Consensus Workshop on Standardization of Reward Processing Tasks

FASEB Campus 9650 Rockville Pike, Bethesda, MD 20814

February 18-19, 2016

Meeting Summary (Final Draft date April 26)

Attendees: Linda Brady (NIMH), Joseph Calabrese (University Hospitals, Case Medical Center), Rosa Canet-Aviles (FNIH), Will Carpenter (University of Maryland; NIMH), Bruce Cuthbert (NIMH), Tiffany Farchione (FDA), Carlos Forray (Lundbeck), Holly Garriock (NIMH), Steven Grant (NIDA), Grag Hajcak (Stony Brook University), Bastian Hengerer (Boehringer Ingelheim), Dorothy Jones-Davis (FNIH), Brian Knutson (Stanford University, *by phone*), George Koob (NIAAA), Joseph Menetski (FNIH), Sarah Morris (NIMH), Jenni Pacheco (NIMH), Diego Pizzagalli (McLean Hospital/Harvard Medical School), Stephane Pollentier (Boehringer Ingelheim), Bill Potter (NIMH), Madhukar Trivedi (University of Texas Southwestern), Uma Vaidyanathan (NIMH) Pat Walicke (NINDS), James Waltz (University of Maryland School of Medicine); Lois Winsky (NIMH); Jared Young (University of California San Diego); David Zald (Vanderbilt University)

Meeting Summary

I. Introduction

This meeting was convened to gather scientific researchers from academia as well as from governmental institutes, regulatory agencies, government funding organizations and pharmaceutical companies together to facilitate communication and accelerate the process of aligning objective functional circuit-based measures of "reward" or "positive valence" in Research Domain Criteria (RDoC) that could be coupled with clinical measures to better define and operationalize the potential drug targets of "anhedonia", "amotivation" and "apathy" across a broad range of CNS and addictive disorders. The major goal of the meeting was to arrive at a list of possible assessments to implement in future drug development trials. We were especially interested in identifying measures that reliably produce a signal in individuals that is robust enough to draw strong conclusions about pharmacological target engagement and that could be used across indications and patient groups. The discussions focused on fMRI, EEG, ERP, and stand-alone behavioral measures, especially those with good translational utility in animal and preclinical studies.

II. Use of fMRI and EEG/ERP as clinical endpoints in pharmacology trials

Before discussing specific tasks and measures, the group considered whether fMRI and EEG are sufficiently standardized, sensitive and reliable for use in clinical trials. There has been extensive work to look at the reliability of these measures across sites and administration protocols, albeit with a predominant focus on functional neuroanatomy, rather than endpoints for measuring drug effects. For some of the tasks, administration procedures are based on empirical results, rather than pragmatic conventions, yet these are not always

published nor adhered to when different sites implement the same test. It was noted that there are established methods for calibrating and standardizing scanners across sites and for combining data collected across multiple sites. It remains to be determined, however, whether treatment targets are best operationalized using self-report, task performance or brain-based measures. Functional MRI, in contrast to PET, for example, does not provide a molecular signal, however, there is evidence of association between BOLD signal and molecular processes, as shown by the observed correlation of the BOLD with raclopride and dopamine release. Brain measures and reported/observed behavior often do not correlate (or correlate only modestly) and in clinical neuroscience experiments, task difficulty is often adjusted to equate behavior across subjects in order to aid interpretation of brain differences, to facilitate regional allocation and circuit detection and to establish categorical distinctions between diagnosis groups and (healthy) controls, rather than using performance in the task as a measure. The general conclusion of the group was that although no specific brain-based measure has been identified as optimal for use as a clinical endpoint in a drug development trial, fMRI and EEG/ERPs are feasible and informative for inclusion in clinical trials. Whether sufficiently informative to be used as go/no-go criteria in early phase drug development could not be confirmed, given the sparsity of studies investigating treatment effects.

