ISCTM 2015 Autumn Conference

27-29 August 2015

Marriott, Amsterdam, the Netherlands

Poster Abstracts

1

The DSM-5 MDD Anxious Distress Specifier: Useful Predictor of Risk: Suicide, Comorbidities, Disability & Treatments?

Wilcox MA[1], Kent J[1], Canuso C[1], Wittenberg G[1]

[1]*Janssen Pharmaceutical Research & Development*

Methodological Question Being Addressed: Is the DSM-5 "With Anxious Distress" specifier a clinically relevant risk indicator in MDD?

Introduction/Aim: Patients with major depressive disorder (MDD) and comorbid anxiety may be at increased risk for suboptimal outcomes. The DSM-5 acknowledges the clinical significance of anxiety with the new anxious distress specifier (ADS) in MDD.
The aim of this work is to describe the relationship between the ADS and: demographics, onset and burden of MDD and anxiety, other psychiatric conditions, suicidal behaviors, disability, and treatment utilization.

Methods: Data from the National Comorbidity Survey-Replication (NCS-R), a representative sample of the U.S. were used. Participants with a lifetime diagnosis of a Major Depressive Episode (MDE) (N=1,091) were included. Constructs included in the DSM-5 ADS were identified using symptoms endorsed during the worst MDE. This allowed the examination of co-occurrence of depression and anxiety symptoms within an episode. Four ADS constructs (feeling keyed up or tense; feeling unusually restless; difficulty concentrating because of worry; fear that something awful might happen) were mapped to items in the NCS-R. The 5th construct (feeling one might lose control of self) did not adequately correspond to any item in the Depression section of the NCS-R. We report the results of a modified specifier comprised of 4 items. Both ordinal (0-4 symptoms) and binary (no anxious distress vs. anxious distress: ≥2 items endorsed) measures are included.

Results: There were no meaningful differences in demographic characteristics across MDD ADS groups. Significant differences ($p<0.05$, **bold** in the table) were found in several clinical characteristics between the ADS and non-ADS groups (Table 1).

	- ADS	**+ ADS**		**- ADS**	**+ ADS**
Disease Burden			Suicide		
Past 12-month MDE	35.3%	**46.9%**	LT attempt	11.0%	**19.0%**
% years w MDE	35.4%	**39.6%**	Attempt during MDE	5.7%	**12.3%**
% years w anxious episodes	42.9%	**51.2%**	Attempts among plans	36.9%	**58.3%**
Comorbid Conditions			Mean # LT attempts	0.18	**0.43**
Panic attack	37.6%	**59.8%**	Disability (Sheehan)		
Social Phobia	18.3%	**35.5%**	Home	1.7	**2.5**
GAD	16.0%	**31.5%**	Relationships	1.7	**2.4**
Specific phobia	19.5%	**30.8%**	Work	1.4	**2.2**
PTSD	13.7%	**22.3%**	Social	1.9	**2.7**
Panic disorder	5.1%	**16.0%**	Treatment		
Sleep			Antidepressants	15.3%	**31.2%**
MDE: Trouble sleeping	64.3%	**84.1%**	Sedatives	6.5%	**15.5%**
Anxiety Episode: Sleep problem	26.8%	**50.2%**	Tranquilizers	4.6%	**11.3%**
			Antipsychotics	0.2%	**3.1%**

	Respondents with LT MDE +ADS had greater disease burden both in terms of past-year episodes and proportion of years between 1st episode and interview for both depressive and anxious episodes. They experienced more psychiatric illness, poorer sleep, greater functional disability, and were more likely to be hospitalized for a MDE, anxiety, or panic attack. This group was almost twice as likely to have attempted suicide in their lifetime or during their worst MDE. Among respondents endorsing a suicide plan, attempts were significantly greater among those with the ADS. **Conclusion:** Respondents with ADS in their worst MDE experienced significantly greater disease burden, comorbid conditions, suicide risk, disability, and treatment utilization. The DSM-5 ADS may be used to select patients who are at higher risk for these factors complicating MDD. **Disclosures:** All of the authors are employees of Janssen, LLC.
2	Withdrawn
3	**Preferences Among Patients for Using Technology to Communicate with Physicians and to Research, Track, and Self-Manage Depression** Durand E[1], Khurana, L[1], Gary ST[1], Otero, A[1], Hall, C[1], Dallabrida, S[1] *[1] ERT (F/K/A PHT Corporation) Boston, MA USA* **Methodological Question Being Addressed**: This study determined how subjects with depression currently use the phone and internet to communicate with their providers and research their disease, and how they think this technology use impacts communication with their providers. **Introduction**: Effective patient-physician communication and consistent self-management is essential for patients with chronic diseases to improve their health, compliance, and engagement in clinical care. Good communication has been shown to improve patient adherence to treatment instructions, especially in treatment for chronic conditions. Frequent physician communication may be more critical for patients with depression, since physicians must persuade patients that the regimen be followed, and patients are responsible for implementing treatments. Using technology options can be one way to facilitate this communication and self-care **Methods**: Subjects aged 23-79 with depression (n=107) were surveyed as part of a mode equivalence study. Subjects answered questions focused on their preferences for communicating with physicians and how they currently use technology to manage their disease and general health. **Results**: Subjects were diverse in age, sex, ethnicity, and technology use. 47% of subjects reported owning a smartphone**,** and 90% of these subjects use their smartphone daily. 47% of subjects have access to a home computer, and 52% have internet at home. 49% reported using the internet daily. In a multi-select question, subjects stated they want their physicians to contact them between visits by phone call (87%), email (21%), or text message (21%). 46% of subjects reported that they regularly monitor or keep track of their disease. Of those subjects that do monitor their depression, 75% use paper and 25% use a computer or online forum, and 90% reported that they share these results with their physicians. If subjects did not understand information they received from their physicians, 49% used internet searches or online forums to get more information. With respect to depression, subjects were most likely to research information on current treatments (67%) and alternative treatments (57%), followed by the impact of diet (53%) or exercise (47%) on depression. When asked how researching their health online might impact discussions with their providers, subjects reported they would likely ask more questions about medication (54%), additional treatment options (47%), alternative therapies (30%), and new technologies (26%) related to their disease. 83% of subjects thought they would ask more questions and increase communication with their physicians as a result of researching their disease. When asked about a potential smartphone application that provides educational messages about depression, subjects would like to have information on the effects of foods/beverages on medications they are currently taking (61%), nutritional value of foods they currently eat (54%), and heathy alternatives to those foods (50%). **Conclusions**: Many subjects with depression currently use technology to gain more knowledge about their diseases (using internet searches and forums), and to communicate with their providers (through phone calls, emails, and text messaging). Subjects think that using technology to research disease management promotes discussion with their physicians. Subjects are most likely to research their treatment options and the impact of diet on disease, and would like similar information provided via smartphone app. Providers should consider these findings when seeking to increase engagement in patients with depression in the clinical care setting. **Disclosure**: The authors are employees of ERT

4	**Development of PK/AE Models in Subjects with Schizophrenia and Healthy Japanese and Non-Japanese Subjects** Tsai M[1], Goldsmith P[2], Xie J[1], Macek TA[1] *[1]Takeda Development Center Americas, Inc., Deerfield IL, USA; [2]Takeda Development Center Europe, Inc., London, UK* **Methodological Question Being Addressed**: Develop studies for within-study comparisons of the tolerability of TAK-063 in Japanese vs. non-Japanese subjects and healthy subjects versus subjects with schizophrenia; characterize the relationship betweenTAK-063 exposure and adverse events to help design future clinical studies. **Introduction**: Current antipsychotics are generally less well tolerated in healthy subjects vs. subjects with schizophrenia and in Asian vs. non-Asian subjects, though very few studies have directly compared antipsychotic tolerability within these populations. The PDE10A enzyme is selectively expressed in the medium spiny neurons of the striatum. Inhibitors of PDE10A like TAK-063 have potential for use in the treatment of schizophrenia. Preclinical data suggests the novel MOA may impart a different propensity for adverse events in humans than current antipsychotics, however the relationship between exposure and adverse events to date has not been reported. **Methods:** Two placebo-controlled, double blind, dose-escalation studies were conducted. The first was a single rising dose (SRD) study of Japanese and non-Japanese healthy volunteers (6 cohorts, n=14 per cohort, 11 subjects (5 Japanese, 6 non-Japanese) TAK-063 and 3 subjects (i.e., 1 Japanese, 2 non-Japanese) placebo). The second was a multiple rising dose (MRD) study that included schizophrenia subjects washed out of their antipsychotic medication for at least five half-lives prior to dosing and healthy Japanese volunteers (n=10 per cohort, 8 TAK-063 and 2 placebo). There were 5 cohorts of subjects with schizophrenia and 3 cohorts of Japanese subjects. Safety and tolerability assessments and intensive PK sampling were collected throughout the trials. The most frequently reported AEs [somnolence and extrapyramidal symptoms (EPS)] were modeled using a logistic regression model: $f[P(AE_i=1)]=\log(p/(1-p))=\beta+f_{exp}$ where AE_i takes a value of 1 if subject i has AE at some time during the study and 0 otherwise. The parameter β denotes the logit for subjects not on drug (placebo). f_{exp} represents the function describing the exposure-response relationship expressed as linear or nonlinear forms, using individual steady-state Cmax and AUC values estimated from non-compartmental methods. **Results:** Somnolence and EPS frequency increased with dose and/or exposure. Linear models using Cmax or AUC values demonstrated adequate goodness-of-fit with no substantial model improvement using Emax functions. Disease status as a covariate was significant for EPS but not for somnolence. Single doses of TAK-063 were similarly well tolerated in Japanese and non-Japanese subjects and somnolence was similar between groups. In the MRD study, the overall rates of adverse events were similar between subjects with schizophrenia and Japanese healthy subjects. At equivalent doses, the incidence of somnolence was similar between each group. However, the incidence of EPS was higher in subjects with schizophrenia than in healthy subjects, despite similar TAK-063 pharmacokinetics across all subjects and between studies. **Conclusions:** Our preliminary results show that PK/AE models described the incidence of somnolence and EPS well. These models will be refined with emerging data to predict AEs for future clinical trial designs for TAK-063, with the aim of understanding the exposure-AE profile associated with PDE10a inhibition. The reasons for the higher rates of EPS in subjects with schizophrenia are unknown. **Disclosures:** M Tsai, P Goldsmith, J Xie, TA Macek: Employees of Takeda
5	**Covariate Adjustment in Analyses of Time-to-Event Endpoints** Mao L[1], Ibrahim Turkoz, I[1], Alphs L[2] *[1]Janssen Research & Development, LLC, Titusville, NJ, USA, [2]Janssen Scientific Affairs, LLC, Titusville, NJ, USA* **Methodological Question Being Addressed:** What is the role of covariate adjustment in analyses of time-to-event endpoints? **Introduction (Aims):** Baseline covariates are routinely collected in randomized clinical trials. Adjustment for important prognostic covariates could help define differential treatment response in patient subpopulations and improve the precision of estimation and statistical inference. In patients with serious mental illness such as schizophrenia, numerous prognostic factors have been reported in the literature as predictors of relapse or treatment discontinuation.[1,2] Unadjusted analyses ignore heterogeneity in patient samples and therefore represent only a crude estimate of treatment effect. Unadjusted methodologies are particularly problematic for assessing the robustness of study findings because the imprecision could lead to conservative conclusions. Adjusting for important prognostic covariates that moderate response may increase

power. In this presentation we illustrate an approach to identify and adjust for important prognostic factors in the analyses of time-to-event endpoints, using data from a randomized, active-controlled study in patients with schizophrenia and a history of incarceration (NCT01157351).

Methods: Although univariate Cox regression analysis is useful for demonstrating basic relationships with time-to-treatment failure, clinical practice suggests that multiple factors can influence the risk of treatment failure. To objectively identify important covariates, we modeled the covariate-outcome relationship separately within each treatment group using Cox regression with a stepwise model-selection procedure. Prognostic baseline variables included age, sex, race, duration of illness, baseline Personal and Social Performance scale score, baseline Clinical Global Impression of Severity scale score, multiple (≥2) prior incarcerations (yes/no), history of substance abuse (yes/no), prior health insurance coverage (yes/no), and being randomly assigned to the same antipsychotic medication taken before randomization (yes/no). Predictive covariates that met model-selection criteria from these independent models were retained in the final Cox regression model.

