Title

Effort-Based Decision-Making Paradigms for Clinical Trials in Schizophrenia: Part 1—Psychometric Characteristics of 5 Paradigms

Journal

Schizophrenia Bulletin, 41(5)

ISSN

0586-7614

Authors

Reddy, L Felice
Horan, William P
Barch, Deanna M
et al.

Publication Date

2015-09-01

DOI

10.1093/schbul/sbv089

Peer reviewed

Schizophrenia Bulletin vol. 41 no. 5 pp. 1045–1054, 2015
doi:10.1093/schbul/sbv089
Advance Access publication July 3, 2015

Effort-Based Decision-Making Paradigms for Clinical Trials in Schizophrenia: Part 1—Psychometric Characteristics of 5 Paradigms

L. Felice Reddy*[,1,2], William P. Horan[1,2], Deanna M. Barch[3], Robert W. Buchanan[4], Eduardo Dunayevich[5], James M. Gold[4], Naomi Lyons[1], Stephen R. Marder[1,2], Michael T. Treadway[6], Jonathan K. Wynn[1,2], Jared W. Young[7,8], and Michael F. Green[1,2]

[1]Department of Veterans Affairs VISN 22 Mental Illness Research, Education, and Clinical Center, Los Angeles, CA; [2]UCLA Semel Institute for Neuroscience & Human Behavior, David Geffen School of Medicine, Los Angeles, CA; [3]Departments of Psychology, Psychiatry, and Radiology, Washington University, St. Louis, MO, [4]Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD; [5]Amgen, Inc., Thousand Oaks, CA; [6]Department of Psychology, Emory University, Atlanta, GA; [7]Department of Psychiatry, University of California, San Diego, CA; [8]Research Service, VA San Diego Healthcare System, San Diego, CA

* To whom correspondence should be addressed; VA Greater Los Angeles Healthcare System, MIRECC 210A, Building 210, 11301 Wilshire Boulevard, Los Angeles, CA 90073, US; tel: (310) 478–3711 x42941, fax: (310) 268–4056, e-mail: lenafelice@ucla.edu

Impairments in willingness to exert effort contribute to the motivational deficits characteristic of the negative symptoms of schizophrenia. The current study evaluated the psychometric properties of 5 new or adapted paradigms to determine their suitability for use in clinical trials of schizophrenia. This study included 94 clinically stable participants with schizophrenia and 40 healthy controls. The effort-based decision-making battery was administered twice to the schizophrenia group (baseline, 4-week retest) and once to the control group. The 5 paradigms included 1 that assesses cognitive effort, 1 perceptual effort, and 3 that assess physical effort. Each paradigm was evaluated on (1) patient vs healthy control group differences, (2) test-retest reliability, (3) utility as a repeated measure (ie, practice effects), and (4) tolerability. The 5 paradigms showed varying psychometric strengths and weaknesses. The Effort Expenditure for Rewards Task showed the best reliability and utility as a repeated measure, while the Grip Effort Task had significant patient-control group differences, and superior tolerability and administration duration. The other paradigms showed weaker psychometric characteristics in their current forms. These findings highlight challenges in adapting effort and motivation paradigms for use in clinical trials.

Key words: schizophrenia/effort/motivation/psychometric

Translational models suggest that performance-based measures of effort-based decision making could be sensitive to deficits in motivation associated with clinically rated negative symptoms. Effort-based decision making has been investigated in preclinical models of motivational negative symptoms, as well as healthy and clinical samples.[1–3] Animal studies demonstrate a general "law of least effort" in which organisms choose to exert the least amount of physical or cognitive effort necessary to obtain a given level of reward,[4] but when reward levels increase, the animal is typically willing to exert more effort. Decisions about whether to exert more effort for higher levels of reward depend on factors such as perceived effort required to complete the associated task, subjective valuation of potential rewards, and likelihood that the reward will actually be received if the task is successfully completed. Effort-based decision-making paradigms attempt to objectively assess the culmination of these processes—ie, motivated behavior defined as how much effort one is willing to exert for different levels of reward. These types of tasks are now being adapted for use in humans and may provide novel and practical tools for objectively assessing motivation in clinical trials of schizophrenia.

Recent studies suggest that people with schizophrenia show disturbances in effort-based decision making. Six of the 8 published studies in schizophrenia have focused on physical effort (eg, motoric or strength-based), with 4 using button-pressing paradigms[5–8] and 2 using hand grip effort tasks.[9,10] The 4 button-pressing studies found patient-control differences in that patients were less willing to exert effort for rewards; the grip effort tasks did not report patient-control differences. The 2 published cognitive effort studies were discordant in that 1 found an effort exertion deficit[11] and the other reported no deficit

in effort exertion in individuals diagnosed with schizophrenia.[12] These tasks are reviewed in more detail in a companion article (Green MF. et al,[13] this issue). While these initial studies suggest that effort-based decision-making disturbances are likely present in schizophrenia, the methods and sample sizes varied considerably, and no study compared multiple effort paradigms. Furthermore, no study systematically evaluated the psychometric properties of the tasks to determine their suitability as outcome measures in clinical trials.

Tests considered as endpoints in clinical trials should be evaluated on the following 5 characteristics: (1) ability to detect differences between individuals with schizophrenia and controls (2) test-retest reliability, (3) utility as a repeated measure, (4) tolerability and practicality, and (5) relationship to external correlates. These characteristics have been identified through a consensus-based approach as the key criteria for evaluating behavioral measures of cognition for clinical trials in schizophrenia.[14,15] The aim of the current study is to evaluate 5 effort-based decision-making paradigms in schizophrenia for their suitability for use in clinical trials. This report focuses on the first 4 characteristics; relationships with external correlates are covered in a companion article (Horan W.P. et al,[16] this issue).

