 2) gradual worsening class (19%), 3) delayed worsening class (31%) in which patients experienced no change in symptoms for about 16 weeks and gradual worsening thereafter and 4) no change in symptoms over the 1-year study period (35%).

Conclusions: Our findings suggest that treatment settings, trial duration, schizoaffective bipolar diagnosis, and baseline level of symptom severity could influence the magnitude of placebo response.

Supported by funding from Pfizer, Inc.

Use of a Linked Hospital Admissions and Healthcare Claims Database in Pharmaceutical Outcomes Research: Results of a Feasibility Study Examining Treatment of Bipolar Mania with Atypical Antipsychotics

Ariel Berger, MPH1, Kafi N. Sanders, MPH2, Jose Alvir, DrPH2, Marko A. Mychaskiw, RPh, PhD2, Angie Qin, MS1, and Gerry Oster, PhD1

1 Policy Analysis Inc., Brookline, MA, 2 Pfizer Inc., New York City, NY

Introduction: In pharmaceutical outcomes research using healthcare claims databases, periods during which patients are hospitalized constitute “black holes”, as these databases do not contain any information on pharmacotherapy received in hospital. While admission-level databases can provide such information, they lack information on pharmacotherapy received outside of hospital. Recently, it has become possible to link these two types of databases. In this study, we explored the potential value of such linked databases to pharmaceutical outcomes research, focusing attention on the use of second-generation antipsychotics (SGA) in patients with bipolar mania.

Methods: Using a linked inpatient/outpatient database, we identified all adults with ≥ 1 admissions for bipolar mania (principal diagnosis [ICD-9-CM] of 296.0X, 296.1X, 296.4X, 296.6X, 296.81) between 1/1/2004 and 9/30/2008. Focusing on each patient’s first admission, we compiled all healthcare claims during the 6-month periods preceding and following hospitalization. As our interest was in the use of SGAs, we limited our attention to patients with evidence of inpatient receipt of oral ziprasidone, aripiprazole, or quetiapine (“study agents”) immediately preceding hospital discharge. We then examined receipt of these agents during the 6-month periods preceding hospitalization and following hospital discharge based on outpatient pharmacy claims. Adherence with study agents following hospital discharge was assessed using proportion of days covered (PDC); patients were deemed nonadherent if PDC fell below 80%.

Results: A total of 67 patients with bipolar disorder were identified who met all study entry criteria. Twenty-two patients (33%) had evidence of receipt of a study agent in the period preceding hospitalization. While all patients had evidence of receipt of study agents following hospital discharge, only 15% were adherent at 6 months.

Conclusion: Linked inpatient/outpatient databases are a promising avenue for future pharmaceutical outcomes research, as they may greatly expand our understanding of the complete chronology of pharmacotherapy and associated outcomes--for many disease conditions.

Source of Funding: Pfizer Inc., New York, NY.

A Meta-Analysis of Computerized Assessment Batteries in Alzheimer’s Disease Treatment Trials

Henry J. Riordan1, Neal Cutler1, Farzin Irani2, & Paul J. Moberg2

1Worldwide Clinical Trials, Beverly Hills, CA.; 2Department of Psychiatry, University of Pennsylvania, Philadelphia, PA

Introduction: Although there have been qualitative reviews of computerized assessment batteries (CAB) in aging and AD there are no quantitative meta-analyses reportedly due to the heterogeneity across studies and CABs. A systematic quantitative overview of the relationship between CABs and AD treatment was established by searching medication trials with a pre- and at least one post-treatment CAB.

Methods: Studies were identified through a search of the MEDLINE, PsychINFO, and PubMed databases. Each CAB website was also examined for relevant publications and test developers were contacted for any relevant publications. CABs included the CDR, CANTAB, CNTB, Cogtest, Cogstate-CT, CNS Vital Signs, Headminder, Integneuro, Mindstreams Neurotrax, Neurocog Trials BAC, Automated Neuropsychological Assessment Metrics and the Penn CNP. When possible, specific tests from each CAB were extracted and grouped into domains of neuropsychological function (processing speed, working memory, etc.). These domains were then collapsed further (i.e., verbal learning + visual learning + face memory = memory) and effect sizes (ES) were calculated.

Results: Analysis of intervention effects on cognitive status (across different medication types) revealed a large overall effect size ($d=0.89$, 95% CI= $1.07 < \delta < 0.65$) that was significantly heterogeneous (QB[32]=291.4, $p < 0.001$). Of the three batteries with the largest associated effects, the CDR had the largest effect size (ES=.91) yielding stronger effects than the CNTB (ES=.85) or the CANTAB (ES=.74) batteries. Measures of attention and working memory yielded the biggest effect sizes with regard to treatment ($d=0.81$, 95% CI= $0.99 < \delta < 0.69$) while learning and delayed recall yielded more moderate effects ($d=0.52$, 95% CI= $0.71 < \delta < 0.38$) and motor function, strategy use and problem solving tasks yielded small to

moderate effects ($d=0.40$, 95% CI= $0.55 < \delta < 0.13$). A small but significant effect of age on effect size ($p=.043$) (i.e., the older the subject the smaller the effect) was noted as was a small but significant effect of education ($p=.051$) (i.e., the more education the larger effect).

Conclusions: This meta-analysis demonstrates that it is possible to determine which CABs and subsequent cognitive domains yield the largest effects with regard to AD treatment (sensitivity), as well as to identify the moderating variables that contribute to, or detract from, this sensitivity.

11 **Can Cochrane Reviews Be Used as a Tool in Setting Priorities for Clinical Trials in Alzheimer's Disease?**

Soares-Weiser K1, Dickersin K2, Smith N1, Davidson M3, Adams CE4

1 Enhance Reviews Ltd, London, UK; 2 US Cochrane Center, Johns Hopkins, Baltimore, USA; 3 Department of Psychiatry, Tel Aviv University, Israel; 4 Cochrane Schizophrenia Group, Nottingham University, UK

Introduction: Cochrane reviews are the golden-standard summaries of current evidence on the effects of health interventions. Although an important aim of these reviews is to indicate need for future research, it is unclear if the recommendations in the reviews are considered when new research is being planned. Seventy-three Cochrane reviews deal with Alzheimer disease, and all provide specific recommendations for future research. We evaluated which recommendations were made, and investigated whether there was evidence of consideration of these recommendations in relevant randomized trials (RCTs) published after each review.

Methods: Cochrane Library (Issue 1, 2010) was searched for reviews in 'dementia, Alzheimer disease or cognitive impairment'. Recommendations from reviews were divided into categories – 'general' (more trials are/are not needed) or 'specific' (interventions, trial methodology, clinical relevant outcomes). Directly relevant RCTs, published after each Cochrane review was then sought on Pubmed. Data on methodology, intervention/control, scales and main outcomes were extracted from RCTs and the findings compared with the original research recommendations of the Cochrane review.

Results: Interventions in 40/73 reviews were not evaluated in new RCTs, although 23 reviews specifically recommended further research in the topic. 121 new RCTs were found dealing with interventions described in 33 reviews. Less than 50% of RCTs described methodological details (e.g., allocation concealment, double-blind, sample size calculation). Cochrane reviews recommended use of relevant clinical outcome measures, such as, mortality, quality of life and burden to care-givers, and suggested for scales validated for this specific population to be preferred. Less than 10% of new RCTs reported clinical relevant outcomes, or reported whether the scale has been validated.