III. Relevant criteria for evaluating reward processing tasks

The group agreed upon the following criteria (listed in approximate order of descending importance) to be considered when evaluating a task/measure for use as a clinical endpoint of reward processing domains:

1. How valid a test of the construct is the task?
2. Does the task have good psychometric characteristics (including high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, and longitudinal stability)?
3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics) standardized on an empirical basis?
4. Is the task free from floor/ceiling effects which would preclude use in subjects with a range of impairment?
5. Does the task have the same performance characteristics across cultures? Is it free from culture- and language-specific features/stimuli?
6. Is the task sensitive to change and to lack and loss of function?
7. Can the task (or an analog) be used in animals?
8. Can the task be used across age groups?
9. Can the task be used in conjunction with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)?
10. Is there consensus on which metric/score should be considered to be primary?
11. Is adequate normative data available across age, gender, education, ethnicity, and socio-economic status?

12. Are the relationships between task performance and neural signal(s) known?
13. Are the relationships between task performance and clinical feature(s) known?
14. Is the task feasible for administration across sites?
15. Is the task feasible for repeated administration (e.g., are alternate forms available, etc.)?
16. Can the task be used as a stand-alone behavioral task?
17. What work is needed to get this task ready for use in clinical trials?

IV. Discussion of specific reward processing tasks

For the purpose of this meeting and on the basis of a large body of research in clinical and normative populations demonstrating distinct neural processes related to different aspects of anticipation of and response to motivationally salient stimuli and outcomes, anhedonia was specifically defined as something more than a simple loss of pleasure, and more broadly related to reward processing. There are many factors involved in a hedonic response (including reward valuation, expectancy, effort valuation, responsiveness to reward, reward learning, and more) and there is likely to be considerable heterogeneity within patients displaying or reporting anhedonia, within and across different disorders, as there could be impairment in one or several of these key processes and so there are several processes that could be tested to measure improvement.

The group's discussion of specific tasks largely focused on the following set of computer-administered measures which were proposed as a straw-man battery:

1. Simple reward task: Participants select from among stimuli, such as cards or doors, which reveal reward or no reward on a random basis. Performance of this task does not require any probabilistic reasoning or learning and thus isolates the immediate neural response to receipt of reward or loss, independent of anticipation or expectation.

2. Monetary Incentive Delay task: Participants are shown an incentive cue which indicates the size of the reward followed by a target stimulus to which the participant must respond in order to win or to avoid losing the indicated reward. The probability and size of the reward and difficulty of response can be varied. This paradigm allows examination of reward prediction, anticipation, and outcome processing.

3. Probabilistic Reward task: Participants are asked to respond to two hardly distinguishable cues, of which one is more frequently rewarded. The ensuing response bias isolates learning-related processes involved in integrating reinforcement outcomes in order to optimize response selection. This paradigm allows examination of reward learning.

4. Effort Expenditure for Rewards task (EFfRT): A button-pressing task in which the participant may choose between a difficult (e.g., button presses with pinkie finger of non-dominant hand for 30 second) or an easy assignment (e.g., button presses with index finger of dominant hand for 7 second) for three levels of reward. The probability of obtaining the reward at successful completion of the assignment is manipulated to 80:20. This task isolates participant willingness to exert effort relative to size and likelihood of reward. This paradigm allows examination of reward and effort valuation.

Each of these tasks was evaluated on the list of criteria provided above. A few examples of the criteria for each of the four tasks are listed below.

Criteria 2: Does the task have good psychometric characteristics (including high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, and longitudinal stability)?

- For the Monetary Incentive Delay task there are data on internal reliability, and for 2+ year test/retest reliability where both the nucleus accumbens and insula activation has shown to have good reliability (rs > .5 for large incentives; Wu et al., 2014).
- There are good psychometric characteristics of the simple reward task, as published in several papers.
- For the EFfRT task, there are data for a short-interval test-retest reliability, and internal reliability, although these have not yet been published. Further psychometrics of the test have been published elsewhere[1].
- For the Probabilistic Reward task, internal reliability has not been evaluated. Test-retest reliability was measured at 0.57 over 38 days in unselected individuals[2]. This was replicated in an independent unselected sample: r = 0.50-0.56 over 39 days[3]. ROC analyses to understand sensitivity/specificity have not yet been performed, although the data are available. There are 5 alternate forms available, which will minimize the practice effects. And limited longitudinal stability has been assessed over a period of 40 days.