Methods

Participants

The sample included 94 individuals with schizophrenia and 40 demographically matched healthy controls. Selection criteria for participants with schizophrenia included (1) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnosis of schizophrenia determined with the Structured Clinical Interview for DSM-IV (SCID-I/P),[17] (2) age 18–60 years, (3) no clinically significant neurological disease, (4) no history of serious head injury, (5) no evidence of substance dependence in the past 6 months and no evidence of substance abuse in past month, (6) no history of mental retardation or developmental disability, and (7) clinically stable (ie, no inpatient hospitalizations for 3 months prior to enrollment, no changes in antipsychotic medication type in the 4 weeks prior to enrollment). Neurocognition and clinical symptom assessments were conducted by interviewers trained according to established procedures that included a library of videotaped interviews developed by the Treatment Unit of the Department of Veterans Affairs (VA) VISN 22 Mental Illness Research, Education, and Clinical Center (MIRECC).

Selection criteria for healthy controls included (1) no psychiatric history involving schizophrenia spectrum disorder (including avoidant, paranoid, schizotypal, or schizoid personality disorders), or other psychotic or recurrent mood Axis I disorder according to the SCID-I and SCID-II, (2) no family history of a psychotic disorder among first-degree relatives based on participant report, and (3) no history of substance or alcohol dependence, and no current substance use. Criteria concerning age, neurological disease, and head trauma were the same as listed for participants with schizophrenia. After providing a complete description of the study to prospective study participants, written informed consent was obtained prior to participation in accordance with approval from the Institutional Review Boards at the VA Greater Los Angeles Healthcare System and UCLA.

Procedures

Participants with schizophrenia were administered the battery of effort-based decision-making paradigms twice (baseline, 4-week retest); healthy controls received the battery once. Clinical symptoms and community functioning were assessed at both testing occasions for the participants with schizophrenia. The effort-based paradigms were administered in 2 different orders that were counterbalanced across participants to minimize possible confounding effects of fatigue or interference from previously administered paradigms (Order A: Grip, Deck, EEfRT [Effort Expenditure for Rewards Task], Perceptual, Balloon; and Order B: Balloon, Perceptual, EEfRT, Deck, Grip). The test administrator recorded reward earnings at the conclusion of each effort task and participants were paid cumulative earnings for the 5 tasks at the completion of the testing session, in addition to hourly payment ($15/h) for participating in the study.

Effort Tasks

Deck Choice Effort Task. This task is based on a paradigm used in healthy samples to measure willingness to exert *cognitive* effort for different levels of monetary rewards.[18,19] In the original version and an adaptation used with participants with schizophrenia, there was no explicit labeling of "hard" and "easy" choices, instead participants had to figure out the hard and easy decks through repeated trials, and then develop an affinity for more or less challenging. In the task adapted for the current study, hard and easy trials were explicitly labeled as such in order to remove any cognitive confound and participants made a series of choices from 1 of the 2 decks of cards (supplementary figure 1). The hard deck is composed of cards that alternate between 2 colors (each color requires a different mental activity) and participants alternate on each trial between making judgments about whether the numbers are odd/even or if they are >5 or <5. The easy deck includes cards that are all the same color (requiring a single mental activity, such as saying whether the numbers are odd or even). Participants learn which judgment is associated with which color during practice rounds which are repeated until over 70% accuracy is achieved. The easy deck always earns $0.10

reward, while the hard deck includes equal number of trials worth $0.10, $0.20, and $0.40. There are 10 cards in each deck and 12 choices of decks in each of the 3 high-demand reward conditions. The primary dependent variable is the frequency of hard choices at each reward level.

Perceptual Effort Task. This task is based on the rodent perceptual effort task designed by Winstanley and colleagues (5 Choice Serial Reaction Time task),[20] based on the 5-choice serial reaction-time task,[21] in which subjects must detect a single stimulus within a short or long duration. Instead of a short or long stimulus duration, the task adapted for this study requires identification of a stimulus that varies in contrast from the visual background (supplementary figure 2). The objective is to correctly identify in which of 7 possible locations a faint stimulus appears. Participants choose between easy and hard trials, and the difficulty level corresponds to the contrast in gray-scale between the stimulus and the background (ie, "easy" has more contrast and is easier to detect). Prior to the effort-choice task, individual perceptual thresholds are measured using a titration task. Contrast level is adjusted using a transformed up-down staircase method to select different grayscale values (ranging from 1 to 256) for various perceptual accuracy levels.[22] Task difficulty is adjusted so that the easy trials include a stimulus that is 98% of the participant's threshold for detection (ie, 2% more visible), and the hard trials have a grayscale value that is 101% of their perceptual threshold (ie, 1% lighter on the gray scale). The easy option is always associated with a small reward (ie, $0.10) and the hard option is paired with variable amounts of reward ($0.10, $0.20, $0.30, $0.40). There are 18 trials at each reward level and the primary dependent variable is the percent hard choices at each reward level.

Grip Effort Task. This task has been implemented in fMRI and behavioral studies, consistently showing it is a sensitive measure of behavioral effort-expenditure decision making.[23,24] The task was completed using a commercially available grip force fiber optic response device (HHSC-1 × 1-GRFC; Current Designs, Inc.) that interfaced with a graphically displayed "thermometer" on a computer screen that presented the amount of grip force exerted in real time. Prior to the choice trials, maximum grip strength is individually calibrated based on participants squeezing the grip device with their maximum strength 3 times with right and left hands. During the actual task, each trial begins with a choice stage in which participants are asked to choose between an "easy" and a "hard" grip task. The easy task involves squeezing the handgrip with 50% of the subject's maximum grip force and the monetary reward is always $0.10. The hard task requires the subject to grip the handgrip with 90% of their maximum strength and the reward for the hard task is paired with variable amounts of reward ($0.10, $0.20, $0.40). After subjects make their selection, there is a 3-s execution stage in which subjects attempt to perform the selected grip force task. Computerized graphics provide feedback about the amount of force exerted on an individually adjusted force-o-meter and inform participants whether they earned a monetary reward (supplementary figure 3). There are 18 trials at each reward level and the primary dependent variable is the percent of hard choices at each reward level.