Results: The final Cox regression model included gender, multiple prior incarcerations, history of substance abuse (yes/no), prior health insurance coverage (yes/no), and being randomized to the same antipsychotic medication taken before randomization as important predictors of time-to-treatment failure, in addition to study treatment. The covariate-adjusted model yielded a more significant difference between treatment groups (hazard ratio [HR], 1.66; 95% confidence interval [CI], 1.25–2.21; $P<0.001$) than the model without covariates (HR, 1.43; 95% CI, 1.09–1.88; $P=0.011$).

Conclusion: Covariate adjustment is an important method to quantify treatment effect more precisely and increase confidence in the generalizability of results. It is especially relevant when conducting analyses of randomized, active-controlled clinical trials. Post hoc identification of covariates should be based on objective criteria in order to minimize biases.

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

References: 1) Hall DL, Miraglia RP, Lee LG, Chard-Wierschem D, Sawyer D, Predictors of general and violent recidivism among SMI prisoners returning to communities in New York state. *J Am Acad Psychiatry Law*. 2012;40:221–231.; 2) Essock SM, Covell NH, Davis SM, Stroup TS, Rosenheck RA, Lieberman JA. Effectiveness of switching antipsychotic medications. *Am J Psychiatry*. 2006;163:2090–2095.

6

Quantifying Myelin Kinetics in Healthy Subjects Using Deuterium Labeling

Kanhai K[1], Goulooze S[1], Stevens J[1], Hay J[1], Dent G[2], Verma A[2], Hankemeier T[3], de Boer T[4], Meijering H[4], Chavez J[2], Groeneveld G[1]

[1]Centre for Human Drug Research, Leiden, the Netherlands, [2]Experimental Medicine, Biogen Cambridge, MA, USA
[3]Netherlands Metabolomics Centre, Leiden, the Netherlands, [4]Analytical Biochemical Laboratory BV, Assen, the Netherlands

Methodological Question Being Addressed: Is it feasible to estimate the myelin turnover in healthy volunteers?

Introduction: Demyelinating diseases, such as multiple sclerosis (MS), are characterized by an increased breakdown of myelin with a subsequent failure of the remyelination process. Enhancement of remyelination may improve recovery after an exacerbation. However, direct enhancement of remyelination can only be shown when myelin turnover rate can be quantified. The turnover rate of biomolecules can be determined by quantification of deuterium labeling after chronic administration of deuterated water. Although the labeling of myelin cannot be determined directly in vivo, typical breakdown products or myelin precursors can be measured in CSF.
The goal of the current study was to demonstrate feasibility of deuterium labeling of galactosylceramides in human cerebrospinal fluid (CSF) after chronic dosing of D_2O (deuteriumoxide, also named heavy water) and to develop a mathematical model to allow estimation of myelin turnover rate based on quantitative measurements of deuterated galactosylceramide.

Materials and methods: Two healthy men and 4 healthy women that met the enrollment criteria consumed 60 mL 70% D_2O twice a day for 70 days. Urine samples were collected weekly to measure the percentage D_2O in body water. The subjects received 5 lumbar punctures for CSF sampling at 35, 70, 94,163, and 547 or 714 days after the first D_2O administration. Timing of sampling was based on literature based translational simulations. A QTRAP® 5500 LC/MS/MS System was used to determine deuterium labeling of galactosylceramide in CSF. Myelin turnover rate in CSF was estimated using non-linear mixed effects modeling.

Results: The D_2O percentage in total body water reached 1.5-3.9%, and the deuterated fraction of galactosylceramide reached 0.05-0.14% indicating ongoing myelin synthesis. Based on the first 4 CSF samples, the apparent myelin half-life was estimated at 1150 days (95% CI = 1077-1223). Results of the 5th CSF sample are being analyzed; they can be used to confirm this estimate and will be presented.

Conclusion: The incorporation of deuterium into galactosylceramide can be measured by sampling CSF after D_2O dosing. Deuterium incorporation rate can be used to estimate the apparent myelin turnover-rate. Future studies will measure myelin turnover-rate in patients with MS and determine the effects of remyelinating therapeutic interventions.

7 Identification of Placebo Responders in Early Phase Clinical Studies

Enschedé D[1], van Amerongen G[1], van Gerven J[1], Groeneveld G1, Hay J[1]

[1]Centre for Human Drug Research (CHDR)

Methodological Question Being Addressed: To investigate the incidence of placebo responders in early phase clinical trials of analgesics and to identify what covariates are associated with a placebo response.

Introduction (Aims): When subjects in clinical trials display a high placebo response, the difference between placebo compared with a novel drug decreases, leading to false negative results and increased costs. When covariates for placebo response can be identified, it makes it possible to exclude placebo responders from clinical trials to increase the difference from placebo of the novel drug. In early phase clinical studies, pain models are routinely used to investigate the analgesic properties of novel compounds. However, these studies also need to take into account the placebo response to ensure optimal analysis of the data. The aim of this study was to investigate the incidence of placebo responders in early phase, clinical trials of analgesics and to identify what covariates are associated with a placebo response.

Methods: Data from 3 double blind, double dummy, randomized, placebo-controlled, 4 or 5 way cross-over studies, investigating analgesic response was used. Subjects were administered novel and established analgesics including a 30-minute intravenous infusion of fentanyl 50 µg/kg, phenytoin 300 mg, (S)-ketamine 10 mg or placebo (NaCl 0.9%), or a single oral dose of imipramine 100 mg, pregabalin 300 mg, ibuprofen 600 mg or placebo. Pain test measurements were performed at baseline and up to 10 hours post-dose. The pain models used consisted of tests eliciting cutaneous electrical, mechanical and thermal (contact heat and cold pressor)-pain and included a UVB model and a paradigm of conditioned pain modulation. To determine which subjects were placebo responders, the average change from baseline (CFB) during each visit was calculated. When the average CFB was highest in the placebo occasion, the subject was identified as a placebo responder. Covariates, including anthropometric and baseline pain responses, were included in a regression model to investigate relationships with placebo response.

Results: The identification of placebo responders found that 10% of the subjects could be considered placebo responders. These results were reproducible in all three studies, which each contained 31, 19 and 19 subjects. The analysis did not identify any significant covariates, which means the magnitude of placebo response cannot be predicted by the covariates used in the analysis.

Conclusion: This analysis demonstrates that approximately 10% of healthy subjects participating in early phase clinical studies of analgesics are placebo responders. However, identifying covariates of the placebo response remains elusive.

8 Patient Self-Assessment in Clinical Development Studies in Severe Mental Illness: Is the Experienced Sampling Method an Option?

van Os J[1], Correll C[2], Leucht S[3]

[1]Department of Psychiatry and Psychology, School of Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, the Netherlands, [2]Hofstra North Shore-LIJ School of Medicine, Hempstead, USA; The Zucker Hillside Hospital, Glen Oaks, USA, [3]Department of Psychiatry and Psychotherapy, Technical University, Munich, Germany

Methodological Question Being Addressed: The use of the experienced sampling method in clinical phase 2-4 drug development trials: what are the advantages and limitations?

Introduction: Clinical trials in severe mental illness generally focus on efficacy and are not designed to assess patient-centred outcomes, let alone individual recovery, which is at odds with the move towards community psychiatry, collaborative care, shared decision making, and personalised care.

To evaluate treatments in clinical trials, we currently ask patients to recall symptoms over the last 2 to 3 weeks using semi-structured interview scales such as the Positive and Negative Syndrome Scale (PANSS). These scales are the gold

	standard, but also have limitations because with recall bias and averaging of data, information is lost. In addition, standard rating scales can be insensitive to change, and have a highly variable interrater reliability despite training of raters. Developing new (self-assessment) instruments is complex. Gaps exist regarding the best ways to define and assess the relevant dimensions of patients' needs, to assess and utilise caregiver- centred outcomes, determine the domains that can reliably be assessed via self-reporting, and to integrate self-reporting scales with healthcare technology service use and delivery. Research shows that e-mental health applications (e.g., web sites, mobile applications) can be used by severely mentally ill people, and have multiple benefits. Thus, to maximise utility and scalability, new self-reporting scales should be integrated with e-mental health platforms. **Methods**: In the last decade, novel m-health approaches such as the Experience Sampling Method (ESM) have been introduced in psychiatry. ESM is a smartphone-based collection of an intensive time series of mental states and behaviours in the context of daily life, typically over the course of one week. ESM has the potential to facilitate the involvement of patients in the process of needs assessment and the evaluation of treatment response at the level of emotional and social adjustment in daily life. In addition, ESM could be used as a basis for add-on psychological interventions that support psychopharmacological approaches. **Conclusion**: Until now ESM has not been used in drug development programmes in psychiatry. Here we further discuss key areas to consider when developing novel self-assessment scales, we elaborate on the rationale for using ESM, and the potential benefits/pitfalls linked to incorporating ESM in phase 2–4 clinical studies. **Disclosures**: The authors report no conflicts of interest for this work
9	**Examining Placebo Effects on MATRICS Battery Measures in Schizophrenia Cognition Clinical Trials** Keefe R[1,2], Davis V[2], Atkins A[2], Harvey P[3], Lombardo, I[4], Bugarski-Kirola D[5], Reid C[6] *[1]Duke University, Durham, NC USA, [2]NeuroCog Trials, Durham, NC USA, [3]University of Miami, Miami, FL USA, [4]Axovant Sciences Inc. New York, NY USA, [5]F. Hoffmann-La Roche Ltd, Basel, Switzerland, [6]Roche Products Limited, Welwyn Garden City, UK* **Introduction**: The Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Battery (MCCB) is described as the "gold standard" by members of the Psychiatry Divisions of the US FDA and the European Medicines Agency. The psychometrics of the MCCB have been increasingly well established. However, the magnitude and predictors of improvement in sequential assessments with placebo treatment is unknown. **Methodological Question Being Addressed**: Identification of predictors of change in the MCCB over successive visits in patients receiving placebo. **Methods**: We combined data from 12 studies that assessed changes in MCCB performance in 751 patients with schizophrenia receiving placebo over 4 to 56 weeks. Change from baseline was investigated using a mixed-effects model of repeated measures controlling for assessment number, baseline score, study, and the baseline score by study interaction. Predictors evaluated included demographics, baseline characteristics and symptoms. Practice effects were examined in a separate model using data from 7 studies (N=517) that measured cognition at screening. **Results**: The overall mean change in the MCCB composite over 56 weeks, adjusting for baseline score, study and their interaction, was 1.9±0.22 T-score points, with a range of 0.6 to 3.3 points across 12 studies of schizophrenia patients treated with placebo. Mean change scores for the 10 subtests comprising the MCCB ranged from 0.2±0.36 (MSCEIT) to 2.3±0.39 (Trail Making) T-score points. Patients scoring higher on the Marder Anxiety and Depression scale at baseline were more likely to show improvement on the MCCB overall composite ($p=0.004$). Practice effect prior to randomization was negatively associated with placebo response ($p<0.001$). **Conclusions**: This new examination of MCCB practice effect over repeated visits in placebo patients will help determine expectations and strategies for placebo controlled trials examining treatments for cognitive impairment in schizophrenia. **Disclosures**: RSE Keefe has currently or in the past 3 years received investigator- initiated research funding support from the Department of Veteran's Affairs, Feinstein Institute for Medical Research, GlaxoSmithKline, NIMH, Novartis, Psychogenics, Research Foundation for Mental Hygiene, Inc., and the Singapore National Medical Research Council. He has received honoraria, served as a consultant or Ad board member for AbbVie, Akebia, Amgen, Astellas, Asubio, AviNeuro/ChemRar, BiolineRx, Biomarin, Boehringer-Ingelheim, Bristol-Myers Squibb, Eli Lilly, FORUM, Helicon,