Conclusion: There are some encouraging signs of the effect of past research on future trials, but they are few. Seeking the best evidence, systematically investigating failures and triumphs of the past, and using all this to inform future study design would seem sensible, but is not yet a reality. Cochrane reviews are of good quality and readily available and should be taken into consideration when conducting a new RCT.

12 **Apparent Effect Size as a Function of Sample Size in FDA-Approved Antidepressants**

Michael Gibertini

PhD INC Research Austin, TX

Introduction: The high failure rate of contemporary antidepressant trials has spurred creative attempts to understand the factors that affect placebo response, error variance, and trial sensitivity. Recently, Turner et al. (2008) presented data from the entire registration trial database of the FDA for 12 contemporary antidepressants. They showed convincingly that compared to the complete database obtained from the FDA review files, the published literature is severely biased, overestimating effect sizes for most antidepressants.

Methods: The Turner compilation is extended and re-examined to help understand the likely population effect size for contemporary antidepressants. Effect sizes on all registration efficacy trials for 13 antidepressants approved by the FDA since 1987 were compiled and analyzed by class of compound, and relationship to study sample size.

Results: The underlying effect size for approved, contemporary antidepressants is, by this analysis, 0.33. There are no differences in effect size among the classes (SSRI, SNRI, other) of compounds examined. Finally, sample size is a major determinant of the likelihood of a trial correctly estimating the underlying antidepressant effect size.

Conclusion: The current results are especially relevant to researchers planning phase II trials for

compounds with expected effect sizes similar to contemporary approved antidepressants. For such future trials, a reliable estimate of effect size, undisturbed by publication bias, will be critical in determining sample size. In addition, the demonstration that small trials will most often under- or over-estimate true effect size should help refine arguments for correct power determination. Hopefully, this re-examination of the FDA database will inform efforts to improve trial design, increase trial sensitivity and decrease the overall trial failure rate. Reference Turner EH, Matthews AM, Linardatos E, et al. Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. *N Engl J Med.* 2008;358:252-60.

13 **Psychometric Properties and Feasibility of the M-3 Checklist: a Brief, Self-rated Screen for Depressive, Bipolar, Anxiety, and Posttraumatic Stress Disorders in Primary Care**

Bradley N. Gaynes, MD, MPH; Joanne DeVeauh-Geiss, MA; Hongbin Gu, PhD; Sam Weir, MD; David R. Rubinow, MD.

University of North Carolina School of Medicine, Chapel Hill, North Carolina

Background: Mood and anxiety are the two most common psychiatric disorders seen in primary care, yet they remain under-detected and under-treated. Screening tools can improve identification, but available instruments are limited by the number of disorders assessed.

Objective: To assess the feasibility and diagnostic validity of the M-3 checklist, a new one-page, patient-rated, 27-item tool developed to screen for multiple psychiatric disorders in primary care. Design, Setting, and Participants: A sample of 647 consecutive participants 18 years and older who were seeking primary care at an academic family medicine clinic between July 2007 and February 2008. A two-step scoring procedure was utilized to make screening more efficient.

Main Outcome Measure(s): Sensitivity and specificity of the M-3 for Major Depression; Bipolar Disorder; Anxiety Disorders; and PTSD. Using a split sample technique, analysis proceeded from determination of optimal screening thresholds to assessment of the psychometric properties of the self-report instrument using the determined thresholds and the Mini International Neuropsychiatric Interview as the diagnostic standard. Feasibility was assessed with patient and physician exit questionnaires.

Results: The depression module had a sensitivity of 84% and a specificity of 80%. The bipolar module had a sensitivity 88%, and a specificity of 70%. The anxiety module had a sensitivity of 82% and a specificity of 78%, while the PTSD module had a sensitivity of 88% and a specificity of 76%. As a screen for any disorder, sensitivity was 83% and specificity was 76%. The M-3 took patients less than 5 minutes to complete in the waiting room, and less than 1% reported not having time to complete it. 83% of clinicians reviewed the checklist in ≤ 30 seconds, and 80% thought it was helpful in reviewing subjects' emotional health.

Conclusions: The M-3 is a valid, efficient, and feasible tool for screening multiple common psychiatric illnesses, including bipolar disorder and PTSD, in primary care. Its diagnostic accuracy equals that of presently used single disorder screens but with the additional benefit of being combined into a one-page tool. The M-3 potentially can reduce missed psychiatric diagnoses and facilitate proper treatment of identified cases

(Supported by M-3 Systems)

14 **PANSS item reliability: Can Standardized Rater Training Improve Negative Subscale Item Reliability?**

Mark Opler, PhD, MPH¹, Sofija Jovic, PhD¹, Stacy Liechti, PhD¹, Christian Yavorsky, PhD¹, Jessica Gordon, MA¹, Larry Yang, PhD¹

ProPhase, LLC¹

Introduction: Paloa et al. (1994) identify the assessment of negative symptoms as important in the assessment of change in overall psychopathology in schizophrenia. The accurate assessment of negative symptoms is also critical in the development of treatment for this debilitating, often under treated and recalcitrant feature of schizophrenia. However the assessment of these symptoms is difficult and numerous studies (e.g., Möller, 2007; Betsen et al, 1996; Norman et al, 1996) have found that negative subscale items in the PANSS are somewhat less reliable than those in the positive and general subscales. Rater error, the subtlety of these symptoms, and the subjective nature of many of these items have been identified as probable culprits. Standardized rater training has been identified as one way to improve overall PANSS reliability. Here we asked if this applied equally to the negative subscale and if these results were durable.

Methods: Results from several large standardized rater training events were analyzed. Training included both applied and didactic components. Raters were asked to rate the PANSS from a video-taped assessment before and after training. The inter-rater reliability was determined using intra-class correlation coefficient (ICC) for pre and post training scores on the negative subscale of the PANSS.

Results: There appeared to be an improvement in the reliable assessment of negative subscale items after training in all cases examined. The ICC range for the PANSS negative subscale before training ranged from .358 (n=11, p<.001) to .615 (n=30, p<.001). The range of post training ICCs were .764 (n=9, p< .001) to .945 (n=30, p<.001).

Conclusion: Based on the improvement in inter-rater reliability it appears that standardized rater training has a significant impact on the assessment of negative symptoms. It is posited that this effect is related to the framework to discuss and engage with the instrument as well as the exposure to clinical examples. Although every rater has unique clinical experience with the target population, the goal of standardized rater training is to encourage the consistent conceptualization of subscale items in the PANSS instrument.

15 **Do Eligibility Criteria Matter? View from a 3rd Blind Eye**

Gary S. Sachs 1, Douglas Vanderberg 2, Daniel Debonis 3, Suzanne Edman 3, Claudia F. Baldassano 4, Bryce Kasuba 5, Louisa Sylvia 6, Eben Goodman 3

1 Concordant Rater Systems, Boston, MA, 2Pfizer, New York, NY, 3Concordant Rater Systems, Boston, Massachusetts, 4University of Pennsylvania, Philadelphia, Pennsylvania, 5Concordant Rater Systems, Philadelphia, Pennsylvania, 6Massachusetts General Hospital, Boston, Massachusetts

Introduction: The high failure rates in CNS studies fuels a search for better solutions. Computer administered assessments offer opportunities to collect data consistently across global study sites and may provide a useful standard for exploring hypotheses regarding the performance of studies and clinical trial sites.