Effort Expenditure for Rewards Task. Based on animal paradigms, EEfRT[25] is a computerized button-pressing game in which the participant chooses easy or hard tasks for variable amounts of reward. The hard task requires an individually calibrated number of button presses to be made within 30 s, with the nondominant pinkie finger. The easy task requires one-third the amount of the individually calibrated hard number of presses to be made within 7 s, with the dominant index finger. The individual calibration phase precedes the practice rounds and choice trials. It requires participants to button-press as many times as possible within 30-s time intervals with both the dominant and nondominant pinkie fingers and after 3 rounds with right and left hands, an average is calculated. The target for the "hard" trials is 85% of this average value; the participant button-presses as rapidly as possible while a computerized graphic illustrates progress toward the goal (supplementary figure 4). There is also a probability component to this task: each trial is preceded by a screen that notifies whether the trial has 50% or 88% probability of being rewarded if it is successfully completed. The specific modifications to the EEfRT task in this study were based on discussions with the task developer (M.T.T.) and included removal of a 12% probability level used in previous iterations (only included 50% and 88%), allowing for individual calibration of "hard" to adapt the required number of button presses, and standardizing the number of trials, so that all participants completed 50 trials. The easy task is always worth $1.00 and the hard task ranges in reward value from $1.24 to $4.30. The rewards are grouped into low, medium, and high bins for analyses and the primary dependent variable is the percent of hard choices within the high probability trials and percent of hard choices at each reward level.

Balloon Effort Task. The Balloon Effort Task involves use of a game controller to press alternating left-right buttons to inflate a balloon on the computer screen until it pops.[6] In the task adapted for this study, participants are told that they will make decisions between 2 response alternatives: an easy option with a lower reward and a more difficult option with a higher reward. We eliminated the 50% probability trials used in a previous implementation of this task, based on published data and correspondence from the task developer (J.M.G.). The easy trials require participants to press the game controller buttons

10 times in order to pop the balloon while the hard task requires 100 button presses to pop the balloon (supplementary figure 5). The easy option is always rewarded with $1.00 while the hard task varies in reward from $3.00 to $7.00. Participants are told that 3 trials will be randomly selected to determine reward amount following the 40 trials of the task; in actuality each participant received a randomized value of either $7.00 or $9.00. The primary dependent variable is the percent of hard choices at each of the 5 reward levels.

Tolerability and Administration Time

Tolerability refers to the participant's view of a test (ie, how much they liked or did not like taking it) and can be influenced by the length of the test, degree of difficulty, or monotony. Participants with schizophrenia were asked immediately after they completed each paradigm to rate on a 7-point Likert scale how unpleasant or pleasant they found it to be 1 = extremely unpleasant; and 7 = extremely pleasant. We also measured administration time for each paradigm to gauge feasibility for use in clinical trials.

Symptom Assessments

The psychiatric symptoms of participants with schizophrenia were evaluated using the Positive and Negative Syndrome Scale (PANSS),[26] the Clinical Assessment Interview for Negative Symptoms (CAINS),[27] and the Negative Symptom Assessment (NSA-16).[28] The PANSS is a 30-item structured interview that yields symptom factor scores on 5 dimensions; we used the positive and negative symptom factor scores to characterize the sample.[29] The CAINS is a 13-item instrument that yields 2 subscales which measure the 2 primary negative symptom factors: Motivation and Pleasure and Expression. The NSA-16 is a global score of negative symptoms, including communication, affect, social involvement, motivation, and retardation. All clinical interviewers were trained through the Treatment Unit of the VA VISN 22 MIRECC. Symptom raters were trained to a minimum intra-class correlation coefficient (ICC) of .80.

Neurocognition

The MATRICS Consensus Cognitive Battery (MCCB)[30] was used to assess neurocognitive functioning. The MCCB includes 10 tests to measure 7 domains of cognition: speed of processing, attention/vigilance, working memory, verbal memory, visual memory, reasoning and problem solving, and social cognition. Standardized *T*-scores were computed for each of the 7 domains, correcting for age and gender. The composite score was based on the average *T*-score from each of the domains, and served as the primary dependent measure in this study.

Statistical Analyses

For each effort task we first calculated the number of inflexible responders (ie, participants who selected all hard across all reward levels or all easy across all reward levels). We conducted all analyses with the entire sample included, and then reanalyzed the data with the all-hard inflexible responders excluded. The focus on the all-hard responses was because they are more problematic for clinical trials: if they are performing at the maximum level of motivation at baseline there is no room for improvement. The results remained largely unchanged after excluding the all-hard inflexible responders, thus the entire sample was retained for the primary analyses. Group differences between participants with schizophrenia and healthy controls were examined using analyses of variance with reward level as a within-subject factor, and effect sizes (ES) were used to compare across tasks.

Currently there is not a consensus on an optimal unitary dependent measure for effort tasks that can be used in correlational or regression analyses. For test-retest reliability and mean-shift analyses, 2 indices served as the dependent measures. The first was calculated by subtracting the percentage of hard choices out of the total possible hard choices at the lowest reward level from the percentage of hard choices out of the total possible at the highest reward level. The second index was the percent of hard choices out of the total possible at the highest reward level only. ICCs were used to examine test-retest reliability in the schizophrenia sample. Practice effects were examined with paired-samples *t* tests; within-group ES were calculated by dividing the mean difference score by its *SD*. Each of these indexes were statistically compared using *z*-tests. We will present the summary descriptive data for tolerability, administration time, and monetary earnings for each task.

