Published in final edited form as:

Psychopharmacology (Berl). 2013 August ; 228(4): 611–622. doi:10.1007/s00213-013-3070-4.

A Critical Examination of Best Dose Analysis for Determining Cognitive-Enhancing Potential of Drugs: Studies with Rhesus Monkeys and Computer Simulations

Paul L. Soto¹, Jesse Dallery², Nancy A. Ator¹, and Brian R. Katz¹

¹Division of Behavioral Biology, Department of Psychiatry and Behavioral Sciences Johns Hopkins School of Medicine

²Department of Psychology, University of Florida

Abstract

Rationale—Best dose analysis involves identifying the dose associated with the greatest improvement in performance for each subject and comparing performances associated with these individually-determined best doses to control performances.

Objectives—The current experiments were conducted to examine whether significant best dose effects might result from the selective analysis of data rather than an actual drug effect.

Methods—Experiment 1 examined effects of nicotine and methylphenidate on delayed matching-to-sample (DMTS) and self-ordered spatial search (SOSS) performances in rhesus monkeys (DMTS: $n=7$; SOSS: $n=6$) to determine the validity and reliability of best dose effects. Experiment 2 used Monte Carlo computer simulations to estimate the likelihood of obtaining a significant outcome when the best dose method was applied to randomly generated data sets for which no difference existed.

Results—Significant effects were obtained when the best dose analysis was applied to performances from non-drug sessions, and best dose performances were not significantly different from the best non-drug performances. The doses identified as best doses from two nicotine dose-response curve determinations were unrelated, and the improvement associated with the best dose observed during the first dose-response curve determination was not reliable when the dose was administered repeatedly. Finally, there was a high likelihood of obtaining a statistically significant difference when no real difference existed.

Conclusions—Best dose analysis for identification of potential therapeutic agents should be replaced by single subject designs.

Keywords

best dose analysis; cognitive enhancer; nicotine; methylphenidate; rhesus monkey; CANTAB; working memory; delayed matching-to-sample; self-ordered spatial search

Corresponding Author: Paul L. Soto, Ph.D., Behavioral Biology Research Center, 5510 Nathan Shock Drive, Suite 3000, Baltimore, MD 21224-6823, Phone: 443.287.4530, Fax: 410.550.2780, psoto@jhmi.edu.

Conflicts of Interest:

We have no conflicts of interest to declare.

Introduction

The effects of drug administration on behavior can vary substantially across individuals. Individual variability in response to drug administration has been recognized for some time in pharmacology and can be traced back at least as far as the concept of initial sensitivity or “initial tolerance” (Collett 1988; Kalant et al. 1971), which is the observation that the amount of drug necessary to produce a specified effect on first exposure can vary across individuals. Inconsistency in drug response is not limited to initial exposure and can occur with repeated administration of a drug (e.g., Kaiyala et al. 2001; Perkins et al. 2008; Smolen et al. 1994). When there are individual differences in drug effects, averaging data may obscure those drug effects. For example, averaging effects that are opposite in direction across individuals may show a null effect when the drug actually is effective in changing behavior for a subset of individuals.

In the area of cognition-enhancing effects of drugs, a common data-analysis approach to address the issue of individual differences is the “best dose analysis.” As the name implies, the goal of the analysis is to identify the most effective, or best, dose for each individual. Some researchers have argued that, in the study of cognition-enhancing drug effects, a best dose approach is necessitated by inter-subject variability in degree of drug efficacy, a narrow range of doses at which therapeutic effects occur, and inter-subject variability in the doses at which therapeutic effects are obtained (Buccafusco and Terry 2004; 2000). Best dose analyses have been used to identify putative cognition-enhancing effects of cholinergic, adrenergic, dopaminergic, and serotonergic drugs (Arnsten and Contant 1992; Arnsten and Goldman-Rakic 1990; Bain et al. 2003; Buccafusco and Jackson 1991; Buccafusco et al. 2003; Buccafusco et al. 1995; Buccafusco et al. 1996; Buccafusco and Terry 2004; Elrod et al. 1988; Franowicz and Arnsten 1998; Gamo et al. 2010; Gould et al. 2013; Katner et al. 2004; Prendergast et al. 1998; Prendergast et al. 1997; Terry et al. 2005).