	Lundbeck, Merck, Mitsubishi, Otsuka, Pfizer, Roche, Shire, Sunovion, Takeda, Targacept. Dr. Keefe receives royalties from the BACS testing battery and the MATRICS Battery (BACS Symbol Coding). He is also a shareholder in NeuroCog Trials. VG Davis and AS Atkins are employees of NeuroCog Trials. PD Harvey has in the last 12 months served as a consultant for Acadia, Boehringer-Ingelheim, FORUM, Lundbeck, Otsuka-America, Roche, Sanofi, Sunovion, and Takeda. I Lombardo is an employee of Axovant Sciences. D Bugarski-Kirola and C Reid are employees of Roche Pharmaceuticals.
10	**What is the Impact of "Professional Subjects" on Medication Efficacy Trials?** McCann, DJ[1] *[1]National Institute on Drug Abuse, National Institutes of Health* **Methodological Question Being Addressed**: What is the impact of "professional subjects" on medication efficacy trials? **Aims**: Professional subjects, defined here as subjects who enroll in clinical trials only for financial gain, have an unknown impact on medication efficacy trials. The current studies were conducted to model their impact. **Methods**: Two types of professional subjects were considered: those "destined to succeed" (DS) and those "destined to fail" (DF). An example of a DS subject is someone who feigns depression to enroll in an antidepressant trial and then answers questions truthfully after randomization (no longer appearing to be depressed). An example of a DF subject is a smoker who enrolls in a smoking cessation trail with no intension of taking study drug or trying to quit. After setting success rates in legitimate subjects at 5% for placebo and 15% for active treatment, modeling studies evaluated the impact of DS and DF subjects on apparent success rates, the number of subjects required for 80% power (α 0.05, two-sided chi squared test), and apparent effect size (odds ratio). **Results**: With no professional subjects, 141 subjects per group yielded 80% power, and the odds ratio was 3.35. With 10% DS subjects, apparent success rates rose to 14.5% and 23.5%, 298 subjects per group were required for 80% power, and the apparent odds ratio fell to 1.81. With 20% DS subjects, apparent success rates rose to 24% and 32%, 494 subjects per group were required for 80% power, and the apparent odds ratio fell to 1.49. In contrast, the impact of DF subjects was modest. With 20% DF subjects in the study population, apparent success rates were 4% and 12%, 180 subjects per group yielded 80% power, and the apparent effect size was 3.27. **Conclusion**: While all professional subjects negatively impact efficacy trials, DS subjects are especially problematic. A small percentage of DS subjects in a study population can greatly reduce apparent effect size, a phenomenon that cannot be overcome by increased sample size.
11	**Enrollment Patterns: Implications for CNS Clinical Trials** Webster N[1], Vardy J[1], Wise-Rankovic A[1] *[1]INC Research* **Methodological Question Being Addressed**: Do clinical trials in CNS follow typical or unique enrollment patterns across the year? How can the enrollment patterns be characterized whether by region or by therapeutic area? What are the likely reasons for observed enrollment patterns? **Introduction (Aims)**: Enrollment patterns are often an important consideration in lengthy CNS trials when projecting enrollment timelines and enrollment rates. The ability to predict and achieve the desired enrollment rate can determine the success of the clinical trial because of the impact on data analysis plans, overall study timeline, project budget, and marketing goals of the sponsor. Many factors can influence enrollment rates. Enrollment projections for clinical trials typically account for an expected seasonal pattern of decreased enrollment in Northern Hemisphere summer (July/August) and winter (December/January) months. Here, using an archive of four years of enrollment data, we more closely examine enrollment patterns in CNS clinical trials by evaluating site enrollment productivity for observable enrollment patterns across regions and therapeutic areas. **Methods**: Data were reviewed on 107 trials enrolling 31,830 subjects at 3,331 investigative sites in CNS trials conducted by INC Research during the years 2011 – 2014, inclusive. Trials were categorized by three CNS therapeutic areas, as defined by business operations (% subjects): psychiatry (55%), analgesia (34%), and neurology (11%). Data were examined in North America and Europe. Enrollment is defined here as subjects consented into the trial. To standardize the evaluation across each of these trials with differing timelines, we developed a productivity ratio formula that summed the number of subjects consented on any given day as an indicator of the number of sites that could produce enrollment

on that day, and calculated a site productivity indicator as the dividend of these two variables (# subjects/# site days = site productivity). By consolidating these productivity data into a single 12 month observation period, we were then able to observe trends in enrollment.

Results: In Table 1, the total site days and screened subjects by therapeutic area are presented for four consecutive years inclusive of 2011 through 2014. Across all 3 therapeutic areas, October represented the month in which the most subjects were screened (3,554 or 11% of enrollment). However, when using the productivity ratio, which accounts for the number of sites that were active, this high performance in global enrollment was greatest for analgesia studies only (see Table 2). Analgesia demonstrated the greatest variability and adherence to the expected seasonal patterns, and demonstrated a a bimodal peak in both April and October. Site activity in analgesia was also at a peak in the Fall; the productivity ratio was consequently higher in April as there were less sites enrolling more subjects. When observing the analgesia site productivity by region, Europe's peak was in September just after the summer respite and North America's peak was in April but not in October. Psychiatry's site productivity followed a flatter distribution. While, psychiatry's enrollment appears to spike in October (Table 1), the number of sites on board also peaked which indicates that the productivity of these sites was not significantly different (see Table 2). Psychiatry's productivity by region, observed separately, re-introduces more variability and a seasonal pattern in Europe. Neurology studies were rather flat in their site productivity pattern as well (see Table 2).

Table 1

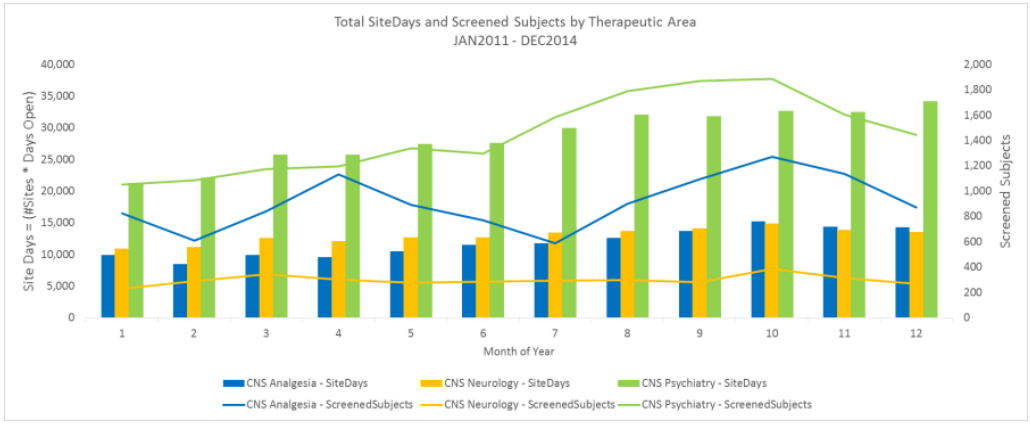
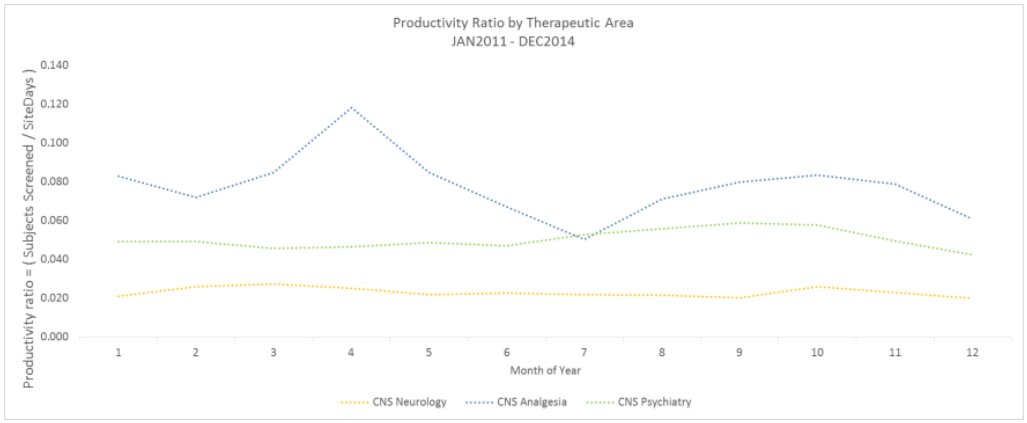


Table 2



Conclusions: Expected seasonal enrollment variations that are often accounted for in clinical trials were observed in part in this large global dataset of enrollment in CNS clinical trials, with the most consistent site productivity peak being in October. An important consideration is the population under study and the measures taken to bolster enrollment, which may lessen the seasonal variation. Further exploration of these patterns is warranted.

Disclosures: Nathan Webster, PMP: INC Research, Julianna Várdy, MBA: INC Research, Alexandria Wise-Rankovic, PhD: INC Research

12	**The Complexion of Risk and Error in Patient Selection for CNS Clinical Trials: Detailed Findings from a 16,000-patient Eligibility Review Database** Nations K[1], Reinhold C[1], Miloslavich K[1] *[1]INC Research* **Methodological Question Being Addressed**: The extent and nature of eligibility decision errors in CNS clinical trials; and how central, collaborative review of subject eligibility an protect against unwanted heterogeneity and error. **Introduction (Aims)**: A growing body of evidence suggests that high CNS trial failure rates are largely attributable to inappropriate patient selection. CNS protocols are universally complicated, with the number of eligibility criteria ranging up to 70 in many trials. Investigators, while clearly dedicated to quality and compliance, interpret protocol criteria differently and have varying levels of tolerance for risk. The subjectivity of every diagnosis in the CNS field, as well as the high number of judgment-driven decisions required to evaluate clinical assessments, naturally translates to compromised internal validity of clinical trials in the absence of careful oversight of incoming subjects. Fortunately, centralized review is both operationally feasible and acceptable to Investigators, when conducted collaboratively leaving final eligibility decisions in the hands of the treating physicians. **Methods**: Eligibility Review is a process conducted at INC Research by a centralized, global team of physicians and doctoral-level clinical scientists. Key screening data in select trials are collected from sites and vendor portals by project management personnel, then compiled and reviewed in a team discussion format. Resulting eligibility concerns and questions are thereafter discussed with Investigators. In the majority of cases, Investigators are able to provide additional clinical history that supports subject eligibility; however, in those cases where Investigators agree the subject is unsuitable, the site proceeds to screen fail the subject. The full process and dialogue takes place within the screening period so that no ineligible patient is randomized inappropriately. **Results**: A total of 16,092 subjects were reviewed for eligibility by a central medical/clinical team, covering 38 trials over a five-year period, 20 in psychiatric indications, 10 in neurology, and 8 in analgesia. Although the submitting Investigators considered all subjects to be eligible to randomize, the review team identified approximately 1,400 (9%) who did not meet eligibility criteria and were ultimately considered unsuitable to enter the trials. Findings differed significantly between therapeutic area: 8.3% of subjects reviewed were considered ineligible in psychiatry trials, 11.6% in neurology trials, and 9.8% in analgesia trials [$F(2,35)= 3.27$, $p=0.04$]. Further examination demonstrated that the proportion of findings was highly dependent on individual trial, where this metric ranged from 4% to 25% in psychiatry trials, 0% to 18% in neurology trials, and 6% to 32% in analgesia trials. The distribution of eligibility-related findings was broad: of the subjects considered ineligible (some for more than one reason), 40% were ineligible for reasons related to laboratory or ECG findings, 40% for medical history, 29% for treatment history or prohibited medications, 17% for psychiatric history, and 7% due to findings related to primary diagnostic validity. All subjects considered ineligible after review were screen failed following collaborative discussions with Investigators. Additional data will be presented on individual indications and regional differences in global trials. **Conclusions**: The current analysis demonstrates that CNS trials are highly susceptible to penetration by unqualified/unsuitable subjects, and that eligibility issues are wide-ranging and not limited to issues of diagnostic validity. Importantly, these data show that no therapeutic area is immune to unwanted heterogeneity in the study sample. Light intervention at the time of screening can homogenize the patient sample to align with protocol criteria, offering protection of statistical power, and rendering trial efficacy and safety results both interpretable and actionable. **Disclosures**: KRN, CKR, and KM are employees of INC Research, the Contract Research Organization responsible for execution of all trials included in this analysis.
13	**eSource Administration of the CDR: Preliminary validation of Internal Consistency Checks** Randolph C[1,2], Weber C[1], Garzio L[1], Negash S[1], Böhm P[1] *[1]MedAvante, Inc., [2]Loyola University Medical Center* **Methodological Question Being Addressed**: Strategies to improve the scoring reliability of the Clinical Dementia Rating scale (CDR) are needed as the CDR is used as a sole primary endpoint in clinical trials of prodromal Alzheimer's disease (AD) and also as a co-primary endpoint in trials of mild-moderate AD. A new tablet-based electronic source (eSource) administration of the CDR with built-in consistency checks was tested with a recently collected clinical trial database.