Results

Participants

Table 1 provides the demographic characteristics. The 2 groups did not differ in age, parental education, sex, or ethnicity. As expected, participants with schizophrenia had significantly lower education than controls. Eighty-two percent of the participants with schizophrenia were taking a second-generation antipsychotic, 12% a first-generation antipsychotic, 1% were taking both, and 4% were not taking an antipsychotic; current medication type was unknown for 1%. A locally developed extrapyramidal side effects scale[31] was used to assess akathisia, rigidity, and tardive dyskinesia. The ratings in this sample were negligible for all types of extrapyramidal side effect (ie, average side effect rating for participants with schizophrenia = 0.13 on a scale of 0–3). Symptom levels for participants with schizophrenia did not differ over the 2 assessments (baseline and 4-week retest). Of note, different tasks have

slightly different sample sizes. There were no systematic reasons for missing data and no group differences between those missing data on various tasks and those with complete data. Order effects were examined for the 2 counterbalanced orders in repeated measures ANOVAs: group × reward level × order, for each task. There were no differences in demographics, cognition, or symptoms between the 2 order groups and there were no main effects of order or interactions between order and reward or group for 4 of the 5 tasks. The only exception was the Grip Effort in which there was a main effect of order and an interaction between order and reward such that both groups chose hard choices more often for the middle level of reward (equal for both orders for low and high levels of reward). The order in which this effect was observed included Grip Effort as the first task, as opposed to the alternate order that included it as last. It should be noted that although the order of tasks was reversed at the 2 administrations, the design did not permit assessment of the sensitivity of each task to order effects. Three of the 5 tasks were individually calibrated for the "hard" tasks: Perceptual, Grip Strength, and EEfRT. We examined group means for calibration levels and found a significant difference for Perceptual such that controls detected more difficult grayscale values than controls, but no group differences in maximum difficulty level between the groups on the Grip Strength or EEfRT tasks.

Table 1. Demographics, Cognition, and Symptoms of the Sample at Baseline

	Schizophrenia (n = 94)	Controls (n = 40)
Sex (% male)	69%	60%
Age (*SD*)	49.1 (11.7)	47.2 (8.1)
Education (*SD*)	13.1 (1.8)	14.4 (1.9)
Parental education (*SD*)	13.5 (2.5)	12.5 (3.3)
Race (%)		
Black	38.3%	32.5%
White	53.2%	60%
Asian	3.2%	2.5%
Hawaiian/Pacific Islander	4.3%	5.0%
Mixed/Other	1.1%	0%
Ethnicity (%)		
Hispanic	16%	20%
Non-Hispanic	84%	80%
MCCB overall composite (*SD*)	31.6 (12.2)	47.0 (8.1)
Symptoms (*SD*)		
PANSS Positive	18.5 (7.5)	
PANSS Negative	16.0 (7.0)	
CAINS Total	21.0 (9.4)	
CAINS MAP	15.6 (7.0)	
CAINS Expression	4.9 (4.1)	
NSA-16 Global	3.5 (1.3)	

Note: MCCB, MATRICS Consensus Cognitive Battery; PANSS, Positive and Negative Syndrome Scale; CAINS, Clinical Assessment Interview for Negative Symptoms; MAP, Motivation and Pleasure; NSA, Negative Symptom Assessment.

Schizophrenia—Healthy Control Group Differences

There was a significant main effect of reward on all 5 effort-based decision-making paradigms, with increased willingness to exert effort at the higher reward levels (table 2). Thus, all 5 tasks worked in the sense that motivation and effort increased in accordance with reward and probability of earnings. There was a statistically significant main effect of group on the Deck Choice Effort and the Balloon Effort in which participants with schizophrenia showed significantly less willingness to exert effort for rewards (figure 1). There was a trend-level group difference on the Grip Effort, in the same direction. There were significant group by reward interactions on the Deck Choice and Grip Effort, with a similar trend for the EEfRT task. For all 3 of these tasks, participants with schizophrenia and controls were roughly equivalent in frequency of hard trial choices at low levels of reward or probability, but controls showed a greater increase than individuals with schizophrenia in willingness to select hard trials as reward/probability values increased. There was also a significant group × probability interaction on the EEfRT task: equal percentage of hard choices between participants with schizophrenia and controls at 50%, but controls significantly greater than schizophrenia at 88% probability. Perceptual Effort was the only task that did not show either a significant main effect of group or a group by reward interaction.

Test-Retest Reliability

Test-retest reliability data are summarized in table 3. Traditionally, ICC values greater than 0.75 represent

Table 2. ANOVAs for Patient-Control Differences at Baseline

Task	Reward	Group	Reward × Group
Deck choice Effort	$F(2, 248) = 80.4$, $P < .001$	$F(1, 124) = 9.7$, $P = .002$	$F(2, 248) = 3.7$, $P = .03$
Perceptual Effort	$F(3, 381) = 49.1$, $P < .001$	$F(1, 127) = 1.1$, $P = .30$	$F(3, 381) = 1.4$, $P = .25$
Grip Effort	$F(2, 258) = 171.9$, $P < .001$	$F(1, 129) = 3.1$, $P = .08$	$F(2, 258) = 3.7$, $P = .03$
Balloon Effort	$F(4, 512) = 35.4$, $P < .001$	$F(1, 128) = 5.1$, $P = .03$	$F(4, 512) = .40$, $P = .81$
EEfRT reward	$F(2, 262) = 82.4$, $P < .001$	$F(1, 131) = 1.1$, $P = .29$	$F(2, 262) = 2.5$, $P = .08$
EEfRT probability	$F(1, 131) = 51.1$, $P < .001$	$F(1, 131) = 1.7$, $P = .20$	$F(1, 131) = 3.9$, $P = .05$

Note: EEfRT, Effort Expenditure for Rewards Task.