A best dose analysis involves determining the effects of a range of drug doses on performance of a cognitive task such as a delayed-matching-to-sample (DMTS) procedure. The dose associated with the greatest accuracy is identified as the best dose on a subject-by-subject basis. For example, three subjects may show greatest accuracy at low, medium, and high drug doses, respectively. A statistically significant difference between the individually-selected best dose performances and vehicle performances would support the conclusion of a drug’s cognitive-enhancing effects.

Of the studies employing a best dose analysis for the assessment of potential cognition-enhancing effects, some also have reported the outcomes of a statistical analysis involving all doses tested (Bain et al. 2003; Buccafusco et al. 1999; Buccafusco et al. 1995; Buccafusco et al. 1996; Buccafusco and Terry 2004; Franowicz and Arnsten 1998; Katner et al. 2004; Terry et al. 1993). In some studies, drug effects were significant according to both a best dose analysis and an analysis including all doses tested (Buccafusco et al. 1999; Buccafusco et al. 1995; Buccafusco et al. 1996; Franowicz and Arnsten 1998), and in other studies, drug effects were significant only according to a best dose analysis (Bain et al. 2003; Buccafusco and Terry 2004; Katner et al. 2004), suggesting that best dose effects may be an artifact of the selective nature of the analysis rather than an actual drug effect.

This paper attempts to critically evaluate the best dose analysis in three ways. In the first experiment, we evaluated the validity and reliability of the best dose analysis using data obtained in rhesus monkeys. In the second experiment, we used Monte Carlo computer simulations to determine the likelihood of obtaining significant best dose effects when a best dose analysis was applied to two sets of data for which no actual difference exists (i.e., all numbers in the data sets were selected randomly from the same distribution).

Experiment 1: Tests of Validity and Reliability in Rhesus Monkeys

The goal of Experiment 1 was to determine the validity of the best dose method for identifying drug effects and to determine both the reliability of the method for identifying effective doses and the reliability of putative improvements produced by identified best doses. To determine the validity of the analysis, performances from non-drug control sessions were analyzed in a manner equivalent to a best dose analysis to assess whether significant effects would be obtained in the absence of drug. Additionally, best dose performances were compared to best control performances. To determine the reliability of the approach for identifying best doses, multiple dose-response curve assessments were conducted and the best doses that were identified from each assessment compared. To determine the reliability of putative improvements, the best doses then were repeatedly administered to assess whether reliable improvements in performance would be obtained. The rationale of Experiment 1 was that significant best dose effects may be questionable if any of the following conditions are met: (a) significant effects are obtained for sessions in which no drug is administered, (b) significant effects are not obtained when all doses are included, (c) significant effects are not obtained when best dose performances are compared to best control performances, (d) best dose identifications are not reliable across repeated determinations, and (e) identified best doses do not yield reliable improvements in performance.

Subjects

Adult male rhesus monkeys (*Macaca mulatta*, N = 13) were singly housed, each in one section of a four-cage housing unit. All caging units were housed in the same temperature- and humidity-controlled vivarium. Lights were on from 7 am to 9 pm. The monkeys were between 11 and 20 years of age at the start of these studies and had been trained previously on the tasks. Weights averaged ~16 kg. The quantities of primate diet (2050 Teklad Global 20% Protein Primate Diet, Harlan Laboratories) that were fed daily were sufficient to maintain weights over the course of the study. Additionally, monkeys received a piece of fresh fruit or vegetables (e.g., ½ orange, ½ apple, etc.) 5 days a week. Feeding occurred approximately 2 hours after completion of behavioral testing. Reverse-osmosis-treated water was available at all times.

Apparatus

Experimental sessions were conducted in the home cage using custom-built mobile devices described previously (Weed et al. 2008). Briefly, each held a computer for control of experimental events (CANTAB software; Lafayette Industries, Lafayette, IN) and two touch screen monitors (Intellitouch, surface acoustic wave technology, ELO TouchSystems, Menlo Park, CA, USA) that allowed two monkeys to be tested simultaneously. A pellet dispenser (BRS/VLE, Laurel, MD, or Med Associates, Inc. St. Albans, VT) was used for delivery of 190-mg food pellets (BioServ, Frenchtown, NJ).