Introduction (Aims): Scoring of the CDR can be challenging and scoring errors are common. We recently developed consistency checks (or "flags") based on a proprietary database containing several thousand expert reviews of CDR assessments to provide raters with real-time queries suggesting they cross-check their scoring with information obtained during the interview, prior to finalizing scores. The goal of this study was to review a sample of CDR administrations performed on paper in order to determine how frequently these flags would have been triggered, and how often a flag would have been associated with a scoring error.

Methods: 200 CDR administrations done within a clinical trial in mild-moderate AD during 2013-2014 were randomly selected. This sample consisted of CDRs completed by a total of 110 site raters from 94 sites in 11 countries. Each CDR source document was reviewed applying the consistency checks that are built into the current eSource system. The administrations that would have triggered any flags were cross-checked against the central review scoring for that administration.

Results: 47.5% of assessments would have resulted in one or more flags. Of the assessments with flags, 63% of these contained a scoring discrepancy. At the domain level, the percentage of flags that were also associated with a scoring discrepancy within that domain ranged from 14% (Home and Hobbies) to 43% (Judgment and Problem Solving).

Conclusions: The consistency checks built into the tablet-based system for the CDR administration would have been triggered on nearly 50% of paper-based CDR administrations. It would appear that these alerts are effective in identifying administration and scoring errors, both within and across domains. Additional research comparing error rates on central review of eSource CDR administrations to paper-based CDR administrations is underway in order to more directly clarify the advantages of the tablet-based administration.

14

Methodological Considerations for Rater Training and Certification In Down Syndrome (DS) Clinical Trials: Overview and Initial Findings from Two DS Studies

Kingery L[1], Ventola P[2], Liogier D'ardhuy X[3], Goeldner C[3], Cohen E[1], Krishna V[1]

[1]Cogstate, New Haven, CT, [2]Yale Child Study Center, New Haven, CT, [3]pRED, F. Hoffmann-La Roche AG, Basel, Switzerland

Methodological Question Being Addressed: Methodological standards for rater training and certification applicable across all CNS indications and scales have not been established, although general recommendations have been proposed (Daniel et al., 2013). Sponsors are therefore responsible for customizing rater training programs based on the study goals and the roles of the scales in the study. This poster summarizes the methods and initial results of a rater training program developed to ensure accurate administration of a unique set of scales in two Down Syndrome (DS) studies.

Introduction (Aims): Ensuring reliable and valid screening and outcome assessments in Down Syndrome (DS) clinical trials is challenging. Scales used in these studies are complex, including psychometric tests, clinical global impression ratings, caregiver interviews, and caregiver-reported outcomes. A further challenge is that many of these assessments require specialized rater skills. This poster summarizes the training methods and results from two DS studies (one observational study and one phase II clinical trial).

Methods: The customized rater training and certification program included rater prequalification, three primary rater roles (clinical, psychometric, and caregiver-scales raters), didactic trainings for all scales, and certification procedures assessing competencies for the psychometric scales and the Vineland-II. All raters completed a Rater Experience Survey (RES) assessing education, clinical experience with DS, and scale-specific experience. Certification procedures included mock scoring exercises, hands-on practice at investigators' meetings (IM), and audio-recorded practice administrations of the psychometric scales and the Vineland-II. Relationships among these variables will be examined using analysis of variance and correlation.

Results: Thus far, 48 raters completed a RES and met pre-qualification criteria. Pre-qualified raters' educational background varied as follows: clinical raters - 1 master's degree (MA), 2 doctoral degrees (PhD/PsyD), and 14 medical degrees (MD); psychometric raters - 6 bachelor's degrees (BA), 8 MA, and 10 PhD/PsyD. Seven raters were prequalified for caregiver-scales only (4 BA, 2 MA, 1 PhD). Overall DS clinical experience varied, ranging from 0 to 37 years (M = 7.8, SD= 8.9; clinical raters M = 11.1, SD=8.3; psychometric raters M = 7.4, SD=9.5; caregiver-scales rater M = 1.1, SD=1.5). Scale-specific experience varied from none to extensive experience, depending on the scale. Rater performance on certification materials for psychometric tests including the Leiter-3, Children's Memory Scale (CMS), and subtests of the CELF-P-2 varied from few errors to many. Analyses examining the relationship between DS clinical experience, education, and performance on the training/certification materials are ongoing and will be reported fully in the final poster. Although previous experience with the scale was associated with fewer scoring errors on the rater assessment materials,

	several experienced raters made unexpected errors during the hands-on training and audio-recorded practice administrations, validating the need for their participation in the certification program. **Conclusions**: Clinical trial raters who meet pre-qualification criteria are highly diverse in their clinical and scale-specific experience, and not always aware of the errors they make despite their experience. Comprehensive trainings and individual assessments are required to prepare and ensure raters testing participants with DS are capable of performing the scales accurately. **References**: Daniel, D., Opler, M., Wise-Rankovic, A., & Kalali, A. (2013). Consensus recommendations on rater training and certification. CNS Summit Rater Training and Certification Workgroup.
15	**Identical Ratings Are An Early Marker of Data Quality Issues** Kott, A[1]; Daniel, DG[2] *[1]Bracket, Prague, Czech Republic, [2]Bracket, McLean, VA, USA* **Methodological Question Being Addressed**: We examined whether the presence of identical PANSS ratings (30/30 PANSS items scored the same across consecutive visits) between the screening and baseline visits predicted identical ratings after randomization **Introduction**: A primary focus of data quality monitoring in clinical trials is early identification of issues that may detract from signal detection. We have previously identified identical PANSS ratings (30/30 PANSS items scored the same across consecutive visits) as markers of poor ratings quality. In 10 global schizophrenia trials, identical ratings occurred in approximately 4.7 % of visits (Daniel and Kott, 2014). In the current analysis, we examined whether the presence of identical PANSS ratings between the screening and baseline visits predicted identical ratings after randomization. **Methods**: We analyzed data from 4,764 randomized subjects into 10 global schizophrenia clinical trials where data was available for the PANSS at the screening, baseline and post-baseline visits. We assessed the association between the presence of identical ratings between the screening and baseline visits and the presence of identical ratings at post-baseline visits utilizing the Chi-square statistic. Analyses were applied to the combined data set as well as individual clinical trials. **Results:** Out of the 4,764 randomized subjects, 440 (9.24%) had their baseline PANSS scores identical to screening scores. 762 (15.99%) subjects had at least 1 pair of identical ratings that included a post-baseline visit. Out of these 762 subjects, 248 (32.55%) had their baseline PANSS scores identical to screening scores. The association between the presence of identical rating at baseline and post-baseline visit was found to be significant ($chi2(1) = 587$, $p < 0.001$), for the whole dataset and for each study individually. The odds of having an identical rating in the post-baseline studies are approximately 9.6 (6.6 – 13.9) times higher for those subjects who had identical rating at baseline compared to those who did not have an identical rating at baseline. **Conclusions**: Our analyses found that across 10 global schizophrenia studies the presence of PANSS identical ratings between screening and baseline robustly predicted identical ratings for that subject after randomization. This represents an important opportunity to identify and address data quality issues prior to randomization. The presence of identical ratings at baseline represents a highly concerning finding as it not only suggests that there will likely be more identical ratings recorded for the subject later in the study but because it likely modifies the subjects' baseline severity, thus distorting the assessment of change from baseline to endpoint. There are numerous options for external review to assure quality of screening and baseline evaluations and appropriate subject selection. Examples include audio/video recordings of site assessments, site completed electronic PANSS and subject validation workbooks and independent telephone assessment of the subject. **References**: Daniel, D.G. & Kott, A. (2014) Is identical scoring of the PANSS across consecutive visits a marker of poor data quality? Presented as a poster at the International Society of Clinical Trials Methodology (ISCTM) Autumn Conference, October 6 – 8, 2014, Boston, MA USA. **Disclosure**: Both authors are full time employees of Bracket.

16 **Is a Computer Simulated Rater Good Enough to Administer the Hamilton Depression Rating Scale in Clinical Trials?**

Sachs G[1], DeBonis D[1], Wang X[1], Epstein B[1]

[1]Bracket

Methodological Question Being Addressed: Are Hamilton depression rating Scale scores obtained by a computer simulated rater within the range expected from site-based raters? Concern about the high rate of failed clinical trials fuel the desire for better measurement techniques.

Introduction: Hamilton depression rating Scale (HDRS) was developed as an instrument which scored multiple symptoms of depression based on information elicited during a face-to-face clinical interview.

Psychometric data suggests several strategies can improve performance of the HDRS including rater training (Muller 2003), structured clinical interviews (Williams 1998) and explicit definition of anchor points (Kalali 2002). These techniques encourage raters to apply a rules based approach to their subjective ratings.

Unfortunately results from actual randomized clinical trials do not always show the benefit of these techniques (Kahn, ACNP poster 2014, Sachs CINP 2014) and video certification exercises commonly produce a distribution of total scores with standard deviation >3.5. The evolution of clinical ratings toward a rules based approach encouraged the authors to develop a computer simulated rater (CSR) based on the scripting and rules taught to site-based raters training to administer and score the HDRS-24.

Methods: Blinded data was harvested from a double blind placebo controlled industry sponsored study. At each study visit that required the site based rater to administer the HDRS-24, the CSR administered the HDRS-24 as a separate independent rating.

The CRS conducts an interactive interview directly with the study subject. An interview algorithm selects probe questions based on the subject's last response and a scoring algorithm maps the subject's responses to a unique anchor point.

Site based raters administering the HAMD were required by the sponsor to have > 2 years of experience with the HAMD 24 and meet training and certification standards based on scoring video tapes viewed at investigator meetings.

Results: Results were obtained from the Bracket HAMD-24 blinded study dataset which included 737 subjects, 112 raters, and 3180 administrations of the paired rater and computer interviews made over the course of a 16 week double blind placebo controlled clinical trial.

Table 1: HDRS and ICC across study visits

Study Visit	Sample Size	HDRS Mean (±s.d.) SBR	HDRS Mean (±s.d.) CSR	p-value	ICC
Screening	723	29.5 (±5.7)	29.9 (±7.8)	0.2401	0.58
Week 0	510	29.3(±5.7)	29.3(±8.4)	0.9266	0.60
Week 2	470	24.1(±7.7)	23.4(±9.0)	0.1816	0.73
Week 8	401	18.7(±9.5)	19.3(±9.9)	0.3658	0.82
Week 9	371	16.8(±9.7)	18.2(±10.2)	0.0547	0.85
Week 12	360	15.5(±9.4)	16.9(±10.2)	0.0288	0.84
Week 16	345	13.2(±9.2)	15.1(±10.2)	0.0598	0.82

Conclusions: SBR and CSR produced comparable mean scores across all time points examined in this RCT. Changes in ICC and p-values observed after randomization may reflect subject practice effects, changes in variance over the course of the study or alteration in rater or respondent behavior after determination of eligibility.