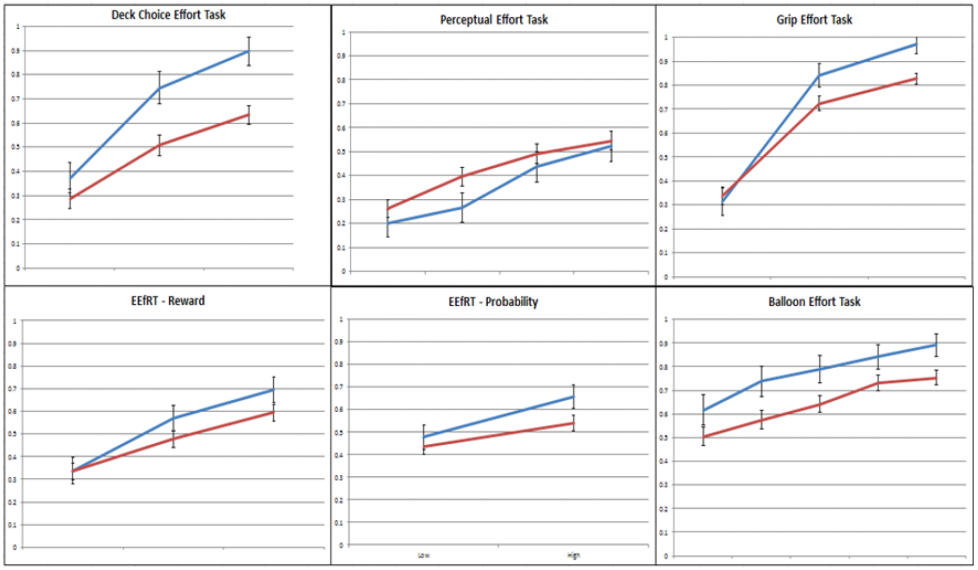

Fig. 1. Patient-control differences in % hard choices by reward level, across the 5 effort-based decision making tasks. Notes: error bars denote standard errors; y-axis denotes % hard choices; x-axis is reward level (from lowest to highest); blue = controls; red = patients.

Table 3. ICCs, Paired Sample *t*-tests, and Test-Retest ES (Cohen's *d*) for Patients at Baseline and 4-wk Follow-up

	Highest Reward-Lowest Reward			Highest Reward Level		
	ICC[a]	*t*-test	*d*[b]	ICC[c]	*t*-test	*d*[d]
Deck Choice Effort	.63**	$t_{82} = -2.12$, $P = .04$	.23	.47**	$t_{82} = -1.39$, $P = .17$	.15
Perceptual Effort	.78**	$t_{78} = -2.47$, $P = .02$	.28	.49**	$t_{78} = 0.53$, $P = .60$	.06
Grip Effort	.63**	$t_{87} = -3.14$, $P < .01$	.34	.59**	$t_{87} = -1.21$, $P = .23$	.13
Balloon Effort	.68**	$t_{86} = -0.09$, $P = .93$	.01	.71**	$t_{86} = 0.70$, $P = .49$	.08
EEfRT reward	.79**	$t_{92} = -1.60$, $P = .11$	.17	.73**	$t_{92} = -0.82$, $P = .41$	.09
EEfRT probability	.53**	$t_{92} = -1.85$, $P = .07$	.19	.69**	$t_{92} = -1.40$, $P = .16$	.15

Note: ES, effect sizes; ICC, Intra-class correlation coefficient.
[a]*Z*-test comparison of ICC's. Significant differences for Highest-Lowest: EEfRT Reward, Perceptual > Deck Choice, Grip, EEfRT Probability.
[b]There were no significant differences between paradigms on this index.
[c]*Z*-test comparison of ICC's. Significant differences for Highest Reward: EEfRT Reward, Balloon, EEfRT Probability > Perceptual, Deck Choice.
[d]There were no significant differences between paradigms on this index.
*$P < .05$; **$P < .01$

"excellent reliability" and values between 0.4 and 0.75 represent "fair to good reliability." As mentioned above, we used 2 dependent variables to calculate reliability coefficients. The high minus low reward dependent variable yielded excellent reliability for the Perceptual Effort Task and EEfRT task. The other 3 tasks were all in the good reliability range according to ICCs. Recalculating the ICCs using the frequency of hard choices at the highest reward level produced ICCs in the fair to good range of reliability for all 5 tasks, with the Balloon Effort Task and EEfRT both having ICC's greater than 0.7. Statistical comparison of the high-low ICCs indicated the EEfRT reward and Perceptual were significantly greater than Deck Choice, Grip Effort, and EEfRT probability. For the highest reward level, the EEfRT reward, probability, and Balloon ICCs were significantly greater than Deck Choice and Grip Effort.

Utility as a Repeated Measure: Mean-Shifts From Time 1 to Time 2

Consideration of tests for repeated assessments in clinical trials includes examination of practice effects that can be problematic if they are large enough to compress the variability. In this case, because there are no "correct" answers, a practice effect refers to a systematic tendency to exert more or less effort when tasks are encountered for the second time. We first analyzed mean changes for the high-low dependent variable for each task. The task with the smallest difference was Balloon Effort, which showed negligible mean shifts from baseline to the 4-week retest (ES = 0.01). The other 4 tasks had small to medium ES, and the Deck Choice Effort, Perceptual Effort, and Grip Effort had significant paired-samples *t* tests

indicating the mean difference between hard choices at the lowest and highest reward levels were slightly higher at retest than at baseline (table 3). When using the percentage of hard choices at the highest reward level all *t* tests were non-significant, the ES were 0.15 or lower. There were no significant differences between tasks in ES.