Procedure

Seven monkeys were trained on a DMTS procedure. Each session consisted of 24 trials with 8 trials at each of three delays (2, 30, or 300 s). The delay on each trial was selected randomly without replacement. Each trial began with presentation, on the center of the screen, of a pseudo-randomly selected sample image from a set of 600 different images (Photo Clip Art 150,000 by Hemera Technologies, Inc.). A touch on the sample image within 30 s (limited hold) turned off the image and initiated the selected delay. After the delay, the original sample image and 2 other unique, randomly selected, images were presented on three corners of the screen. A touch on the image that “matched” the original sample image produced a food pellet, followed by a 5-s period with the screen darkened. If

the monkey did not touch the sample image within the 30-s limited hold, did not touch one of the 3 choice images within 30 s, or if the monkey touched one of the 2 “non-matching” images, the trial ended without pellet delivery, followed by a 10-s period with the screen darkened.

Six monkeys were trained on a self-ordered spatial search (SOSS) procedure. Each session consisted of 54 trials and each trial involved presentation, on the touch screen, of a configuration of a number of small blue boxes within 16 possible screen locations. The number of boxes in the stimulus configuration varied among 2, 3, and 4 boxes (18 trials of each). Each non-repeating touch produced a food pellet. If the monkey made a repeat touch (incorrect) or failed to make a touch within 30-s of trial onset or from the time of the previous touch (omission), the trial ended and a 9-s period followed during which the screen remained blank and touching the screen produced no scheduled consequence, followed by a new trial. If the monkey touched all the boxes without repetition, the trial ended, was defined as correct, and was followed by 5-s period with the screen darkened before the next trial.

Drug Testing

Nicotine bitartrate (Sigma-Aldrich, St. Louis, MO) and methylphenidate hydrochloride (Mallinkrodt Pharmaceuticals, Inc.) were dissolved in 0.9% sodium chloride (saline). Drugs were administered intramuscularly in the thigh at a volume of 0.2 – 0.8 ml, depending on drug solubility, 15 min prior to the session. Drug solutions were typically prepared fresh each testing day, but occasionally were prepared from stock solutions that were no more than 1 week old.

Experimental sessions were conducted 5 days a week, Monday - Friday, starting at approximately 10 or 11 am. Drug test sessions usually occurred on Tuesday and Friday, if subjects completed at least 85% of the trials at each delay (DMTS) or number of boxes (SOSS) in the preceding session. Baseline (no injection) sessions occurred on Mondays and Wednesdays (Wednesday baseline sessions were excluded from analysis to avoid possible carryover effects from Tuesday’s drug administration). Vehicle sessions usually occurred on Thursday. During nicotine and methylphenidate dose-response curve determinations, doses were studied in a pseudo-random order with the restriction that the highest two doses were tested after the lower doses had been tested. In the SOSS monkeys, the nicotine dose-response curve was re-determined twice (~1 year after the original determination), with an approximately 3-week period between those two re-determinations. During those 3 weeks, the best dose identified from the first of those re-determinations was administered six times (on Tuesdays and Fridays).

Data Analysis

Overall session accuracy was calculated for each monkey for each session by dividing the total number of correct trials by the total number of trials completed. Also, the percentage of correct trials for each trial type (trials with delays of 2, 30, and 300 s in the DMTS procedure or trials with 2, 3, and 4 boxes in the SOSS procedure) was calculated for each monkey for each session. Percentages of trials correct were converted to proportions and arcsine square root transformed to increase normality for statistical analysis (McDonald 2009). Several analyses were conducted to assess the validity of the best dose approach. In the first analysis, a two-way repeated measures ANOVA using factors of dose and task parameter (delay for DMTS; number of boxes for SOSS) was conducted using all the doses tested. In the second analysis (Best Dose Analysis), the best dose session for each monkey was identified as the session during which overall accuracy (collapsed across task parameter value) was the highest compared to the other doses tested. Percentage correct values from

the best dose session were compared to percentage correct values from vehicle sessions using a two-way (factor of delay or number of boxes and a factor of treatment including best dose and vehicle) repeated measures ANOVA. In the third (Best Vehicle Analysis) and fourth (Best Baseline Analysis) analyses, the best vehicle and best baseline session for each monkey was identified as the session during which overall accuracy was highest compared to other sessions of the same type and the percentage correct values from the best vehicle and baseline session were compared, via two-factor ANOVA, to the percentage correct values obtained during the remaining vehicle and baseline sessions, respectively. In the fifth analysis (Best Session Analysis), the accuracy values associated with the best dose, best vehicle, and best baseline sessions were compared via two-factor ANOVA.