Criteria 7: Can the task (or an analog) be used in animals?

- The EFfRT task can be done in animals, by placing a rat in a T-maze and observing the selection of either a high effort option (needing to scale a barrier) for a high reward (additional food pellets), or a low effort option (no barrier) for a low reward (less food pellets)[4].
- A conceptually identical version of the Probabilistic Reward Task has been developed for rodents (9). The human and rodent versions have yielded the same patterns when testing pharmacological challenges (9; 10), nicotine withdrawal (11), or stressors (12; Der-Avakian et al., in preparation).

Criteria 14: Is the task feasible for administration across sites?

- For the Monetary Incentive Delay task, there is evidence of utility across sites. It is currently being used in the IMAGEN study[5], a large multisite study in Europe, as well as the FAST-MAS trials.

In summary, it was thought that these tasks covered many of the key processes that have been shown to be related to reward processing and would provide coverage of the various mechanisms that may give rise to anhedonia across various mental disorders. This set of tasks was, however, suggested only as a stimulus for starting and focusing the discussion of specific tasks that could be considered for use in clinical trials, and the group did not reach a final consensus on tasks to be included in a standard battery for reward processing circuitry. Many other tests which may be appropriate for such use were not discussed in detail but have been reviewed elsewhere[1,6,7].

V. Conclusions and next steps

At the conclusion of the meeting it was clear that although there were no specific behavioral tasks or neuroimaging protocols unanimously recommended for immediate use in pharmaceutical trials targeting anhedonia, there are several tasks that are quite close to being ready for use in trials as measures of engagement of neural or behavioral targets or, potentially, as endpoints. The following issues require resolution via future task development efforts:

- Additional studies are required to complete the knowledge base for each of the individual paradigms. This includes not only the generation of additional data on reliability and validity, but also establishing the data sets that allow to decide on basic procedural aspects and the standardization of the task (e.g., number of trials per run, number of runs, standardization of cues (e.g., symbols, digits, letters) and type and amount of rewards.

- Many anhedonia-related tasks and measures have been developed and studied within individual diagnostic groups. Further studies of the relationships between behavioral and/or neural signals obtained using these tasks and anhedonic pathology across diagnostic groups of psychiatric, neurologic and addictive disorders are needed in order to determine whether the observed relationships are robust measures of reward circuitry disturbances independent of disorder.

- Given the relative isolation by which experience with most of these paradigms has been collected, an effort should be undertaken to create a centralized, accessible repository where interested investigators can share and upload information on technical, procedural, analytical and outcome related aspects.

- Many of the existing studies on anhedonia-related tasks and measures have been conducted using cross-sectional designs. It will be important to determine, via longitudinal research, the extent to which these tasks are able to reliably detect change in clinical pathology. It may be that some measures can detect changes prior to the manifestation of clinical symptoms, other changes may be state-independent and serve as trait/disease markers rather than measuring symptom severity. Longitudinal research and observations at different disease stages can support the clinical relevance of the observed changes.

- As outlined above, there was consensus that multiple tasks are needed to assess the various aspects of anhedonia but the candidate tasks have largely been studied in isolation, rather than complementary. Similar to the elegant study by Reddy, Horan et al (ref. 1,6), who compared five effort-based decision making paradigms in one clinical trial, conducting a trial using different reward processing paradigms within one setting would certainly advance the field. As with cognition, one could well envision a reward processing test battery, which could provide an objective quantification of anhedonic complaints and provide guidance as to the relevant impaired subdomain/circuitry in the different disorders. Obviously, if the intent would be to use this in a general clinical setting, the paradigms would be required to be valid as stand-alone tests, without the additional complexity of simultaneous fMRI or EEG assessments and some tests have not been optimized for "stand-alone" administration.