Methods: Computer administered YMRS assessments (YMRSComp) were added to a double-blind protocol investigating low (MS+ZLo) and higher doses of ziprasidone (MS+ZHi) vs placebo (MS+PBO) added to mood stabilizer (MS) treatment for manic and mixed episodes. The primary outcome variable in this protocol was change from baseline to endpoint YMRS score based on the site-based rater's (SBR) YMRS scores. The analysis compares SBR versus computer ratings overall and subgroups defined by computer determination of key eligibility criteria (e.g. At screen and baseline subjects meets DSM IV criteria for a manic or mixed episode, YMRS ≥ 18 and YMRS change from Screen to baseline < 25%). The analysis plan called for one way ANOVA to compare the three treatment groups for both SBR assessments and computer assessments; overall and for comparisons of subgroups defined by the computer assessment of key eligibility criteria.

Results: Of 505 enrolled subjects, <37% met protocol eligibility requirements based on the computer assessments. No statistically significant differences were found between MS+ZL or MS+ZH and MS+PBO based on YMRSSBR or YMRSComp. A numerically larger change from baseline to day 21 (LOCF) was seen among subjects with a valid diagnosis compared to those without a valid diagnosis.

Conclusion: The failure to demonstrate efficacy of adjunctive ziprasidone for acute mania is surprising given its proven efficacy as monotherapy. Several issues may contribute to the lack of observed efficacy. The computer assessment data suggests the study was impacted by enrollment of many subjects that, based on computer assessments, did not meet the protocol eligibility requirements. The results show numerically greater separation for MS+ZL subjects meeting each eligibility requirement than those considered ineligible by the computer assessments.

16 **Effect of Language-Specific Training on Rater Performance in Assessment of PANSS Items and Subscales**

Stacy Liechti, PhD¹, Evgenia Ivanova, MA¹, Sofija Jovic, PhD¹, Jessica Gordon, MA¹, Mark G. A. Opler, PhD, MPH¹, Christian Yavorsky, PhD¹

ProPhase, LLC¹

Introduction: Globalization of CNS clinical trials requires assessments be administered in multiple languages, but monolingualism in rater training methods and performance monitoring persists. This causes biases as training materials and methods may dramatically affect inter-rater reliability assessment. There is a need to develop culturally adequate training materials in relevant languages spoken by investigators. Evidence suggests this will increase the quality of investigator's training, monitoring and ultimately improve studies.

Methods: Data was pooled from a series of training programs for the Positive and Negative Syndrome Scale (PANSS), including English and Russian-language materials. A total of 118 investigators were included. Investigators' performance was compared between groups, including (1) those with significant prior experience (2 years or more) and (2) those with limited or no prior experience using the PANSS. Results from four videos, including two filmed in English with subtitles or with translated transcripts were compared with performance on two videos filmed entirely in Russian (i.e. with Russian-speaking interviewers and subjects). Agreement versus gold-standard scores was calculated within each group of raters for each video, taking experience level and language into account. F-tests were conducted on within-sample variance, and standardized measures of internal consistency (ICC) were taken. Regression models were fitted for item-specific factors.

Results: Within the less experienced group, the average level of agreement with gold standard was relatively low as compared to more experienced raters for both English and Russian-language videos ($p < 0.01$). Negative symptoms demonstrated the lowest scores on measures of agreement ($p < 0.05$) and internal consistency, but continued to demonstrate significant differences when Russian-language materials were utilized, regardless of rater prior experience. Findings suggest that some individual items may be affected by linguistic or cultural factors, while others seem to be dependent on experience.

Conclusion: Our results confirm that assessment in negative symptoms can be improved through the use of professional, adapted, standardized materials in native languages for global investigators. While experience and prior training does affect overall performance, the differences shown in the assessment of negative symptoms and subsets of symptoms from the general psychopathology scale persist. Further research on key symptoms and select items should be carried out.

17 **Placebo Response Assessed by Site and Blinded Remote Centralized Raters in a GAD Trial**

Williams JBW.1,2, Dunn J.3, Kobak KA.1, Giller E.1, Curry L.4, Wilson P.4, Detke M.1,5,6.

1MedAvante, Inc., 2College of Physicians and Surgeons, Columbia University, 3Flexion therapeutics 4Sepracor, Inc 5Indiana University School of Medicine, 6McLean Hospital & Research Associate, Harvard Medical School,

Introduction: High placebo response in anxiety studies contributes to the 50% study failure rate. Remote centralized raters offer a solution that may address several problems. We compared the rate of placebo response assessed by site vs. central raters.

Method: This double-blind, placebo-controlled, multi-center study examined the efficacy and safety of two doses of an experimental compound to treat Generalized Anxiety Disorder. The primary outcome measure was the Hamilton Anxiety Scale (HAM-A). Site raters assessed subjects 6 times over an 8 week period. Remote centralized raters also independently rated subjects on the HAM-A at baseline and week 6. 119 site raters were trained and qualified by United Biosource Corporation (UBC). 22 remote centralized raters were trained and calibrated by MedAvante, and maintained high interrater reliability throughout the study with quarterly group calibrations and regular observations by trainers.

Results: Site raters admitted 122 subjects to the placebo arm of the study. Of these, remote centralized raters would have admitted 59 (48%) and excluded 63 (52%), based on their HAM-A ratings. At baseline, site raters' mean HAM-A score was 24.04 (SD= 3.29; N=122), compared to 19.83 (SD=6.037; N=122) for remote raters. In addition, site raters' mean scores were higher than remote raters' on all of the individual HAM-A items at baseline. At endpoint, remote raters' scores were similar to site raters' (14.70 vs. 13.95). Exploratory analyses found the mean placebo change by site raters was -9.3. This was significantly higher than the -5.9 point mean placebo change as measured by the remote raters. Site raters classified 47 (39%) subjects as placebo responders ($\geq 50\%$ reduction in the HAM-A score), as compared to 29 (24%) by remote raters in the site-admitted cohort (nominal $p = .015$).

Discussion: Blinded remote raters showed a 36% reduction in placebo response compared to site raters. This may be due to the blinded remote ratings being independent from the sites' enrollment process. A decrease in placebo response may improve signal detection in clinical trials. In the absence of a positive control, the impact on effect size is unknown.

18 **Quantifying Rater Drift in an International Sample of Investigators Participating in Standardized Rater Training Events: Is PANSS Reliability Maintained Over Time?**

Christian Yavorsky, PhD¹, Mark Opler, PhD, MPH¹, Evgenia Ivanova, MA¹, Jessica Gordon, MA¹, Sofija Jovic, PhD¹, Larry Yang, PhD¹

ProPhase, LLC¹

Introduction: Most studies that employ psychometric instruments to measure change in pathology over time implement some form of training in order to address the ongoing integrity of ratings following initial training. Rater drift has long been understood to negatively impact trial results but to date there have been few studies that have sought to systematically quantify this. In this study we looked at training data for the PANSS (Positive and Negative Syndrome Scale) from several time points to determine if drift occurred in this sample and if it affects certain items or subscales more than others.

Methods: Raters participating in standardized training scored the PANSS based on a video-taped interview. Data from the initial training session was compared to that obtained during a “refresher” session 8 months later to determine if drift had occurred. Inter-rater reliability was obtained by the use of the Intra-class correlation coefficient (Shrout & Fleiss, 1979) and compared at the two time points. Concordance with gold standard ratings was also compared for the same time points.

Results: Intra-class correlation coefficients (ICC) for raters (n=96) following initial training were in the good range at ICC = .868 (p<.001) and concordance with gold standard rating was high. At the 8 month refresher training session reliability had fallen to ICC = .738 (p<.001) with gold standard concordance only in the moderate range. Positive, negative and general subscale reliabilities fell proportionally and observed/subjective items were most susceptible to decrease in reliability.