Computer administered scales may offer important advantages not because a CSR is a better than the average site-based rater, but because the computer is consistent, fast, and frugal. By simulating the judgment of a human rater the CSR offers an alternative to reliance on self-report measures.

Disclosures: All Authors are full time employees of Bracket.

17 **A Methodology for Evaluating Clinical Trial Sites and Raters Based on Performance Data**

Miller D[1], Feaster T[1], Allen S[1], Gratkowski H[1] and Butler A[1]

[1]Bracket, Wayne, PA, USA

Methodological Question Being Addressed: Is it possible to use past training, certification, and clinical outcome ratings data to enrich clinical trial site and rater selection?

Introduction: The selection of investigative sites to participate in clinical trials is often focused on evaluation of historical therapeutic experience, past subject recruitment and regulatory compliance records. The evaluation of clinical outcome performance criteria could be an important component of improving study execution. In this study, a methodology was developed to evaluate historical experience and clinical outcome administration performance data as a mechanism to enhance site and rater selection.

Methods: A proprietary database of sites and raters who had participated in recent clinical trials (trailing 3 years) was compiled. The database included historical experience with ratings scales, performance data on certification programs, and performance data based on quality assurance programs implemented to ensure quality ratings were performed. Quality assurance measures included blinded review of audio/video ratings, worksheet reviews assessing accuracy of scoring and rater scoring analytics. Each rater's experience, certification and quality assurance measures were assigned weighted numerical values. Each rater at a site contributed to the site's overall score. Each site was classified as "Recommended", "Moderately Recommended" or "Not Recommended." Once sites were categorized, a clinical review was conducted of the data and rankings based on experience with sites and raters as well as overall performance. Sites in the "Recommended" category were those that had overall positive scores contributing and were in the top 56% of sites evaluated. Sites in the "Moderately Recommended" group had lesser positive weighting and contributed to 40%. Sites in the "Not Recommended" category had scores that were in the lower 5% of sites.

Results: Data were evaluated for 13,600 unique raters covering 2,195 research sites in 49 countries and across 21 different clinical trials. Certification and experience data were evaluated for every rater. Performance data included 27,277 scale administrations resulting in 10,188 rater contacts for potential quality assurance issues. Of those issues that required contact with the site rater, 4,352 resulted in a remediation (re-training or targeted review of potentially problematic data). Sites and raters were required to complete a minimum number of trials and ratings to be weighted and evaluated. A total of 2,198 sites were given a final classification based on those criteria. 1,223 (56%) were classified as "Recommended", 873 (40%) were classified as "Moderately Recommended", and 102 (5%) were classified as "Not Recommended."

Conclusions: CNS clinical trials almost always rely on subjective, clinician-rated outcomes as primary endpoints. The importance of training those raters and monitoring their subsequent performance is well understood. Systematic tracking and evaluation of experience and performance data is routinely utilized to assist in clinical trial site feasibility processes. This data frequently relies heavily on past patient recruitment and site data monitoring outcomes, and rarely proactively references past clinical ratings performance data. Utilizing this data may be useful in identifying the highest quality clinical trial sites and raters to conduct future research programs.

Disclosures: This poster is financially supported by Bracket. The authors report no conflicts of interest for this work.

18 **The Interview for Clinician-Rated Dimensions of Psychosis Severity (IPS)**

Daniel D[1], Velligan D[2], Fu DJ[3], Brown B[3], Alphs L[3]

[1]Bracket Global LLC, [2] University of Texas Health Science Center, San Antonio, [3] Janssen Scientific Affairs, LLC

Methodological Question Being Addressed: Rationale for creation of a semi-structured interview and manual to facilitate accurate and reliable use of the CRDPSS

Introduction: The Clinician-Rated Dimensions of Psychosis Symptom Severity (CRDPSS) is an eight-item dimensional clinician completed assessment of symptoms commonly observed in psychotic disorders. (Barch DM, Bustillo J, Gaebel W et al, 2013;. DSM-5) This scale is designed to be used by clinicians to monitor treatment progress and enhance clinical decision-making. Employing the CRDPSS scale in clinical trials can potentially help to interpret and apply clinical trial results in real world practice. High levels of inter- and intra-rater reliability are critically important to the success of clinical trials. The psychometric properties of the CRDPSS in clinical and research settings have not yet been reported in the peer reviewed psychiatric literature. However, semi-structured interviews and more detailed instructions provide enhanced inter-rater reliability of numerous existing assessments utilized in clinical trials of psychotic disorders. To enhance

reliability, we designed and are field testing a semi-structured interview and instruction manual for the CRDPSS entitled The Interview for Clinician-Rated Dimensions of Psychosis Severity (IPS).

Method: Principles guiding design of the IPS included: 1) standardization of the definition of each domain; 2) standardization of thoroughness and content of information applied to rating of each dimension; 3) standardization of the method of assigning a rating of dimensional severity; 4) cultural neutrality; 5) applicability to "real world" clinical assessment as well as formal clinical trials. Segments of the IPS were inspired by the BPRS Expanded Version 4.0 (Ventura J, Lukoff D, Nuechhterlein KH et al, 1993), Cognitive Assessment Interview (CAI) (Ventura J, Reise SP, Keefe RS et al, 2010) and the NSA-16 Manual (Axelrod BN, Goldman RS, Alphs LD, 1993.)

Results: In order to assure consistent and thorough interviews and instructions for scale completion the IPS outlines: definitions for each dimension, sources of information and reference groups for evaluation of symptom severity. This specifically includes:

1) Time frame instructions
2) Sources of information specifications for each item
3) A list of interview questions in semi-structured format
4) Additional anchor points addressing the frequency, intensity and impact on functioning of each domain of symptoms
5) Instructions for reference group comparison for evaluating cognition and negative symptoms.
6) Instructions for assessing the combined effect of avolition and restriction of affect in evaluating negative symptoms
7) A case example with explanation of correct scoring.

Discussion: The CRDPSS is a recently-introduced assessment measure for the severity of psychotic illness included in DSM-5. The eight items of the CRDPSS are designed to be able to be rated in clinical interviews. However, their psychometric qualities have yet to be reported in the peer-reviewed literature. The IPS has been designed to enhance the accuracy and reliability of the CRDPSS for clinical trial use. The IPS is currently undergoing field testing in a large real world style clinical trial in three countries and languages. Utilizing this sample, assessment of the psychometric qualities of the CRDPSS and IPS will be reported in future presentations.

Acknowledgements: The authors appreciate the advice and feedback of Dr. Joe Ventura.

19

Characterization of Cognitive Function with the CANTAB in Individuals with Amnestic MCI in Relation to Hippocampal Volume, Amyloid and Tau Status: Preliminary Baseline Results from the Pharmacog/European-ADNI Study

Nathan P[1,2], Galluzzi S[3], Marizzoni M[4], Cotelli M[4] Babiloni C[5], Bartres-Faz D[6], Bordet R[8], de Anna B[10], Didic M[11], Farotti L[12], Forloni G[13], Jovicich J[14], Marra C[15], Marzano N[5], Molinuevo J[16], Nobili F[17], Pariente J[18], Parnetti L[12], Payoux P[19], Picco A[17], Ranjeva JP[20], Roccatagliata L[21], Rossini P[15], Salvadori N[12], Schonknecht P[22], Berwig M[22], Hensch T[22], Soricelli A[23], Tsolaki M[24], Vecchio F[25], Visser P[26], Wiltfang J[27], Orlandi D[3], Abbott R[28], Blin O[7], Frisoni G[29]

*[1]Inventiv Health Clinical, Cambridge, UK, [2]Department of Psychiatry, University of Cambridge, [3]IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy, [4]IRCCS Fatebenefratelli, Brescia, Italy, [5]University of Rome, Rome, Italy, [6]Universitat de Barcelona and IDIBAPS, Barcelona, Spain, [7]Mediterranean Institute of Cognitive Neurosciences, Marseille, France, [8]Universite Lille, Lille, France, [9]Hospital Clinic Barcelona, Barcelona, Spain, [10]Aix-Marseille Universite, Marseille, France, [11]Service de Neurologie et Neuropsychologie, Marseille, France, [12]Ospedale Santa Maria della Misericordia, Perugia, Italy, [13]Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy, [14]University of Trento, Trento, Italy, [15]Catholic University, Rome, Italy, [16]ICN Hospital Clinic Universitari and Pasqual Maragall Foundation, Barcelona, Spain, [17]Clinical Neurophysiology, Department of Neurosciences, Ophthalmology and Genetics, University of Genoa, Genoa, Italy, [18]Department of Neurology, CMRR and INSERM U825, Toulouse, France, [19]Institut National de la Santé et de la Recherche Médicale, Toulouse, France, [20]CIC-UPCET, CHU La Timone, AP-HM, UMR CNRS-Universite de la Mediterranee, Marseille, France, [21]Department of Neuroscience, Ophthalmology and Genetics University of Genoa, Genoa, Italy, [22]University of Leipzig, Leipzig, Germany, [23]Fondazione SDN per la Ricerca e l'Alta Formazione in Diagnostica Nucleare, Naples, Italy, [24]Aristotle University of Thessaloniki, Thessaloniki, Greece, [25]AFaR Association for Biomedical Research, Rome, Italy, [26]Alzheimer Centre VUMC, Maastricht, Netherlands, [27]University of Duisburg-Essen, Essen, Germany, [28]**Cambridge Cognition, Cambridge UK,** [29]LENITEM (Laboratory of Epidemiology, Neuroimaging and Telemedicine) IRCCS - Istituto Centro S. Giovanni di Dio - Fatebenefratelli, Brescia, Italy, Memory Clinic and LANVIE - Laboratory of Neuroimaging of Aging, University Hospitals and University of Geneva, Geneva, Switzerland*

Methodological Question Being Addressed: Identifying a homogeneous population of patients is critical for prodromal Alzheimer's disease (AD) clinical trials. The findings of this study describe biomarkers that could be utilised at screening to identify a more homogenous population of patients in clinical trials with fronto-striatal and hippocampal dependent

attention and memory deficits, neurodegeneration and CSF biomarker abnormalities consistent with prodromal AD populations.

Introduction: MCI is a heterogeneous condition with differential underlying pathophysiologies. Accumulation of beta amyloid (A□) and/or Tau in the brain is associated with greater neurodegeneration and cognitive decline and a prelude to Alzheimer's disease (AD). Understanding MCI populations for hippocampal specific memory deficits and biomarker abnormalities will help identify a more homogeneous population with a greater risk of developing AD.

Methods: Participants were recruited from the PharmaCog (E-ADNI; work package 5), European multicentre study. 150 individuals underwent clinical and cognitive evaluation using the CANTAB tests, high resolution 3T MRI with MPRAGE and lumbar punctures for the assessment of cerebrospinal fluid (CSF) levels of A□42, tau and p-tau. Individuals were divided into A□+ (CSF-POS) and A□- (CSF-NEG) based on CSF A□42 levels (cut off; 550 pg/ml). The data reported here is the preliminary analysis of the baseline data.

Results: At baseline, CSF-POS individuals showed worse performance relative to CSF-NEG individuals on hippocampal dependent memory tasks (effect sizes ranging from -0.12 to -0.66). Age and education adjusted performance on the paired associate learning (PAL) task of episodic memory was associated with hippocampal volume (Right hippocampus β=-0.03,p=<0.01;R2=0.26; Left hippocampus β=-0.03,p=<0.01;R2=0.20) and CSF levels of tau (β=0.04,p=<0.01;R2=0.12) and p-tau (β=0.24,p=0.02;R2=0.06) with greater errors on the PAL task associated with reduced hippocampal volume and higher CSF levels of tau and p-tau. Similarly, worse performance on the spatial recognition memory (SRM) task was associated with low CSF levels of Ab42 (p=0.01) and higher CSF levels of tau (p=0.05), p-tau (p=0.03) while worse performance on the pattern recognition memory (PRM) task (delayed) was associated with reduced left (p=0.04) and right hippocampal (p=0.05) volume.