Tolerability/Duration/Earnings

Participants with schizophrenia considered each measure tolerable with no significant difference noted between tasks (see table 4). Mean ratings ranged from 4.6 to 5.5 across paradigms (scale range: 1 = extremely unpleasant to 7 = extremely pleasant). Mean administration times ranged from 14.1 to 24.9 min. These administration times include instructions, practice trials, and individual calibrations. Average earnings for each task ranged from $4.20 to $9.40; thus task earnings equated roughly to hourly earnings for study participation (ie, approximately $30 for tasks and participation for 2 h).

We compared inflexible responders across tasks and found no participants from either group who picked all hard or all easy across the 5 tasks (table 4). There were 78 unique inflexible responders (ie, inflexible on only 1 task): 54 participants with schizophrenia and 24 controls. To examine characteristics that might separate inflexible responders from other participants, we calculated the number of participants who were inflexible on at least 2 tasks and found 10 participants with schizophrenia who were inflexible on 2 tasks. However, looking closely, we found no consistency in task, and no significant differences from the remaining schizophrenia sample in terms of demographics, cognition, or clinical symptoms.

Discussion

This article presents the first large-scale psychometric evaluation of multiple effort-based decision-making paradigms in schizophrenia. We assessed 5 effort tasks in terms of 4 criteria for potential use in clinical trials: ability to differentiate patients and controls, test-retest reliability, utility as a repeated measure, and tolerability/practicality. The analyses yielded several notable results. First, all of the tasks were considered valid and behaved as expected: there were significant main effects of reward (and probability) level for all tasks, indicating that participants increased their willingness to exert effort for higher rewards. Second, 4 of the 5 tasks significantly separated participants with schizophrenia from controls. Third, the tasks showed relatively good reliability and small mean changes across the 2 time points. Fourth, the tasks were well tolerated, though some tasks had relatively longer administration times. In summary, the results suggest that all tasks were valid measures of effort expenditure, but each task showed a particular pattern of strengths and weaknesses. We will discuss them according to the 4 criteria.

Regarding ability to discriminate individuals with schizophrenia and controls, 2 tasks had main effects of group and 3 tasks had significant interactions. Notably, the group by reward interactions may reflect a problem in matching one's effort to the reward, as opposed to having a more general aversion to exerting effort. Alternately, the interaction could simply reflect the fact that under low reward conditions, control performance was so poor that the schizophrenia and control groups were compressed towards the floor, and that high levels of control performance (eg, high reward and probability conditions) are needed to see group differences. This issue is carefully examined in the animal literature, and a translational discussion of its relevance can be found in the companion article (Young J.W. and Markou A.,[32] this issue). The cognitive Deck Choice Effort Task showed the largest group differences and this is consistent with the previously published task[11] that used a progressive ratio breakpoint design, but not the cognitive effort task that used a similar demand selection task.[12] In the progressive ratio breakpoint study, the individual trials were relatively easy but the effort exertion was related to repetition of trials and increasing cognitive load. The Grip Strength task showed a significant group by reward interaction. It is notable that the 2 previous studies of grip strength used somewhat different methods from the task adapted

Table 4. Tolerability for Patients; Duration, Reward, and Inflexible Responders for Patients and Controls at Baseline

						Inflexible Responders, %			
	Tolerability[a]	Duration (min)		Reward ($)		All Hard		All Easy	
	(1–7)	SCZ	HC	SCZ	HC	HC	SCZ	HC	SCZ
Deck Choice Effort Task	4.6 (2.0)	17 (6)	14 (4)	4 (3)	7 (2)	10	15	9	0
Perceptual Effort Task	5.4 (1.6)	14 (4)	13 (4)	9 (3)	8 (3)	11	5	10	3
Grip Effort Task	5.5 (1.8)	17 (4)	16 (4)	9 (3)	11 (2)	10	8	3	0
Balloon Effort Task	5.5 (1.7)	25 (11)	22 (7)	8 (.8)	8 (.8)	26	33	0	0
EEfRT	5.3 (1.6)	24 (7)	23 (5)	9 (0)	9 (0)	10	8	11	5

Note: SCZ, schizophrenia; HC, healthy controls.
[a]There were no significant differences in tolerability ratings between tasks.

for this study and either did not find group differences or did not examine them.[9,10] The Balloon Effort Task had a main effect of group, but no interaction; the original publication showed a trend for a group by reward interaction and no main effect of group. We found trend-level group × reward and group × probability interactions for the EEfRT task. The 3 previous administrations of this task in schizophrenia also found similar interactions indicating decreased willingness to exert effort among schizophrenia participants at high reward and high probability levels. Importantly, we adapted the task to include individual titration to adjust the "hard" for subjective levels of difficulty and we standardized the number of trials each subject received, unlike previous versions of the task. Given similar results across studies with these procedural differences, it appears that group differences on this paradigm were not solely driven by general motor speed or dexterity differences between participants with schizophrenia and controls. Overall, the paradigms yielded evidence of group differences; however, there are complexities inherent in the categorical sensitivity criterion such as potential subgroups within patients and variability in domains such as cognition that may reduce sensitivity at the group level. The accompanying manuscript (Horan W.P. et al,[16] this issue) closely examines potential patient subgroups in terms of negative symptom severity, cognition, and other external correlates that may be related to effort and motivation.

Test–retest reliability is a critical construct for clinical trials because this property is critical for detecting changes with treatment. It translates directly into statistical power: as test–retest reliability decreases, a larger sample size is required to detect the same amount of change. For test-retest reliability in this study, the tasks performed reasonably well, though the results differed somewhat by task and by the particular dependent measure used. The high-low dependent measure indicated better reliability than the high reward dependent variable for most paradigms. One possible explanation is that the sensitivity to reward is a more stable characteristic than only scores at the highest level. The Deck Choice and Grip Strength had ICC's that exceeded .60 for the high-low variable, putting them at an acceptable level of reliability.