- It will be important, once such a test battery is designated, to optimize test administration procedures to minimize subject burden and maximize psychometric properties. This process should include collecting normative data from a geographically and ethnically diverse participant group which includes a broad age range. It will be important to bear in mind that when a single test is used to measure a single aspect of anhedonia, interpretation of changes in that aspect may be an artifact of psychometric properties of that task (e.g., greater practice effects) rather than differential treatment effects.

- Attention to a bench-to-bedside approach involving animal to human translation is important and should be aligned in terms of test paradigms to the extent possible. While not all aspects of reward processing may have a pre clinical correlate, there is evidence supporting the conservation of some mechanisms across species. Aligned assays of components of reward between animals and humans should facilitate both target identification and validation with application to understanding pathophysiology as well as novel drug development. Selection and funding of preclinical studies (mainly by NIH institutes) could take into account which paradigms were prioritized for human studies to accelerate the rate at which more predictive animal studies of human effects are developed.

- When considering the use of a novel endpoint in a clinical trial (with the intention of seeking an indication based on that endpoint) the recommendation is to engage the regulatory agencies early on in the process to ascertain whether the endpoint

would be considered acceptable and, if not, what data may need to be generated to make it acceptable.

- Ultimately, the clinical relevance and meaningfulness of changes in these paradigms will drive regulatory acceptance, be it for diagnostic or treatment evaluation purposes. At the current stage, while the knowledge base is growing, the link between these technical assessments and the clinic is not sufficiently established. Additional work on reward processing paradigms should therefore always keep a focus on clinical aspects, by ensuring that the clinical correlates of the behavioral tasks, if not in the primary focus, at least are well documented to enable future analyses.

1. Horan, W. P. *et al.* Effort-Based Decision-Making Paradigms for Clinical Trials in Schizophrenia: Part 2—External Validity and Correlates. *Schizophr Bull* **41,** sbv090–1065 (2015).
2. Pizzagalli, D. A., Jahn, A. L. & O'Shea, J. P. Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biological Psychiatry* **57,** 319–327 (2005).
3. Santesso, D. L. *et al.* Individual differences in reinforcement learning: Behavioral, electrophysiological, and neuroimaging correlates. *NeuroImage* **42,** 807–816 (2008).
4. Salamone, J. D., Cousins, M. S. & Bucher, S. Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit procedure. *Behavioural Brain Research* **65,** 221–229 (1994).
5. Schumann, G. *et al.* The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry* **15,** 1128–1139 (2010).
6. Reddy, L. F. *et al.* Effort-Based Decision-Making Paradigms for Clinical Trials in Schizophrenia: Part 1—Psychometric Characteristics of 5 Paradigms. *Schizophr Bull* **41,** sbv089–1054 (2015).
7. Young, J. W. & Markou, A. Translational Rodent Paradigms to Investigate Neuromechanisms Underlying Behaviors Relevant to Amotivation and Altered Reward Processing in Schizophrenia. *Schizophr Bull* **41,** 1024–1034 (2015).
8. Wu CC, Samanez-Larkin GR, Katovich K, Knutson B. Affective traits link to reliable neural markers of incentive anticipation. *Neuroimage, **84**:279-89* (2014).
9. Der-Avakian A, D'Souza MS, Pizzagalli DA, Markou A: Assessment of reward responsiveness in the response bias probabilistic reward task in rats: implications for cross-species translational research. Transl. Psychiatry 2013; **3**:e297
10. Pizzagalli DA, Evins AE, Schetter EC, Frank MJ, Pajtas PE, Santesso DL, Culhane M: Single dose of a dopamine agonist impairs reinforcement learning in humans: behavioral evidence from a laboratory-based measure of reward responsiveness. Psychopharmacol. 2008; **196**:221–232
11. Pergadia ML, Der-Avakian A, D'Souza MS, Madden PA, Heath AC, Shiffman S, Markou A, Pizzagalli DA: Association between nicotine withdrawal and reward responsiveness in humans and rats. JAMA Psychiatry 2014; **71**:1238–1245
12. Bogdan R, Pizzagalli DA: Acute stress reduces reward responsiveness: implications for depression. Biol. Psychiatry 2006; **60**:1147–1154