Conclusions: These findings show associations between hippocampal and fronto-striatal dependent memory performance assessed using the CANTAB tests, CSF biomarkers and hippocampal volume in biomarker positive amnestic MCI individuals. The findings have implications for identifying MCI patients at risk of developing AD and enriching a more homogenous population for clinical trials with memory deficits, neurodegeneration and A□/tau biomarker abnormalities consistent with prodromal AD populations.

20 **Distinguishing Adults with ADHD from Adults without ADHD Symptoms with Computerized Cognitive Tests**

Wolters J[1 2], Sambeth A[1], Riedel W[1]

[1]Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands, [2]Department of Pedagogy, Fontys University of Applied Sciences, Eindhoven, The Netherlands

Methodological Question Being Addressed: In what way can computerized cognitive tests reliably distinguish adults with ADHD from adults without symptoms of ADHD?

Introduction: Attention Deficit Hyperactivity Disorder (ADHD) is a developmental disorder characterized by behavioural dysfunctions such as inappropriate inattention, impulsivity and overactivity. ADHD diagnosis is based on observable behaviour criteria as stated in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V). However, the cognitive impairments seen in child and adult ADHD can be neuropsychologically assessed, that is, measured by standardized tests. In diagnostic- and treatment research in adult ADHD, many different neuropsychological tests are used as indices of symptom severity. This leads to the question which tests are reliable indices in adult ADHD-studies. The aim of our present study is to validate cognitive tests as diagnostic markers of ADHD-symptoms by studying differences in cognitive task performance between adult subjects with diagnosis ADHD (diagnosed group), subjects without diagnosis but meeting at least four out of six ADHD diagnose criteria (symptom group) and subjects with a maximum of one out of six ADHD diagnose criteria (controls).

Methods: The experiment is still ongoing, but preliminary analysis is reported on a sample of 62 participants (age mean=21.97, SD=3.08, range 18-31; 75.8% females). The ADHD group consisted of 15 participants, the symptom group 23 and the control group included 24 participants.
Main outcome measures were commission errors and reaction time of correct responses on the continuous performance test (CPT) as measures of attention and impulsivity respectively, percentage long term rewards on the choice-delay test (CDT) as a measure of impulsivity and number of immediate correctly recalled words on the 30-word visual verbal learning task (VVLT) as measure of working memory capacity. Subjects were tested three times: the first session was for familiarization. The second and third sessions were identical and were used to determine test-retest reliability.

Results: The main outcomes of the CPT are presented in table 1. Reaction times of the control group resemble closely those reported by Epstein et al. (2003): 394 ms (SD=69). Test-retest analysis of the reaction time showed a correlation of $r = 0.89$, $p < .01$. and $r = 0.54$, $p < 0.01$ for commission errors.

Table 1: mean performances on CPT outcome measures

	Reaction time ms	Reaction time ms	Commission errrors	Commission errors
	Session 2	Session 3	Session 2	Session 3
ADHD	390 (SD=65)	392 (SD=57)	1.31 (SD=1.25)	1.85 (SD=2.04)
Symptom	388 (SD=62)	381 (SD=62)	1.35 (SD=1.34)	0.87 (SD=1.40)
Control	403 (SD=103)	404 (SD=93)	0.85 (SD=1.53)	1.10 (SD=1.77)

The percentages of long term reward of the CDT are presented in table 2. Test-retest analysis showed a correlation of $r = 0.60$, $p < 0.01$.

Table 2: percentages of long term reward of the CDT

	Long term reward	
	session 2	session 3
ADHD	90.0 (SD=18.0)	74.4 (SD=28.5)
Symptom	78.4 (SD=25.9)	77.3 (SD=27.8)
Control	86.0 (SD=26.6)	88.4 (SD=24.0)

The number of correctly recalled words of the VVLT is presented in table 3. Test-retest analysis showed a correlation of $r = .44$, $p < .05$ for the first trial, $r = .70$, $p < .01$ for the second trial and .76, $p < .01$ for the third trial.

Table 3: number of correctly recalled words per trial per session

	Session 2			Session 3		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
ADHD	10.0 (SD=2.5)	15.0 (SD=4.2)	17.9 (SD=4.5)	10.7 (SD=3.2)	14.87 (SD=4.1)	17.7 (SD=6.1)
Symptom	9.9 (SD=3.8)	14.1 (SD=6.2)	16.5 (SD=7.2)	11.8 (SD=4.5)	14.8 (SD=6.2)	17.6 (SD=8.0)
Control	10.3 (SD=3.3)	15.1 (SD=3.8)	18.8 (SD=5.1)	10.8 (SD=4.0)	15.6 (SD=5.0)	18.3 (SD=5.9)

Conclusions: Based on preliminary results, the dependent measures of the cognitive tasks show promising reliability outcomes, especially the reaction time on the CPT and number of correctly recalled words on the second and third trial. The CDT outcomes show mixed results and a lesser reliability than the CPT and VVLT, which might indicate that the CPT is less useful as a cognitive measure of ADHD symptoms. Whether the discussed differences in outcome measures of all tests are statistically significant, is still inconclusive, since presented results are preliminary.

References: Epstein, J.N., Erkanli, A., Conners, C.K. , Klaric, J., Costello, J.E., & Angold, A. (2003). Relations between continuous performance test performance measures and ADHD behaviors. *Journal of abnormal child psychology, 31*(5), 543-554.

21

The Utility of Patient Reported Cognitive Impairment Deficits in Late Life Depression

Nelson C[1], Bickford D[1], Mackin R[1]

[1]*University of California, San Francisco*

Methodological Question Being Addressed: Are patient-reported cognitive scales useful in late life depression?

Introduction (Aims): Cognitive problems are common in patients with Late Life Major Depressive Disorder (MDD). Cognitive impairment can be the result of depression, early degenerative disease, or both. Neuropsychological (NP) test performance is the most objective method of assessing cognition; however, patient reports may be useful screening tools or may be more sensitive to subtle cognitive changes. The current study is a hypothesis generating exploratory study of a patient-reported cognitive questionnaire (PDQ). Specifically we examined how patient reports correlate with NP test performance, depression severity, and structural brain indices in 24 regions of interest.

Methods: Patients aged ≥ 60 years with SCID diagnosed MDD and a Hamilton Depression Rating Scale (HDRS) score ≥ 19 participated in this study. Patients completed a 4T MRI brain imaging battery, a NP test battery and the 20 item PDQ questionnaire. PDQ items were scored 0-4 on a Likert scale ranging from 'never' to 'almost always'. Subscales for memory and executive dysfunction each included 7 items. The NP tests for memory were the Hopkins Verbal Learning Test, delayed recall (HVLT-DR) and the Wechsler Adult Memory Scale, Logical Memory subtest (WMS-LM). Tests of

executive function included Trail Making Test part B, the Stroop Color Word Test, the Delis Kaplan Executive Function Test (DKEFS) and the Symbol Digit Modalities Test. NP scaled scores were calculated based on age. Patients with at least one NP test score ≥ one SD below the mean were considered "impaired." PDQ scores and the NP test scores were correlated with depression severity, with each other, and with 14 regions of cortical thickness and 10 regions of brain volume. Because this was exploratory, we did not control for multiple comparisons.

Results: 86 patients completed the PDQ. Mean age was 72.2 years (range 60-86); 74% were female. HDRS depression severity was modestly correlated with PDQ memory (r=.22, N=85, p=0.04) and PDQ executive function (r=.35, N=85, p=0.001). Patient reported memory difficulty modestly correlated with HVLT-DR (r = -.28, N=83, p=.01) but not the WMS-LM. PDQ executive function scores were weakly correlated with DKEFS (r=-.25, N=85, p=0.02) but not meaningfully or significantly with any other NP executive function tests. 32 patients scored in the cognitively impaired range, but the PDQ global scale was not significantly associated with the impaired/not impaired distinction. PDQ memory scores and executive function scores were only significantly correlated with one of 24 brain regions (left temporal pole volume and left temporal thickness respectively).

Conclusions: The utility of patient-reported cognitive impairment on the PDQ is not established. Patient reported measures were associated with depression and may reflect perception rather than actual performance. Patient reported measures were weakly or not correlated with NP test performance. The PDQ global scale was not useful for distinguishing impaired from unimpaired patients. PDQ scales were only associated with structural changes in one of 24 brain regions. Future studies will need to assess the predictive value and sensitivity of patient reported cognitive measures during treatment of depression.

Disclosures: In the past 12 months Dr. Nelson has served as an advisor or consultant to Corcept, Eli Lilly, Genentech, Lundbeck, Otsuka, Pfizer; he has received honoraria from BMS Canada, Otsuka Asia, and Genentech; he receives research materials from Avid; and owns stock in Atossa (breast cancer device). Mr. Bickford and Dr. Mackin have no conflicts of interest.

22

The Daily Activity Report (DAR) a Novel Measure of Functional Outcome for Serious Mental Illness

Velligan D[1], Sierra C[1], Martin M[2], Fredrick M[1], Maglinte G[3], Corey-Lisle P[4]

[1]Department of Psychiatry, University of Texas Health Science Center San Antonio, San Antonio, TX, [2]Health Research Associates, Inc., Seattle, WA, USA, [3]Amgen Inc., Thousand Oaks, CA, USA, [4]EMD Serono, Rockland, MA

Methodological question being addressed: Are there better ways to assess real-world functional outcomes in short term clinical trials for patients with schizophrenia?

Introduction: The assessment of real-world functional outcomes in clinical trials for medications targeting negative symptoms and cognitive impairment is extremely important. We tested the psychometric properties of the Daily Activity Report (DAR), a novel assessment of productive daily activity.

Methods: We administered the DAR and additional assessments of functional outcome, functional capacity, cognition and symptomatology to 50 individuals with schizophrenia and 25 healthy controls at two time points, one month apart. The DAR records a person's daily activity for seven consecutive days based upon phone calls made three times a day. A total score and scores in three domains; instrumental domestic, social and nondomestic work or school related activities are generated for the DAR.

Results: Test retest reliability based on the Pearson correlation was .67 and the ICC for the DAR total across one month was .80. Reliabilities are higher with shorter time intervals. The total DAR score as well as scores for social activity and non-domestic work/school differed significantly between control and patient participants (p<.0001). Instrumental activity significantly differed by group only on weekends (when control participants were not at work). Total DAR score was significantly correlated with negative symptoms and community functioning (both r=.42 p<.003). DAR scores were only weakly related to positive symptoms.

Conclusions: This study provides preliminary support for the reliability and validity of the DAR using interviewer administration. The development of a patient reported version of the DAR using smart phone technology is the next step.

Disclosures: This study was funded by a research grant from Amgen Pharmaceuticals

23

Exploring Psychomotor Sensitization in Healthy Volunteers Following Repeated Amphetamine Exposure

van Gorsel H [1], van der Aart J[1], de Kam M[1], Timmers M[2], de Boer P[2], van Gerven J[1]

[1]Centre for Human Drug Research, Leiden, the Netherlands, [2]Janssen Research and Development, a Division of Janssen Pharmaceutica N.V., Beerse, Belgium

Introduction: Repeated use of psychostimulants can alter dopaminergic neurotransmission, leading to behavioral and neurochemical 'sensitization'[1]. Sensitization might serve as a biomarker to test the effects of novel compounds which have the potential to reverse dopaminergic hypersensitivity. The goal of this study was to explore the possibility of reliably eliciting and predicting amphetamine-induced behavioral changes following four doses of oral dexamphetamine.

Methods: 16 healthy amphetamine-naive males (mean age 32 years, range 25-44) underwent 5 test days with the NeuroCart® test battery which includes adaptive tracking, body sway, saccadic and smooth pursuit eye movements, finger tapping and the stop signal task (SST). Subjective effects were assessed using Visual Analogue Scales, the Profile of Mood States and the amphetamine sub-scale of the Addiction Research Centre Inventory. Baseline performance was assessed on Day -1. Subsequently, ten subjects received oral dexamphetamine 20 mg (capsule) on study Days 1, 3, 5 and 14. As a control, six subjects received dexamphetamine 20 mg on Days 1 and 14 and placebo on Days 3 and 5. Subjects were blinded to treatment allocation. On each study day, the NeuroCart® test battery was performed twice pre-dose and 1, 2, 3, 4 and 6 hours post-dose. In addition, to investigate the relationship between individual dopamine levels and performance on psychomotor tasks, dopamine D_2/D_3 receptor occupancy was measured with two [^{11}C]raclopride positron emission tomography (PET) scans: at baseline and at one hour after the first dose of dexamphetamine (Day 1). Repeated measures task data were compared with a mixed model analysis of variance with fixed factors treatment, time and treatment by time, random factor subject, subject by treatment and subject by time and the average pre-value per day, and the average of all Day -1 values as covariate. Dopamine D_2/D_3 receptor occupancy was calculated using a simplified reference tissue model.