Regarding utility as a repeated measure, we examined practice effects and found the 2 dependent variables again yielded different results across tasks. For this criterion, the high-low variable had somewhat larger ES than the high reward level dependent variable differences between Time 1 and Time 2, although there were no significant differences between tasks within each criterion. Another relevant consideration for suitability for clinical trials is the number of inflexible responders. Participants who choose all hard are problematic for clinical trials because there is no possibility of improvement over time. The frequency of inflexible responders who picked hard trials across all reward levels was substantial for some tasks, particularly Balloon Effort. Importantly, we did not observe an increase in inflexible responders at the second administration, thus inclusion or exclusion of the extreme high scores did not affect overall group mean shifts.

For tolerability and practicality, we considered tolerability ratings from the participants with schizophrenia, administration time, and actual earnings/payouts. The 5 paradigms had very similar and high tolerability ratings. Administration time variability across tasks could be caused by a variety of sources. Three tasks (ie, Deck Choice, Perceptual Effort, and Grip Strength) had trial duration controlled, so irrespective of hard or easy choices, the trials lasted only a certain number of seconds. The Balloon Effort and EEfRT, on the other hand, allowed unlimited time and only controlled for the total number of trials. These tasks were considerably longer, and duration was dependent on choice difficulty and response speed. In addition, all tasks included practice trials, but only the Perceptual Effort and Deck Choice Effort required participants to repeat practice rounds until acceptable accuracy levels were achieved. The individual calibration phases of the Perceptual Effort, Grip Strength, and EEfRT tasks also added time to overall administration duration. For the tasks in which effort choice impacts total duration of the task, it should be considered that temporal discounting represents a potential confound and variations in task length could affect effort exertion.

Regarding the reward payouts, the tasks differed in reward structure. For Deck Choice, Perceptual Effort, and Grip Effort, participants received the exact amount of money they earned on all of the trials combined (ie, for each correctly completed trial). For the Balloon Effort Task and EEfRT, participants were instructed prior to the task that earnings would be based on a set of randomly selected trials after completion of the task. In actuality these 2 tasks were programmed so all participants received a prefixed, roughly equivalent, amount of money. The implications of these differences for repeated administrations is unknown, but future studies may want to query participants as to whether recall of earnings (expected vs actual) from previous administrations affects subsequent administrations. Additionally, participants received hourly compensation for participating in the study. Individual differences in monetary valuation can also impact effort-reward processing and this issue is closely examined in the accompanying manuscript (Horan W.P. et al,[16] this issue).

In summary, our findings highlight some of the complexities of adapting effort-based paradigms for repeated measures in clinical trials. Several aspects of the results are promising. In particular, the tasks appear to be valid indicators of reward motivation. Of note, the sample is representative of an urban, older, chronic, stable outpatient sample with a wide range of negative symptoms and

it is unclear whether the results are fully generalizable to early illness patients, less stable patients, or patients recruited for a negative symptom clinical trial. In terms of the criteria to evaluate paradigms for use in clinical trials, this article examined the first 4 of the 5 criteria and found specific strengths and weaknesses among the tasks. Future comparative studies could use receiver operating characteristic curves to compare the sensitivity and specificity of the different tasks. In terms of detecting group differences, reliability, and utility as a repeated measure, EEfRT is the most promising paradigm. If practice effects are less of a concern, the Grip Effort Task performed very well in terms of group differences, tolerability, and duration. Based on the psychometrics evaluated in this article, all 5 paradigms have potential and several will likely be acceptable for clinical trials with some modifications and refinements. The fifth criterion, external validity and relation to relevant correlates such as functioning and clinician-rated measures of negative symptoms, is the topic of the companion article (Horan W.P. et al,[16] this issue).

Supplementary Material

Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Funding

This project was an investigator-initiated research collaboration funded by Amgen, Inc.

Acknowledgments

Dr Barch reports she has been a paid consultant for Roche, Amgen, Pfizer and Takeda, and has received research funds from Pfizer and Lurasidone. Dr Buchanan reports he is a DSMB member for Pfizer, and he has been a consultant to AbbVie, Amgen, EnVivo, Roche, and Takeda. Dr Dunayevich is a full-time employee of Amgen. Dr Gold reports he receives royalty payments from the BACS and he has been a consultant to Amgen, Hoffman LaRoche, Pfizer, Merck, Astra Zenaca, Solvay, and Glaxo Smith Kline. Dr Marder reports having been a consultant to Abbott, Roche, Genentech, Otsuka, Bristol Meyers Squibb, Pfizer, Lundbeck, and BoehringerIngelheim, and he has received research funds from Amgen, Psychogenics, Sunovion. Dr Treadway reports having been consultant to the Boston Consultant Group. Dr Young has received consulting compensation from Amgen and Arena Pharmaceuticals as well as research grant support from Cerca, Omeros, Lundbeck Ltd, NIMH, and the U.S. Veteran's Administration VISN 22 Mental Illness Research, Education, and Clinical Center. Dr Green reports having been a consultant to AbbVie, DSP, Forum, and Takeda, he is a member of the scientific board for Mnemosyne, and he received research funds from Amgen and Forum. All other authors report no financial disclosures.