Conclusion: In this sample we found that over an 8 month period rater drift did appear to occur as gauged by the metrics of reliability and concordance. This tendency to produce idiosyncratic or incorrect ratings seemed most prevalent in those items that were observed or subjective. Because this drift introduces random error into the data and this can have implications for sample size and power (Muller & Szegedi, 2002), it is imperative that standardized rater training occur at regular intervals.

19

Assessing Interview Quality and Scoring Accuracy in Clinical Trials with Centralized Quality Control & Calibration

Brown BR.1, De Santi S.1,2, Detke M.1,3,4, Williams JBW.1

1MedAvante, Inc., 2NYU Langone Medical Center, 3Indiana University School of Medicine, 4McLean Hospital & Research Associate, Harvard Medical School, 5College of Physicians and Surgeons, Columbia University

Background: CNS clinical trials fail more often than their a priori powering indicates they should. Quality assurance/quality control (QA/QC) safeguards for clinical (including primary) outcome measures have rarely been utilized. The large number of raters performing assessments in multi-site trials increases the possibility of variability in ratings. Rater drift over time is well-documented and common¹, and superior interview performance as measured by the Rater Applied Performance Scale (RAPS) is associated with drug-placebo separation². We report the first findings using Centralized Quality Control & Calibration (CQC), a new approach to monitoring and remediating the administration and scoring of clinical outcome measures.

Method: 17 Calibrated Quality Reviewers were rigorously trained and continuously calibrated on scale scoring and interview quality. This cohort was tightly calibrated on the MADRS, HAM-A and HAM-D, ICCs = .91-.93. Data from multiple on-going clinical trials were pooled. Site raters audio recorded all MADRS, or HAM-A and HAM-D administrations and uploaded the recordings to a central server. A priori scoring accuracy and RAPS interview quality criteria were established. Calibrated Quality Reviewers independently scored 492 site raters’ assessments and rated interview quality using the RAPS. Only after scores and RAPS were submitted was the Calibrated Quality Reviewer given access to the site raters’ scores. Feedback was provided to the site raters on both interview quality and scoring accuracy before their next reviewed assessment.

Results: 492 assessments were reviewed to date. At the first review of 110 site raters, 51% met the a priori criteria for scoring accuracy, 60% for interview quality and 35% met both criteria. By review six or later (n=55) there were substantial improvements: 69% met criteria for scoring accuracy, 87% for interview quality and 65% met both criteria. Improvements generally occurred by month four of the study. Analysis of RAPS domains showed Adherence and Follow-up difficulties commonly compromised interview quality. Rating sleep and sadness presented the greatest scoring challenges.

Conclusion: QA/QC of clinical assessments identified significant scale administration and scoring issues. Repeated feedback improved rater performance substantially. Study outcomes will be evaluated to determine if continuous QA/QC of study assessments assists sponsors in identifying risks that contribute to

CNS trial failures.

References

1. Kobak K, Kane JM, Thase ME, Nierenberg AA. Why do clinical trials fail? The problem of measurement error in clinical trials: Time to test new paradigms. *Journal of Clinical Psychopharmacology* 2007; 27: 534-535.
2. Lipsitz J, Kobak K, Feiger A, Sikich, D, Moroz BAG, Engelhardt N. The Rater Applied Performance Scale: Development and reliability. *Psychiatry Research* 2004; 127: 147-155.

20

Results of an H3 Receptor Antagonist Clinical Trial in Adults Diagnosed with ADHD

Stacey Boyer*, Donna Palumbo*, Phillip Chappell*, Bryce Li*, and Garry Pinter*

**Pfizer Global Research and Development*

Introduction: A potential alternative approach to the therapy of adult ADHD is the blockade of histamine H3 receptors. Histamine plays a role in centrally mediated activities, including memory functions, arousal, and wakefulness. Preclinical data indicates that numerous H3 antagonists are reported to increase wakefulness (e.g. BF2.649 during lights-off (active) period and decreased DREM episodes in orexin (-/-) mice, Ciproxifan in cats). This study aimed to investigate the efficacy and safety of a selective H3 receptor antagonist in improving the symptoms of adults diagnosed with ADHD.

Methods: A multi-site, double blind, placebo-controlled, three treatment, twoperiod crossover trial was conducted with an H3 antagonist. Each subject was randomized to one of the four treatment sequences: (1) 1 mg QD → Placebo; (2) Placebo → 1 mg QD; (3) Flexible (0.5 mg titrated to 2 mg) QD → Placebo; and (4) Placebo → Flexible (0.5 mg titrated to 2 mg) QD. The duration of each treatment period was three weeks with a one-week washout phase between treatment periods. The primary efficacy endpoint was the change from baseline at week three on the adult attentiondeficit/ hyperactivity disorder investigator symptom rating scale total score (AISRS). Two key secondary endpoints were the change from baseline at week three on the AISRS subscales (inattention and hyperactivity/impulsivity).

Results: Sixty-six male subjects were assigned to treatment, with forty completing the study. The study subjects had an average age of 37.4 years and an average baseline AISRS total score ranging from 38.3 to 36.8, indicating a moderate to severe baseline level of symptomatology. Both AISRS scores indicated no statistically significant difference between active treatment and placebo. Insomnia was the all-causality adverse event with the highest frequency of occurrence in subjects receiving treatment.

Conclusions: Treatment appeared to be generally safe and well tolerated as dose limiting adverse events were mostly mild to moderate and spontaneously resolving. This therapeutic approach may not be beneficial in the treatment of symptoms of ADHD in adult patients meeting DSM-IV criteria. However, the increased incidence of insomnia during active treatment periods suggests the possibility of an obscured efficacy signal. These data are suggestive of H3 antagonists potential for the treatment of other CNS-related disorders.

21

White Matter Hypoplasia is a Biomarker for Familial Major Depression

MJ Dubin, MM Weissman, D Xu, R Bansal, HT Zhu, X Hao, J Liu, V Warner, BS Peterson

Introduction: Major depressive disorder (MDD) is a debilitating illness with genetic and environmental causes. Structural and functional neuroimaging data suggest that the frontal lobe plays a significant role in the pathophysiology of familial MDD (Drevets, et.al., 1997). High familial risk for MDD is reflected in frontoparietal cortical thinning, even in high-risk individuals who do not develop depression. (Peterson, et.al., 2009). Herein, we describe white matter hypoplasia correlated with familial risk for MDD.

Methods: We analyzed samples from a three generation study spanning more than 20 years, of families in which the children (G2) and grandchildren (G3) were at either high or low risk for depression. Risk was defined as the major depression status of the first generation (G1) (Weissman, et.al., 2005). Structural images were acquired from a 1.5T scanner. Whole-brain analysis comparing high (N=66) and low risk (N=65) subjects was performed using Volume Preserved Warping, which detects local volume expansion and contraction between groups. (Xu, et.al., 2007). Correction for multiple comparisons was accomplished using False Discovery Rate (FDR).

Results: An extensive region of white matter within the frontal lobe and overlapping significantly with the corona radiata, frontostriatal, and frontolimbic projections, was smaller on average in the G2 and G3 members of the high-risk group. This was also true when G2 and G3 subjects without a history of depression were compared according to risk status. This result survived correction for multiple

comparisons and common environment and continued to be significant when age and gender were used as covariates. Post hoc analyses show that a sub-region correlated with measures of inattention and visual memory.