Results: On the first dosing day, subjects showed significant improvement on all NeuroCart® performance tasks. The different parameters on the SST indicated an improvement in both reaction time and accuracy. Although between-subject variability was large, within-subject performance was highly consistent across all study days. There were no significant group differences (2 vs 4 doses) between Day 1 and Day 14 on any of the tasks. The subjective measures also showed a similar pattern of within subject consistency across the study days. Statistically significant [^{11}C]raclopride dopamine receptor occupancy ($p<0.05$, single-sided paired T-test) was observed in the striatum following dexamphetamine when comparing pre- and post-dose group means. However, up to 7 subjects showed negligible occupancy. The number of missed Go-trials on the SST, measured after PET, correlated negatively with D_2/D_3 receptor occupancy in the caudate nucleus.

Discussion and conclusion: Dexamphetamine consistently improved performance on the Neurocart® performance tasks, but there were no signs of potentiation after repeated dosing. Therefore, this study did not reproduce earlier findings suggestive of response sensitization. This study does, however, provide evidence of consistent and sustained acute amphetamine effects as well as a positive relationship between SST performance and amphetamine-induced elevated brain dopamine levels.

Reference: [1] Featherstone et al. "The Amphetamine-induced sensitized state as a model of schizophrenia"; Progress in Neuro-Psychopaharmacology & Biological Psychiatry 31 (2007) 1556-1571.

Acknowledgement: The authors would like to acknowledge the Department of Nuclear Medicine of the VU Medical Center, Amsterdam, the Netherlands, for acquisition of the PET scans and support in PET data analysis.

Disclosure: HC van Gorsel, J van der Aart and JMA van Gerven were fully employed by the Centre for Human Drug Research (CHDR) at the time this study was carried out. P de Boer and M Timmers are fully employed by Janssen Research and Development. CHDR received a research grant from Janssen Research and Development as co-funding for this study.

24

A Meta-analysis of Pharmacodynamic Testing with the NeuroCart Used in the Early Phase Drug Development of Antidepressants, Stimulants, and CNS Depressant Agents

Hay J[1], van Gerven J[1]

	[1]*Centre for Human Drug Research, Leiden, The Netherlands* **Methodological Question Being Addressed**: To investigate if class- and/or compound-specific profiles can be identified, with a standardized multimodal test battery that is designed to examine drug-sensitive CNS-domains. **Introduction**: The aim of this study was to perform a common analysis, to compare the pharmacodynamic (PD) profiles of different treatments including antidepressants, stimulants and CNS depressant agents. **Methods**: The results from 38 studies with six different CNS active drugs were analyzed. All studies used the NeuroCart®, a test battery consisting of saccadic eye movement, smooth pursuit, adaptive tracking, body sway, visual analogue scales and electroencephalography. A mixed model with average baseline value taken as covariatewas used. Furthermore the least square mean (LSM) was calculated. Estimate of differences (EOD) were calculated by comparing the LSMs of the treatment and the placebo. **Results**: EODs for saccadic peak velocity and body sway (respectively) were for caffeine 135 mg PO (+24%, +115%), citalopram 30 mg PO (+78%, +7%), reboxetine 4 mg PO (-45%,+16%), lorazepam 2 mg PO (-130%, -129%), ethanol 0.6 g/L TCI (-109%, -30%), zolpidem 10 mg PO (-52%, -51%), and diphenhydramine 50 mg PO (-40%, -16%). These two tests from the multidimensional test battery already provided some differentiation between drug classes. Other PD-parameters (smooth pursuit, adaptive tracking, visual analogue scales and electroencephalography) contributed to further diversification of pharmacodynamics profiles. **Conclusion**: Results showed that by using a broad range of CNS tests , drugs can be profiled with unique CNS 'fingerprints'. More drugs need to be compared to determine the relationships between PD effects and mechanism of action.
25	**Is ADAS-cog5 a Suitable Instrument to Assess Cognitive Decline in Patients with Early AD?** Podhorna J[1], Krahnke T[2], Shear M[1], Harrison J[3], for the Alzheimer's Disease Neuroimaging Initiative* *[1]Boehringer Ingelheim Pharma GmbH & Co. KG Germany, [2]Cogitars GmbH Germany, [3]VU University Amsterdam, Netherlands* **Methodological Question Being Addressed**: To test whether the ADAS-cog5 is the optimal instrument for assessing decline in patients with early Alzheimer's disease. **Introduction (Aims)**: Development of new treatments for Alzheimer's disease (AD) has broadened into interventions at early stages which require treatment of individuals who tend to have only modest cognitive impairment with a slow decline over time. The ADAS-cog11 is the most commonly employed instrument for clinical trials but was developed to measure cognition in advanced stages of the AD dementia. Attempts to improve its psychometric properties for early stages by removing items prone to ceiling effects and/or by adding cognitive measures that are known to be impaired early yield a number of ADAS-cog variants. **Methods**: Using Alzheimer's Disease Neuroimaging Initiative (ADNI) data, we tested whether combining the above mentioned approaches, which yields the ADAS-cog5, would provide an optimal instrument for assessing decline in subjects with early AD, if compared to other ADAS variants. Given the interest in enrichment strategies we also examined this latter issue with focus on CSF enrichment markers defined by the diagnostic criteria for early AD. **Results**: The results of our investigations indicate that the decline over time defined by change from baseline on any of the ADAS-cog variants was minimal in subjects with Mild Cognitive Impairment (MCI). Approximately half of MCI subjects fulfilled enrichment criteria for positive AD pathology (t-tau/Abeta ratio >0.39). Impact of enrichment was detectable but subtle in MCI. The annual decline in Mild AD was more pronounced but still modest. Majority of Mild AD subjects (>90%) fulfilled the criteria for positive AD pathology. Direct cross-comparison of the ADAS-cog variants was performed using the signal-to-noise ratio (SNR) with higher values reflecting increased sensitivity to detect change over time. The SNRs were low in MCI but not in Mild AD. The numerically largest SNRs were seen for the ADAS-cog5 in MCI, and, for both the 5-item and 13-item ADAS-cog variants in Mild AD, although associated confidence intervals were large. **Conclusions**: Our analyses showed that the possible utility of ADAS-cog expansion or reduction is less than compelling, particularly in MCI. Thus in Mild AD, it seems that the difference between variants is more likely driven by the addition of items (delayed word recall and digit cancellation) rather than by removing items on which early AD subjects score close to ceiling. **Disclosures**: Jana Podhorna, Michael Shear and Tillmann Krahnke are employees of Boehringer Ingelheim. John Harrison is an advisor for Boehringer Ingelheim but has no conflict of interest to report with regard to this publication. *Data used

	in preparation of this article were obtained from the ADNI database. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf
26	**Suicide Ideation and Behavior Assessment Tool (SIBAT): A Novel Measure of Suicidal Ideation and Behavior and Perceived Suicide Risk** Alphs L[1], Canuso C[2], Williamson D[1], and the SIBAT Consortium *[1]Janssen Scientific Affairs, LLC, Titusville NJ, USA, [2]Janssen Research and Development LLC, Titusville, NJ, USA* **Methodological Question Being Addressed**: Can a tool to assess suicidal ideation, behavior, and risk that is based on comprehensive clinician- and patient-reported assessments, provides a flexible modular structure, and captures changes resulting from intervention be developed? **Introduction**: Suicide is one of the leading causes of preventable death. Clinicians wanting to monitor suicidal ideation, behavior, and risk require a tool that includes all of these components. Ideally, it should allow assessment of change as the result of intervention. The SIBAT is based on an established measure of suicidal ideation and behavior: the InterSePT Scale for Suicidal Thinking–Plus (ISST-Plus). The SIBAT comprises items from the ISST-Plus that have been reorganized into 10 modules to allow for efficient, comprehensive data collection. The SIBAT is divided into patient-self-report and clinician-rated sections. Its modular structure allows for customization, and clinicians can adjust the administration of specific modules to best meet their needs. Thus, responses less susceptible to change (eg, demographics, medical history) are segregated into modules distinct from responses more likely to fluctuate (eg, current suicidal ideation). **Methods**: The SIBAT was created by the SIBAT Consortium, a group of clinical trial and academic experts in scale development, suicidology, and clinical management of suicidal patients. **Results**: The SIBAT Consortium met regularly over 18 months and developed a modular instrument based on clinician consensus, a review of suicide literature, and the ISST-Plus. During revisions of the provisional version of the SIBAT scale, modules were added and item wordings refined. A draft version agreed upon by the SIBAT Consortium was reviewed by 14 patients from a psychiatric clinical research setting and by 686 members of Patients Like Me, an online patient community who self-identified as being at risk for suicide. All participants evaluated items from the patient-reported modules of the SIBAT in terms of semantic clarity, relevance of questions, and adequacy of response choices. This feedback was incorporated and approved by the SIBAT Consortium. A validation study is planned to examine reliability and validity of a computerized version of the instrument, and this study will also include exploratory factor analyses and item response theory analyses. Modifications of selected SIBAT items based on these cognitive interviews will be presented. **Conclusions**: The SIBAT facilitates the comprehensive assessment of suicidal ideation, behavior, and clinician assessment of risk by combining a flexible modular structure and comprehensive patient-reported assessments. Its patient-reported modules include a broad, standardized background of information that is efficiently collected on a computer. This system then provides clinicians with a robust basis for clinical judgments of imminent and long-term suicide risk. The validation of the SIBAT, which incorporates extensive cognitive reviews from diverse sources, will ultimately allow it to be broadly applied across patient populations. **Disclosure**: Support: Janssen Scientific Affairs, LLC. One or more authors report potential conflicts, which are described in the program. **References**: Lindenmayer JP, Czobor P, Alphs L, Anand R, Islam Z, Pestreich L. The InterSePT Scale for Suicidal Thinking (ISST): a new assessment instrument for suicidal patients with schizophrenia. *Schizophr Res.* 2001;49(suppl 1-2):5. Sheehan DV, Alphs LD, Mao L, Li Q, May RS, Bruer EH, Mccullumsmith CB, Gray CR, Li X, Williamson DJ. Comparative validation of the S-STS, the ISST-Plus, and the C-SSRS for assessing the Suicidal Thinking and Behavior FDA 2012 Suicidality Categories. *Innov Clin Neurosci*. 2014;11(9-10):32-46.
27	**Design of the Schizophrenia Disease Recovery Evaluation and Modification (DREaM) Study** Fu DJ[1], Turkoz I[2], Alphs L[1] *[1]Janssen Scientific Affairs, LLC, Titusville, NJ, [2]Janssen Research & Development, LLC, Titusville, NJ*