References

1. Markou A, Salamone JD, Bussey TJ, et al. Measuring reinforcement learning and motivation constructs in experimental animals: relevance to the negative symptoms of schizophrenia. *Neurosci Biobehav Rev*. 2013;37:2149–2165.
2. Salamone JD, Correa M, Farrar A, Mingote SM. Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology (Berl)*. 2007;191:461–482.
3. Salamone JD, Correa M, Farrar AM, Nunes EJ, Collins LE. Role of dopamine–adenosine interactions in the brain circuitry regulating effort-related decision making: insights into pathological aspects of motivation. *Future Neurol*. 2010;5:377–392.
4. Solomon RL. The influence of work on behavior. *Psychol Bull*. 1948;45:1–40.
5. Fervaha G, Graff-Guerrero A, Zakzanis KK, Foussias G, Agid O, Remington G. Incentive motivation deficits in schizophrenia reflect effort computation impairments during cost-benefit decision-making. *J Psychiatr Res*. 2013;47:1590–1596.
6. Gold JM, Strauss GP, Waltz JA, Robinson BM, Brown JK, Frank MJ. Negative symptoms of schizophrenia are associated with abnormal effort-cost computations. *Biol Psychiatry*. 2013;74:130–136.
7. Barch DM, Treadway MT, Schoen N. Effort, anhedonia, and function in schizophrenia: reduced effort allocation predicts amotivation and functional impairment. *J Abnorm Psychol*. 2014;123:387–397.
8. Treadway MT, Peterman JS, Zald DH, Park S. Impaired effort allocation in patients with schizophrenia. *Schizophr Res* 2014;5:00701–00704.
9. Hartmann MN, Hager OM, Reimann AV, et al. Apathy but not diminished expression in schizophrenia is associated with discounting of monetary rewards by physical effort. *Schizophr Bull*. 2015;41:503–512.
10. Docx L, de la Asuncion J, Sabbe B, et al. Effort discounting and its association with negative symptoms in schizophrenia. *Cogn Neuropsychiatry*. 2015;20:172–185.
11. Wolf DH, Satterthwaite TD, Kantrowitz JJ, et al. Amotivation in schizophrenia: integrated assessment with behavioral, clinical, and imaging measures. *Schizophr Bull* 2014;40:1328–1337.
12. Gold JM, Kool W, Botvinick MM, Hubzin L, August S, Waltz JA. Cognitive effort avoidance and detection in people with schizophrenia. *Cogn Affect Behav Neurosci*. 2015;15:145–154.
13. Green MF, Horan WP, Barch DM, Gold JM. Effort-based decision making: A novel approach for assessing motivation in schizophrenia [published online ahead of print June 18, 2015]. *Schizophr Bull*. doi: 10.1093/schbul/sbv071.
14. Carter CS, Barch DM, Buchanan RW, et al. Identifying cognitive mechanisms targeted for treatment development in schizophrenia: an overview of the first meeting of the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia Initiative. *Biol Psychiatry*. 2008;64:4–10.
15. Green MF, Nuechterlein KH, Gold JM, et al. Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. *Biol Psychiatry*. 2004;56:301–307.

16. Horan WP, Reddy LF, Barch DM, et al. Effort-based decision making paradigms for clinical trials in schizophrenia: Part 2 - External validity and correlates. *Schizophr Bull*.
17. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders - Patient Edition*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1997.
18. Kool W, McGuire JT, Rosen ZB, Botvinick MM. Decision making and the avoidance of cognitive demand. *J Exp Psychol Gen*. 2010;139:665–682.
19. McGuire JT, Botvinick MM. Prefrontal cortex, cognitive control, and the registration of decision costs. *Proc Natl Acad of Sci U S A*. 2010;107:7922–7926.
20. Cocker PJ, Hosking JG, Benoit J, Winstanley CA. Sensitivity to cognitive effort mediates psychostimulant effects on a novel rodent cost/benefit decision-making task. *Neuropsychopharmacology*. 2012;37:1825–1837.
21. Robbins TW. The 5-choice serial reaction time task: behavioural pharmacology and functional neurochemistry. *Psychopharmacology (Berl)*. 2002;163:362–380.
22. Hairston WD, Maldjian JA. An adaptive staircase procedure for the E-Prime programming environment. *Comput Methods Programs Biomed*. 2009;93:104–108.
23. Cléry-Melin ML, Schmidt L, Lafargue G, Baup N, Fossati P, Pessiglione M. Why don't you try harder? An investigation of effort production in major depression. *PLoS One*. 2011;6:e23178.
24. Kurniawan IT, Seymour B, Talmi D, Yoshida W, Chater N, Dolan RJ. Choosing to make an effort: the role of striatum in signaling physical effort of a chosen action. *J Neurophysiol*. 2010;104:313–321.
25. Treadway MT, Bossaller NA, Shelton RC, Zald DH. Effort-based decision-making in major depressive disorder: a translational model of motivational anhedonia. *J Abnorm Psychol*. 2012;121:553–558.
26. Kay SR, Opler LA, Lindenmayer JP. The Positive and Negative Syndrome Scale (PANSS): rationale and standardization. *Br J Psychiatry*. 1989;155(suppl 7):59–65.
27. Horan WP, Kring AM, Gur RE, Reise SP, Blanchard JJ. Development and psychometric validation of the Clinical Assessment Interview for Negative Symptoms (CAINS). *Schizophr Res*. 2011;132:140–145.
28. Axelrod BN, Goldman RS, Alphs LD. Validation of the 16-item Negative Symptom Assessment. *J Psychiatr Res*. 1993;27:253–258.
29. Marder SR, Davis JM, Chouinard G. The effects of risperidone on the five dimensions of schizophrenia derived by factor analyses: Combined results of the North American trials. *J Clin Psychiatry*. 1997;58:538–546.
30. Nuechterlein KH, Green MF. *MATRICS Consensus Cogntive Battery*. Los Angeles, CA: MATRICS Assessment, Inc.; 2006.
31. Bratti IM, Kane JM, Marder SR. Chronic restlessness with antipsychotics. *Am J Psychiatry*. 2007;164:1648–1654.
32. Young JW, Markou A. Translational rodent paradigms to investigate neuromechanisms underlying behaviors relevant to amotivation and altered reward processing in schizophrenia. *Schizophr Bull*.