Conclusions: These findings support an association between frontal white matter hypoplasia and vulnerability to depression, independent of the presence or absence of depression. As such, it is an anatomical endophenotype and is correlated with cognitive measures of inattention and memory. (Cannon, et al. 2006). As the ROI overlaps with frontostriatal and frontolimbic projections, it may contribute to the development of MDD by disrupting white matter pathways critical for emotional and cognitive regulation.

References: Cannon, T (2006), 'Endophenotypes in the Genetic Analysis of Mental Disorders', Annual Review of Clinical Psychology, vol. 2, pp. 267-290. Drevets, W (1997), 'Subgenual prefrontal cortex abnormalities in mood disorders', Nature, vol. 386, no. 6627, pp. 824-827. Peterson, B (2009), 'Cortical Thinning in Persons at Increased Familial Risk for Major Depression', Proceedings of the National Academy of Sciences, vol. 106, no. 15, pp. 6273-6278. Weissman, M (2005), 'Families at High and Low Risk for Depression: A 3-Generation Study', Archives of General Psychiatry, vol. 62, no. 1, pp. 29-36. Xu, D (2007), 'Unifying the Analyses of Anatomical and Diffusion Tensor Images Using Volume- Preserved Warping', Journal of Magnetic Resonance Imaging, vol. 25, no. 3, pp. 612-624.

22

Data Monitoring and Ongoing Individualized Rater Education Diminishes ADAS-Cog Administration and Scoring Errors

Marian A. Ormont MD, Theresa A. Bromley PhD

ePharmaSolutions

Introduction: The ADAS-Cog is the most frequently utilized primary outcome measure in clinical trials targeting Alzheimer's dementia. Significant variance in intra-rater ADAS-Cog experience, administration procedures and scoring guidelines exists. Monitoring ADAS-Cog data during a dementia trial, and contacting raters regarding data errors is important to diminish rater drift, minimize scoring errors, and promote inter-rater reliability. Rater education regarding questionable ADAS-Cog worksheets improves their ability to administer and score the scale as demonstrated by diminished errors over the course of a study.

Methods: 36 raters from Hungary, India, Lithuania and South Africa administered the ADAS-Cog in a clinical trial. Raters' highest degree achieved ranged from BA (11%) to a MA, MSW, RN (34%), to MD or PhD (55%). Prior experience with the ADAS-Cog ranged from none to >50 administrations. The ADAS-Cog was administered at 5 study visits per subject. 597 submissions were reviewed by doctoral level clinicians. 32% of the submissions were found to contain at least one error. The most significant number of errors was found in Remembering Test Instructions (22%), Mazes (10%) and Number Cancellation (9%) respectively. Identified errors were individually communicated to the rater. Raters were educated regarding the correct administration and scoring of the scale, and were also provided with a standardized written description of the item and tutorial. The ratio of ADAS-Cog worksheet errors to ADAS-Cog worksheet submissions will be determined per rater per month. The rate of change in the ratio will be determined per rater. Rates of change will be correlated with country, educational level and ADAS-Cog experience.

Results: Raters administered from 3 to 109 ADAS-Cogs, with an average of 30.4 administrations per rater. Preliminary data indicates that ADAS-Cog errors decreased in a significant number of the raters after individualized clinical contact.

Conclusions: Data monitoring improved raters' abilities to administer and score the ADAS-Cog. Number of errors per rater decreased over time thus improving the data quality. Data Monitoring with individual rater contact is effective in studies utilizing the ADAS-Cog as a primary outcome measure.

23

The MARQ-AD: A New Measure of Interview Quality in Alzheimer's Disease Trials

Weber CJ1, De Santi S1,2, Kirby LC4., Williams JBW1,3

1MedAvante, Inc., Hamilton, NJ, 2NYU Langone Medical Center, 3College of Physicians and Surgeons, Columbia University, 4Critical Path Institute, Phoenix, AZ

Introduction: Clinician-administered cognitive assessments in clinical trials of Alzheimer's disease (AD)

have multiple sources of variability which can collectively diminish or obscure a treatment effect. Both structured (e.g., ADAS-cog) and semi-structured (e.g., CDR) assessments have significant variability in administration and interpretation of subject responses between sites and between raters. This variability in key outcome measures can undermine the detection of a treatment effect in a randomized Clinical Trial. Compensating for low signal detection requires increased numbers of subjects, increased time and costs, and may constitute a significant factor in the recent failures seen in AD drug trials.

The challenges of standardized administration and scoring of AD assessments such as the ADAS-cog have been documented¹. Even experienced raters working on multiple dementia protocols have a substantial likelihood of administration/scoring inconsistency, which decreases signal detection¹. Furthermore, long trials incur increased rater turnover that can exacerbate these standardization issues. This poster presents a new instrument designed to evaluate variability in the administration of AD outcome measures.

Methods: The MedAvante Analysis of Rating Quality - Alzheimer's Disease (MARQ-AD) was developed to quantify critical domains of rater performance and address reasons for administration and scoring variability encountered in AD assessments. The MARQ-AD was designed to assess raters' clinical interview skills, as a clinical trial qualification and as an ongoing performance assessment tool. AD trainers with at least 10 years of AD assessment experience perform the evaluations with the MARQ-AD. The trainers undergo extensive administration and scoring training on the MARQ-AD and study-specific scales being monitored. Trainers are highly calibrated for administration and scoring on each scale.

Results: Trainer calibration and use of the MARQ-AD provides consistent evaluation and feedback on all reviewed assessments, which, in the context of a research clinical trial, makes detection of a true efficacy signal more likely.

Conclusions: The evaluation of rating quality has been effective in other disease states that utilize subjective assessment instruments². The benefit to AD and other dementia clinical trials of the MARQ-AD scale is improved interview standardization and quality, resulting in decreased rater drift over the course of a trial and higher rater concordance.

References:

- (1) Connor D, Sabbagh, M. Administration and Scoring Variance on the ADAS-Cog. 2008 Journal of Alzheimers Disorders: 15:461-4.
- (2) Jeglic E, Kobak KA, Engelhardt N, Williams JB, Lipsitz JD, Salvucci D, Bryson H, Bellew K. A novel approach to rater training and certification in multinational trials. (2007) International Clinical Psychopharmacology.; 22: 187-91.

24

Do Recruitment Sources Impact Study Outcome: Assessment by Computer and Site-Based Raters

Bryce Kasuba¹, GS Sachs², Douglas Vanderberg³, Kathleen Ice⁴, Michael Zelasky⁴, Dan DeBonis⁵, Suzanne Edman⁵, Jenn Sturgis⁵, Eben Goodman⁵

1 Concordant Rater Systems, Havertown, PA, 2Massachusetts General Hospital, Boston, MA, 3Pfizer, New York, NY, 4Pfizer, New London, CT, 5Concordant Rater Systems, Boston, MA

Introduction: Concerns about slow sample accession in clinical trials have prompted a variety of strategies that attempt to bolster recruitment. Recent large studies often extend local recruitment efforts by adding resources from experienced national recruitment vendors. Little data is available however, that compares outcomes for patients recruited from different sources (e.g. usual care providers versus mass media). Speculations about the high failure rates in CNS studies include concern that subjects are not always representative of patients seeking treatment for the disorder under study.

Computer administered assessments offer opportunities to collect data directly from subjects across global sites and provide a useful standard for exploring hypotheses regarding the performance of recruitment strategies. Analysis of data comparing ziprasidone to placebo on the site-based rater's (SBR) and computer ratings found no significant differences between the treatment groups. This presentation provides the outcomes for subjects based on the referral source.