	Methodological Question Being Addressed: What impact do long-acting injectable (LAI) antipsychotics have on disease progression and disease modification in patients with recent-onset schizophrenia or schizophreniform disorder? **Introduction**: For patients with recent-onset schizophrenia, recurrent relapses and persistent cognitive deficits can contribute to clinical and functional deterioration. Therefore, providing adequate treatment within the first 5 years following the onset of symptoms represents a critical period in the pathophysiology of the disease. LAI antipsychotics may be an effective treatment option for recently diagnosed patients with schizophrenia because they provide certain knowledge of adherence. However, few comparative effectiveness studies evaluating the use of LAIs and oral antipsychotics (OAs) in this population have been completed. The objective of the DREaM study is to compare the efficacy of a once-every-3-months LAI vs OAs in disease progression and the potential for disease modification in patients with recent-onset schizophrenia. Data supporting disease modification will be based on a totality-of-evidence approach that evaluates the course and pathophysiology of the illness with measures of symptoms, functioning, and biological change. **Methods**: DREaM (NCT02431702) is a prospective, matched-control, randomized, open-label, flexible-dose study in subjects with recent-onset schizophrenia or schizophreniform disorder that will compare disease progression and disease modification following treatment with LAI or OAs. The DREaM study includes 3 treatment phases: 2-month open-label run-in (Part 1), 9-month disease progression (Part 2), and 9-month extended disease progression/modification (Part 3). After completing a run-in with OA treatment, patients will be randomized in a 1:2 ratio to flexible-dose LAI or OAs, respectively. After 9 months, the OA group will be further randomized 1:1 to LAI or continued on flexible-dose OAs. Changes in cognition, patient functioning, and volume of brain intracortical myelin will all be assessed as measures of disease progression (Parts 2 and 3) and modification (Part 3). The overall primary endpoint to establish evidence of disease progression will be time to first treatment failure, defined as psychiatric hospitalization due to worsening symptoms; deliberate self-injury, suicide ideation, or violent behavior; new arrest/incarceration; discontinuation of antipsychotic treatment due to inadequate efficacy or safety; treatment supplementation with another antipsychotic; or increase in psychiatric services. **Results**: Approximately 250 subjects will be randomly assigned for this study. Analysis will require distinct approaches to each of the major endpoints related to symptoms, functioning, and biological changes for establishing evidence of disease progression and modification. **Conclusions**: Key innovations of the DREaM study include randomized matched-control of patients with a randomized delayed-start design. It is anticipated that study results based on these design parameters will identify important insights into disease progression and potential disease modification in recent-onset schizophrenia. The study will evaluate whether LAI treatment can slow disease progression and possibly modify disease course compared to OAs by tracking changes in cognition, functioning, and brain imaging. **Disclosures**: Study supported by Janssen Scientific Affairs, LLC. Several authors report potential conflicts of interest, which are described in the program. **References**: 1)Fu DJ, Bossie CA, Sliwa JK, Ma YW, Alphs L. Paliperidone palmitate versus oral risperidone and risperidone long-acting injection in patients with recently diagnosed schizophrenia: a tolerability and efficacy comparison. *Int Clin Psychopharmacol*. 2014;29(1):45-55. 2) Bartzokis G, Lu PH, Raven EP, et al. Impact on intracortical myelination trajectory of long acting injection versus oral risperidone in first-episode schizophrenia. *Schizophr Res*. 2012;140(1-3):122-128.
28	**Bipolar Depression: Acute Stable Response to Medication as a Predictor of Long-term Treatment Response** Iosifescu D[1], Tsai J[2], Pikalov A[2], Kroger H[2], Loebel A[2] *[1]Icahn School of Medicine at Mount Sinai, New York, NY, [2]Sunovion Pharmaceuticals Inc., Fort Lee, NJ and Marlborough, MA* **Methodological Question Being Addressed**: Whether 2 consecutive weeks of response prior to study endpoint can be used as an early indicator of the likelihood of long-term efficacy in studies of bipolar depression; if so, this may be a meaningful endpoint in future studies in this patient population. **Introduction (Aims):** The aim of the current *post hoc* analysis was to evaluate the predictive value of acute treatment response on long-term treatment response at 6 months, in patients receiving monotherapy medication for bipolar disorder. **Method:** This observed cases analysis included patients with bipolar I depression (DSM-IV-TR criteria and baseline Montgomery Åsberg Depression Rating Scale [MADRS] score ≥20) who completed 6 weeks of double-blind, placebo-

controlled treatment with lurasidone (fixed-flexible doses of 20–60 mg/d and 80–120 mg/d; dosages combined) followed by 6 months of open-label treatment with lurasidone (flexible doses of 20-120 mg/d). Response was defined as ≥50% reduction from double-blind baseline on the MADRS total score. In addition, acute stable response was defined as 2 consecutive weeks of meeting response criteria during the initial 6 week double-blind treatment, and partial response was defined as ≥25% but <50% reduction on the MADRS. The proportion of responders at 6 months was determined among patients who were non-responders, partial responders, and responders at each week during weeks 2-6. The predictive value of acute stable response at any time during the initial 6 week treatment and at weeks 5-6 specifically, for long-term response at 6 months, was evaluated with sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the ROC (receiver operative characteristic) curve (AUC) . Additionally, we determined the proportion of responders at 6 months with initial improvement in CGI-BP-S (clinical global impression of severity, bipolar depression) ≥1 during weeks 1-6 of acute treatment.

Results: The majority of patients who completed 6 months of open-label lurasidone treatment following 6 weeks of acute double-blind lurasidone treatment were responders at endpoint, regardless of their response status at weeks 2-6. Among non-responders, partial responders, and responders at 6 weeks, the proportion of responders at the end of 6 months of extension treatment was 84% (16/19), 79% (30/38), and 96% (93/97), respectively. The ability of 2 consecutive weeks of acute response to predict long-term response at month 6 was similar, regardless of whether acute stable response occurred at any time during weeks 1-6 (sensitivity, 61%; specificity, 80%; PPV, 97%; NPV, 18%; AUC_{ROC}, 0.776), or stable response occurred specifically at weeks 5-6 (sensitivity, 58%; specificity, 87%; PPV, 98%; NPV, 18%; AUC_{ROC}, 0.788). 139 patients completed 6 months of extension treatment, of whom 94% experienced improvement in CGI-BP-S ≥ 1 point during the 6 weeks of initial double-blind treatment with lurasidone.

Conclusions: Based on the AUC_{ROC}, the predictive value of acute stable response was fair-to-moderate. Most patients who responded after 6 months of treatment had demonstrated clinically meaningful improvement by the end of initial acute treatment. The majority of responders and partial responders at the end of acute treatment achieved full response at the end of extension treatment.

Disclosure: Sponsored by Sunovion Pharmaceuticals Inc.
Clinicaltrials.gov identifier: NCT00868452

29

Extrapolating Outcomes from Clinical Studies for Economic Evaluation: A Case Study in Schizophrenia

Rajagopalan K[1], Trueman D[2], Crowe L[2], Patel P[1], Loebel A[1]

[1]Sunovion Pharmaceuticals Inc., Marlborough, MA, [2]Abacus International, Bicester, Oxfordshire, United Kingdom

Methodological Question Being Addressed: How should outcomes from clinical trial data be extrapolated for use in economic evaluations?

Introduction (Aims): Cost-effectiveness evaluations of new technologies that treat chronic conditions adopt lifetime horizons to model long-term benefits associated with their use. Therefore, such evaluations require the extrapolation of outcomes beyond the duration of clinical studies. Cost-effectiveness evaluations of atypical antipsychotic treatment have suggested that a key determinant of cost-effectiveness is the rate of relapse. We aim to demonstrate the life-time extrapolation of 12-month data from a double-blind, non-inferiority study of two atypical antipsychotics in adults with schizophrenia.

Methods: The National Institute for Health and Care Excellence Decision Support Unit guidance was used to extrapolate the 12-month observed risk of relapse to generate life time transition probabilities of time-to-relapse. The extrapolation was performed using the following 4-step process: 1) Log-cumulative hazard plots were compared to determine the validity of the proportional hazards assumption; 2) Parametric survival models using alternative distributions for the baseline hazard (exponential, Weibull, Gompertz, loglogisitic, lognormal and generalized gamma) were fitted to the data; 3) The models were compared using the Akaike information Criterion (AIC), Bayesian Information Criterion (BIC) and Martingale residuals, and finally 4) Extrapolations from best fitting models were validated through clinical expert review to determine the final model used to generate transition probabilities.

Results: Inspection of the log-cumulative hazard plot suggested hazards were not proportional prior to day 100, but became more proportional thereafter, and therefore all model were estimated assuming proportional hazards or proportional odds. AIC and BIC statistics suggested that exponential, lognormal and Gompertz models fitted the data similarly well. This finding was supported by visual inspection of the Martingale residuals, which suggested the

generalised gamma, lognormal, and Gompertz model all provided reasonable fits to the data. Following review of the candidate extrapolation assumptions by clinical experts, the lognormal distribution was selected for the base-case of the economic evaluation, with the Gompertz distribution used in sensitivity analysis. The lognormal, Gompertz and exponential distributions provided 3 year risks of experiencing relapse of 47%, 32% and 57%, respectively.

Conclusions: Extrapolation of short-term clinical outcomes to the lifetime horizon of patients is necessary to evaluate the long-term economic benefits of antipsychotics that are used to treat chronic conditions such as schizophrenia. The assumptions selected can have large effects on model outcomes. This analysis suggests that using a systematic approach for extrapolation of shorter-term data can help minimise uncertainty in health technology assessments presented to payor decision-makers.

Disclosures: This research was sponsored by Sunovion Pharmaceuticals Inc. One or more authors report potential conflicts which are described in the program: KR, PP and AL are employees of Sunovion Pharmaceuticals Inc; DT and LC are employees of Abacus International, which has received consultancy fees for this research from Sunovion Pharmaceuticals Inc.

30

Open Translational Science in Schizophrenia

Wilcox M[1], Savitz A[1]

[1]*Janssen Pharmaceuticals*

Methodological Question Being Addressed: What can be learned about therapeutic efficacy/safety, natural history of disease, and methods development by analyzing 17 Janssen paliperidone clinical trials and 10 NIH-funded genetic studies about schizophrenia together in an open science framework?

Introduction: Data from pharmaceutical clinical trials and NIH funded studies about schizophrenia have never been analyzed together before. The Open Translational Science in Schizophrenia (OPTICS) project is a new development that is designed to be a true interdisciplinary approach to addressing fundamental questions about the disorder and treatment of patients with schizophrenia.

Aims: Conduct a pilot project to demonstrate the value of an open-science approach using pharmaceutical clinical trial and federally-funded observational data to:

1. Advance efficacy and safety of medicines for schizophrenia;
2. Increase understanding of schizophrenia, including disease natural history, subtypes, and causes; and
3. Contribute to the development of analytic and design methods for disparate data types, including novel statistical methods and research designs.

Methods: Data: Collections of Janssen's paliperidone clinical trials (N=17 trials) and NIH genetic/genomic data about schizophrenia (N=10 studies) are being used. The collections are available on the Yale Open Data Access site (YODA) and the NCBI Database of Genotypes and Phenotypes (dbGaP) sites.

Process: An open invitation has been issued to researchers worldwide to collaborate in the analyses. The Harvard Catalyst Clinical and Translational Science Center will issue an RFA to provide funding for researchers. At the conclusion of the analysis period (Q4 2016), researchers will meet to discuss results and prepare the publication. All results passing peer review will be published in an open-access online journal. Finally, the pilot will be evaluated with the goal of replicating it for other neuropsychiatric disorders.

Note that this is not the establishment of a data repository with a common data model for use in perpetuity; instead, it is a time-limited open-science collaboration.

Results/Discussion: This is the first time data about the causes of the disorder and data from clinical trials of therapies will be available to researchers in one place. The ability to analyze these datasets together will enable researchers to address questions about the disease, therapies, and analytic methods in ways not possible before now.

The project was announced in June of this year. Members of the project's advisory board include Yale University School of Medicine, Rutgers University, Harvard T.H. Chan School of Public Health, the National Institute of Mental Health (Genomics Branch), and Janssen Pharmaceutical Research & Development.

Participant researchers will agree to meet the data access and use requirements of all data owners. They also must agree that any knowledge generated from this project (e.g., publications, models predicting outcome) will be dedicated to the

	public and will be free for everyone to use. One of the goals of the project is to encourage collaborations among industry and academia, including those that will exist outside traditional disciplines, such as econometrics, bioinformatics, and computer science. **Disclosures**: Drs. Wilcox and Savitz are employed by Janssen, LLC. https://sites.google.com/site/opticsschizophrenia/ http://yoda.yale.edu/optics-trial-bundle http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/collection.cgi?study_id=phs000887.v1.p1