Methods: Computer administered assessments (Diagnostic Validation, Ham-DComp, MADRSComp) were included in a double-blind protocol that randomized 303 Bipolar I subjects to adjunctive treatment with ziprasidone vs placebo for acute depression. The primary outcome variable was change from baseline to endpoint MADRSSBR score. Computer diagnostic assessment interviews were scored (0-100) using the Bipolarity Index, a measure of diagnostic confidence. Quality ratings were defined based on the absolute value of the difference between the MADRSComp and MADRSSBR at baseline and categorized on a 0-4 ordinal scale.

Results:

Referral Source	N	Baseline Ratings						Endpoint Ratings (LOCF)		
		Ham-D		MADRS		Bipolarity Index	Rating Quality	Change from baseline MADRS		Improvement
		SBR	Comp	SBR	Comp			SBR	Comp	
Psychiatry office	85	25.3	27	28.4	28.5	73.3	0.87	-9.6	-8.5	30.50%
Other medical office/posting	28	24.9 (3.4)	28.1 (7.1)	28.8 (5.3)	29.4 (7.3)	72.7	0.39	-11	-11.3	39.30%
Mass media (print, radio, TV)	10	24.9 (3.0)	23.2 (4.2)	29.4 (5.1)	28.1 (11.4)	71.5	1.3	-18.9	-16.6	50.00%
Support (personal/group)	5	23.2 (3.9)	26.2 (6.9)	30.6 (4.3)	31.0 (10.9)	71	1	-14.5	-14.3	54.80%
Other	15	24.7 (3.2)	27.9 (5.7)	30.1 (5.8)	33.4 (9.8)	68.3 (11.0)	0.93	-16.3 (12.4)	-15.4 (15.4)	33.30%

Conclusion: Subjects that entered the trial based on the recommendation of the psychiatrists' office or another medical office had tandem SBR and computer assessment ratings that were more concordant than subjects responding to mass media. These subjects also had Bipolarity Index scores demonstrating greater confidence in a Bipolar I diagnosis.

Computer assessments provide an opportunity to directly collect information from subjects that might facilitate design and testing of interventions to mitigate the potential for study failure.

25 Heritability in cognitive performance in large-scale studies can be assessed by computer-based testing

Aaron S. Hervey, Kathryn Greenfield, Sandeep Vaishnavi, C. Thomas Gualtieri

NC Neuropsychiatry Attention & Memory Centers, Chapel Hill, Charlotte, and Raleigh, NC

Introduction: Many psychiatric diseases, such as schizophrenia and Attention Deficit Hyperactivity Disorder (ADHD), are complex polygenic disorders with cognitive deficits. Quantitative trait locus analyses in genome-wide scans may be particularly useful to study such disorders. Large sample sizes are necessary for these studies, and it is necessary to have an efficient way to measure cognitive variables. Here, we test the hypothesis that computerized neurocognitive testing will be able to efficiently identify heritability in cognitive performance in a large psychiatric sample.

Methods: A total of 267 parent-child dyads (N=534) were obtained from a database of results from a computerized neurocognitive test battery, CNS Vital Signs (CNSVS). Subjects with acquired brain disorders were excluded, allowing for focus on polygenic psychiatric disorders such as mood disorders, schizophrenia, and ADHD. Correlations were determined between parent-child dyads, as well as between the same child group and non-related adults. The non-related adults were matched to the parent group for age, gender, education, and race. Univariate regression analyses were done to determine the magnitude of performance in children accounted for by their parents.

Results: Multiple significant positive correlations in neurocognitive test performance were found in parent-child dyads, in 14 standard score test domains, including an overall composite index, as well as the domains of Psychomotor Speed, Cognitive Flexibility, Nonverbal Reasoning, Working Memory, and Executive Function (all p 's < .001). By comparison, significant positive correlations were found between the matched parent and child groups for only one domain: Psychomotor Speed ($p < .01$). For each of the domains found to be significantly correlated, parent performance accounted for 5-10% of child performance. Parent performance accounted for a greater proportion of variability in child performance than did non-parent adult performance for each domain.

Conclusions: These findings are consistent with the results of twin studies that demonstrate significant heritability in tests of general neurocognitive ability. Computerized testing is an effective tool to efficiently identify heritability in cognitive performance, and may be useful in large-scale studies that attempt to identify genetic bases of polygenic diseases and predict response variability in pharmacological trials.

26 An Interview Guide for the Scale for Assessment of Negative Symptoms (IG-SANS)

¹Joep Schoemaker, MSc, ²Richa Gaur, PhD, ³Luis Ramirez, MD, ⁴Nathan Lee, MSc & ⁴Hossein Kaviani, PhD.

1 Schering-Plough, Oss, The Netherlands, 2 The Cognition Group, Gurgaon, India, 3 Quality Outcomes Training, Ohio, USA, 4 The Cognition Group, Dartford, United Kingdom.

Introduction: The Scale for the Assessment of Negative Symptoms (SANS) is regularly used as a primary outcome measure in clinical trials. In order to attain sufficient inter-rater reliability on SANS it is important that raters are trained on the scale, as well as patient interviewing skills. Available instructions on the use of SANS only provides an explanation of individual SANS symptoms but do not contain a set of questions for proper evaluation of these symptoms. For the use of SANS in a multicenter, randomized, clinical trial (RCT), an interview guide was developed to increase scale item reliability

Methods: The guide was developed by a schizophrenia expert; it contained twelve questions to obtain the information needed to rate the 25 items of the SANS. Three raters reviewed and provided face validity. The SANS, using the interview guide, was administered and video captured in 214 subjects. The videos were rated independently by trained site raters (SRs) and independent raters (IRs) at baseline and endpoint. The IR also used the rater performance scale (RAPS) to evaluate the “adherence to the structured interview guide” by the SR. Only interviews that followed the interview guideline were used to assess the test–retest reliability between SR and IR using an intra-class correlation

Results: The overall quality of interviews for SANS was rated by IR as ‘good-excellent’ in the majority of cases, with no meaningful differences for interviews done at baseline and endpoint. The Inter-rater reliability was found to be substantial for the composite score and moderate for all global items

Conclusions: The development of an interview guide for SANS is a first attempt to assist raters in querying negative symptoms of schizophrenia in a consistent manner. IG-SANS facilitated rater training on a scale they were not much familiar with, providing investigators with explicit instructions and specific interview questions. An innovative test-retest method allowed close monitoring of rater performance during the conduct of a clinical trial. With this method it could be confirmed that substantial Inter-rater reliability was attained on the SANS rating scale for notoriously difficult to rate (negative) symptoms of schizophrenia.

27

Computerized Cognitive Testing in Idiopathic PD with Motor Fluctuations: Patient Experiences.

Smita Pandey Bhat1 , Ph.D., Rita Kakati 2 , Saurabh Gupta 1 , MBBS, DPM, Amit Sharma 1, Geetika Nath 1, Hossein Kaviani 2, Ph.D. Susan De Santi 3, Ph.D.

1The Cognition Group, Gurgaon, India, 2 The Cognition Group, Dartford, U.K, 3NYU School of Medicine, Centre for Brain Health, New York, USA.

Introduction: Cognitive impairment in Parkinson’s disease (PD) is well established. Computerized cognitive batteries have been used with PD in clinical trials. The ability of PD patients to use such tests has not been formally examined. We examined early to late PD patients’ capacity to perform computerized cognitive testing. We measured the percent of PD patients failing to complete the battery, identified the specific tests not completed, and reasons for incomplete tests.

Methods: 854 PD patients ranging in duration from 2 to 20 years, participating in a multinational clinical trial were administered; six tests from the Cogtest library including Auditory Number Sequencing (ANS), Symbol Digit Substitution (SDS), Strategic Target Detection (STDT), Spatial Working Memory (SWM), Tower of London (TOL), and Word List Memory (WLMT). Four assessments took place (screen, baseline, week-12, and week-24) over a course of six months. Tests not completed were coded electronically with explanations.

Results: On average 11% of PD patients failed to complete testing; 13% at screen, 10% at baseline, 11% at week-12, and 10% at week -24. The reasons reported for failing to complete testing included; fatigue (4%), refusal to continue experiment (3%), equipment problems (2%), and unable to understand test concept (1%). The tests difficult to perform were SWM (2%), TOL (2%), WLMT (1%), and STDT (1%) suggesting problems in working memory, executive function, and memory.

Conclusion: 89% of PD patients at varying severity levels successfully completed computerized cognitive testing (Cogtest) suggesting this battery can be efficiently incorporated into clinical trials.

28

A Multicenter, Add-On Trial of D-Serine for Negative and Cognitive Symptoms of Schizophrenia

Mark Weiser^{1,2, 11,13,14}, Uriel Heresco-Levy⁴, Michael Davidson², Daniel C. Javit,³ Nomi Werbeloff³, Ari A. Gershon³, Yehuda Abramovich^{5, 6}, Daniela Amital ⁷, Adiel Doron^{8, 9, 10}, Shai Konas, Yehiel Levkovitz¹¹, David Liba Alexander Teitelbaum Yosef Zimmerman¹²,

1Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, 2Chaim Sheba Medical Center, Tel Hashomer, Israel, 3Ezrath Nashim-Herzog Memorial Hospital, Jerusalem, Israel, 4Beer Ya’akov Mental Health Center, Beer Ya’akov, Israel, 5Geha Mental Health Center, Petah Tikva, Israel, 6Kfar Shaul/Eitanim Mental Health Center, Jerusalem, Israel, 7Lev Hasharon Mental Health Center, Pardesia, Israel, 8Mazra Mental Health Center, Acre, Israel, 9Nes Ziona Mental Health Center, Nes Ziona, Israel, 10Shalvata Mental Health Center, Hod Hasharon, Israel, 11Jaffa Mental Health Center, Jaffa, Israel, 12Nathan Klein Institute, Orangeburg, New York, 13New York University, New York

Background: Observations that antagonists of the NMDA receptor of glutamatergic neurons can mimic symptoms of schizophrenia have raised the hope that NMDA agonists can improve symptoms. Based on encouraging but tentative results of trials in which NMDA agonists were added to anti-psychotics drugs we conducted an adequately powered RCT adding d-serine an NMDA modulator to antipsychotics.

Methods: This was a 195 patient, multicenter, double-blind, randomized, placebo-controlled 16 week trial of d-serine 2 gm/day as an add-on to anti-psychotics. Subjects had schizophrenia or schizoaffective disorder, inpatients and outpatients, stabilized on anti-psychotics, with persistent negative symptoms. The primary outcomes measures were changes in negative symptoms and cognition as measured by the Scale for the Assessment of Negative Symptoms (SANS) and the MATRICS battery respectively.

Results: Both d-serine and placebo group improved on the SANS and MATRICS scores but there were no significant differences between groups: (d-serine 12.9% , placebo 17.3% improvement, $F[1, 147] = 1.16, p=0.28$; and d-serine 7.0% , placebo 6.1% improvement, $F [1, 125] = 0.65, p=0.42$] respectively. D-Serine was well tolerated .

Discussion: This study does not support the use of add-on d-serine at 2 gm/day for schizophrenia; however, the results are limited by a large placebo response,. Future studies will administer higher doses, and/or attempt to effect the NMDA receptor using other mechanisms, such as agonists of the pre-synaptic mglu2/3 receptor, or glycine re-uptake inhibitors.

29

Methodological Considerations Underlying Demonstrated Antidepressant Efficacy of Extended Release Quetiapine Fumarate Monotherapy

Kalali A,¹ Montgomery S,² McIntyre RS,³ Earley WR,⁴ Nyberg S,⁵ Eriksson H⁵

¹Quintiles Inc., San Diego, and University of California, San Diego, CA, USA; ²Imperial College School Of Medicine, University of London, UK; ³Department of Psychiatry, University Health Network, University of Toronto, Toronto, Ontario, Canada; ⁴AstraZeneca Pharmaceuticals, Wilmington, DE, USA; ⁵AstraZeneca R&D Södertälje, Sweden

Introduction: Pivotal registration placebo-controlled trials for currently approved antidepressants are successful <50% of the time underscoring the high rate of negative/failed trials in new drug discovery.¹ However, the clinical development program for once-daily extended release quetiapine fumarate (quetiapine XR) in patients with major depressive disorder (MDD) demonstrated antidepressant efficacy in 80% of acute monotherapy studies. Therefore, some design features of these studies may merit further investigation.

Methods: Quetiapine XR monotherapy (50-300 mg/day) was evaluated in five randomized, double-blind, placebo-controlled, parallel group studies. Two studies employed fixed-dose and two studies used modified fixed-dose strategies; one study used flexible dosing. Two studies included active controls to ensure assay sensitivity. Inclusion criteria stipulated that patients have HAM-D \geq 22. Presence of anxiety symptoms at screening did not lead to exclusion; however, patients with Axis-I/Axis-II disorders were excluded. Patients were randomized equally across treatment groups.

Results: 4/5 of acute monotherapy studies were significantly in favor of quetiapine versus placebo on the primary outcome measure (Figure 1); one was a failed study where both quetiapine XR and active control (escitalopram) failed to differ significantly from placebo. Sites in non-elderly monotherapy studies were monitored with an intent to recruit \geq 40% patients with HAM-D \geq 28. Study populations had moderate-to-severe depression, thereby increasing the likelihood that depressive symptoms were of sufficient severity to minimize placebo response. Potential rater inflation or inherent bias of the scales was potentially counteracted by using different scales for inclusion criteria (HAM-D) and outcome measures (MADRS).

Conclusions: Several factors, including disease severity, are reported in the literature to be influential in determining antidepressant efficacy in studies. The successful quetiapine XR clinical development program in MDD included several methodological maneuvers that may have enabled its success, including enrolling only patients with moderate-to-severe depression, presence of anxiety symptoms, exclusion of Axis-I/Axis-II comorbidities, minimizing rater inflation or potential bias of scales, fixed dosing, and multiple arms. Comparison between methodologies/populations from this clinical development program and other antidepressant registration trials is expected to provide valuable insights, and may lead to revised recommendations for design and conduct of future antidepressant studies.

Reference

1. Khan A, Schwartz K. *Essent Psychopharmacol* 2005; 6: 221-226.

Funded by AstraZeneca Pharmaceuticals

One or more authors report potential conflicts which are described in the program

Preferred presentation type: poster

The Disruptions to Cognition, Everyday Function and Quality of Life in Oncology Patients: A Therapeutic Opportunity?

Keith A. Wesnes, Chris Edgar and Helen Brooker

United BioSource Corporation, Goring.

Introduction: UK Background Patents with a variety of forms of cancer often report cognitive difficulties. Cognitive testing has confirmed the presence of impairments to attention and memory in several oncology populations. However there has been little systematic study of these deficits or of their relationships to everyday function and quality of life.

Methods: The study population comprised 432 females (mean age: 57.6 years, range 18 to 86) and 456 males (mean age: 60.6 years, range 18 to 88), diagnosed with either Hodgkin's Disease (n=136), Non-Hodgkin's Lymphoma (n=348), Chronic Lymphocytic Leukemia (n=141) or Multiple Myeloma (n=263). An automated cognitive test battery (the CDR System) was administered, yielding measures of attention, working and episodic memory. The ECOG Performance Status Score as well as Quality of Life (QOL) scales including the FACT-AN and the Cancer Linear Analogue Scales (CLAS) were also administered. Analysis The profile of cognitive deficits was determined by comparison to the CDR System normative data base of over 6000 healthy individuals. The deficit profile in the patients over core domains of function was determined using Cohen's d effect sizes in five age cohorts. Functional ability on the ECOG was related to cognitive function in these age cohorts, and the relationships between cognitive function and various QOL measures were also examined.

Results: Patients in all age bands showed marked impairments to the Power of Attention measure from the CDR System, with large effect sizes (Cohen's d range 1 to 1.6), together with notable slowings in the speed of retrieval of information from memory (range 1.1 to 1.7). The older patients also showed deficits in other aspects of attention, working and episodic memory. There were clear associations between cognitive function and various quality of life indices. A good overall relationship was identified between the ECOG Performance Status Score and cognitive function; the difference between those patients with scores of 0 (Fully active, no restriction) and 2 (unable to work) showed a strong relationship to attention in the entire population (d=0.8), and again grew with age.

Discussion and Conclusions: There is a clear cognitive deficit profile in cancer patients, and with ageing both the range and magnitude of these cognitive deficits increase. The strong relationship of the CDR attentional measures to the everyday function is consistent with findings in other clinical populations (eg dementia, Bronnick et al, 2006). Further, the disruptions to attention and aspects of memory were associated with reductions in various QOL measures. Cognitive function is a therapeutic target in many diseases, and the extents of the deficits seen in this study make such pharmaceutical intervention an attractive proposition in oncology. Early results with modafinil and epoetin alfa in oncology patients have been encouraging in this respect (Kohli et al, 2009; Littlewood et al, 2001; Tesch et al, 2006), suggesting that there is therapeutic potential for improving cognitive function, quality of life and everyday function in patients with various forms of cancer.

Improvements to Cognitive and Visuo-Perceptual Task Performance in Dementia with Lewy Bodies with Galantamine

Chris J Edgar¹, Keith A Wesnes¹, Andrea Zangara², Keith Edwards³ ¹United BioSource Corporation, Goring-on-Thames, UK

²Flordis Natural Medicines, St. Leonards, Australia; ³Alzheimer's Diagnostic and Treatment Center, Bennington, Vt, USA.

Background: The McKeith Consensus Criteria for Dementia with Lewy Bodies (McKeith IG et al, Dementia with Lewy bodies. Lancet Neurology 2004, 3: 19-28) identified deficits to attention and visuospatial ability to be especially prominent central features of the condition. Galantamine is effective in the symptomatic treatment of Alzheimer's disease (AD). In addition to acetylcholinesterase inhibition, galantamine acts as an allosteric modulator of neuronal nicotinic ACh receptors. Recent studies have suggested that this additional nicotinic action may underlie the early on-set benefits of galantamine to aspects of attention seen in AD patients (Vellas et al, Curr Med Res Opin 2005, 21: 1423-1429), which have been shown to significantly exceed those of donepezil (Galvin et al, Alzheimer's Disease and Related Disorders 2008 22: 30-38).

Methods: This was a multicenter, investigator-initiated, open-label, flexible-dose (8-24 mg/day), 24-week study of galantamine in 50 patients with mild to moderately severe Dementia with Lewy Bodies (DLB). Mean age was 76.5 years (range 50 to 91). There were 29 male and 21 female patients, and the mean MMSE at baseline was 20.8 (range 7 to 30). The CDR System [www.unitedbiosource.com] was

administered to assess specific aspects of cognition including major aspects of attention, working memory and episodic memory. In addition the CDR Visuo-perceptual battery was administered to assess aspects of perception and the discrimination of various visual images.

Results: All aspects of cognitive function assessed with the CDR System improved with treatment; these reaching statistical significance for Power of Attention at 12 weeks and Quality of Episodic Memory at 12 and 24 weeks. Small effect size (0.2-0.46) improvements were seen to other aspects of cognitive function and to performance on the visuo-perceptual battery.

Discussion & Conclusions: The improvements to Power of Attention and Quality of Episodic Memory in DLB seen in this study have been seen previously with other anticholinesterases (eg McKeith et al, 2004). The improvements to visuo-perceptual performance suggest that such assessments may be useful in future work. These findings support the recommendations of McKeith et al (2004) that outcome measures in DLB trials should include specific measures of attention and visual perception. The similarity of the attention deficits in DLB to those in Parkinson's disease dementia (PDD) has been recognised (McKeith et al, 2004), and rivastigmine was found to produce a marked improvement on the CDR Power of Attention score in the pivotal registration trial for the compound in PDD (Emre et al, New England Journal of Medicine 2004, 351: 2509-2518). The nicotinic action of galantamine may suggest that the various nicotinic receptor agonists under development could be useful in treating the attentional deficits in both DLB and PDD. Funded by an unrestricted grant from Jansen-Cilag (protocol GAL-DLB-OL)

32

The Implications of the Cognitive Deficit Profile in Schizophrenia for Therapeutic Strategies

Howard Hassman¹, Richard Fruncillo¹, Steven Glass¹,
David Krefetz¹, Maria Pinho¹, Keith Wesnes²

¹CRI Worldwide, Mount Laurel, NJ, United States, ²United BioSource Corporation, Goring on Thames, Oxon, United Kingdom

Background: There is a large body of work showing a range of impairments to cognitive function in schizophrenia. The MATRICS initiative has considered this research, and proposed a number of domains of cognitive function which are potential targets for treatment, including speed of processing, attention/vigilance, working memory and verbal/visual learning.

Methods: The CDR System comprises a set of automated tests of cognitive function and has been extensively used in worldwide clinical trials. The various tests enable a detailed evaluation of information processing and attention, as well as various aspects of executive control, working memory and episodic memory. In this study the system was administered to 60 males and females aged 18-65 years with a primary DSM-IV diagnosis of schizophrenia or schizoaffective disorder.

Results: The CDR System identified a wide range of cognitive deficits, the largest being to the ability to focus attention and the time taken to retrieve information held in both working and episodic memory. All other parameters of attention were compromised, including early stages of attention, cognitive processing, sustained attention and variability of attention. Further the ability to hold and retrieve information in both working and episodic memory was compromised. These deficits, particularly the attention deficits were related to the severity of illness.

Discussion: This work has confirmed previous findings with computerised tests in schizophrenia that a broader range of impairments to attention exists than is generally recognized. Further, the evidence of notably slowed retrieval of information from working and episodic memory in schizophrenia is commonly overlooked with non-automated tests.

Conclusions: For pharmaceuticals which may not have a broad potential cognitive enhancement profile, one or more of the major deficits identified in this work could be considered as the initial therapeutic targets. However, even for those compounds which may be expected to have broad benefits, it may be sensible to restrict the primary cognitive outcomes to one or two major domains.