For each of the two nicotine dose-response curve re-determinations conducted in the SOSS monkeys, a best dose analysis was conducted. A Pearson correlation coefficient was calculated to identify whether the correlation between best doses identified from the two dose-response curve re-determinations was statistically significant. A Spearman rank order correlation coefficient was calculated on rankings assigned to the monkeys based on their identified best doses to determine whether a monotonic dependence existed between the best doses identified in the two re-determinations. Finally, for the repeated best dose administrations, performances from best dose sessions were averaged together for each subject and compared by repeated-measures ANOVA to baseline and vehicle session averages that occurred during the same period and to baseline sessions from the approximately 3-week period without any drug administration that occurred prior to the first nicotine dose-response curve re-determination.

In all analyses, post-hoc comparisons using the Holm-Sidak method at a family-wise $\alpha = 0.05$ were conducted to compare accuracy within each trial configuration (e.g., to compare best dose to vehicle at each delay in the DMTS best dose analysis).

Results

Mean lengths (and SDs) of the DMTS sessions during the nicotine and methylphenidate dose-response curve determination were 48.3 ± 3.7 and 45.8 ± 8.4 min, respectively. Nicotine and methylphenidate produced no statistically significant change in DMTS accuracy at any of the three delays (Fig. 1). In contrast to the lack of effects obtained when all doses were used in the statistical analysis, DMTS performances associated with best doses of nicotine and methylphenidate were significantly better than DMTS performances following vehicle administration (Fig. 1; Table 1). Similarly, DMTS performances associated with the best vehicle and best baseline sessions were significantly better than performances associated with the remaining vehicle and baseline sessions, respectively, for both nicotine and methylphenidate (Fig. 1; Table 1). Finally, DMTS best dose performance was not significantly better than best vehicle or best baseline performance for either nicotine or methylphenidate (Fig. 1; Table 1).

Mean lengths of SOSS sessions during the nicotine and methylphenidate dose-response curve determination were 15.4 ± 0.9 and 18.1 ± 7.7 min, respectively. In the SOSS procedure, nicotine and methylphenidate again failed to produce a statistically significant change in accuracy when all doses tested were included in the analysis (Fig. 2). SOSS performances associated with nicotine and methylphenidate best doses were significantly better than vehicle performances (Fig. 2; Table 1). As with the DMTS procedure, SOSS performances associated with the best vehicle and best baseline sessions of nicotine and methylphenidate were significantly better than the remaining vehicle, or baseline, performances, respectively (Fig. 2; Table 1). SOSS performances associated with the best doses of nicotine and methylphenidate were not significantly better than best vehicle or best baseline performances (Fig. 2; Table 1).

The best doses of nicotine identified from the first of the two nicotine dose-response curve re-determinations in the SOSS group ranged from 0.001 to 0.056 mg/kg (Table 2). There was a statistically significant difference between best dose and vehicle accuracy values ($F_{1,5} = 12.032$, $p=0.018$). The best doses of nicotine identified from the second re-determination ranged from 0.003 to 0.056 mg/kg. For monkey RQ6101, there was a tie for the best dose in the second dose-response curve determination (0.003 and 0.056 mg/kg, both associated with 96.3% correct overall). There was a statistically significant difference between best dose and vehicle accuracy values when the two best dose performances of RQ6101 were averaged ($F_{1,5}=68.307$, $p<0.001$) and when they were used separately in the ANOVA ($F_{1,5} = 48.004$, $p<0.001$ using performance after 0.003 mg/kg; $F_{1,5} = 71.255$, $p<0.001$ using performance after 0.056 mg/kg). For no subject was the best dose of nicotine identified in the two dose-response curve determinations equal (Fig. 3; Table 2). Further, the correlation between the best doses of nicotine from the two dose-response curve determinations was not significant when 0.003 mg/kg ($r = -0.32$, $p = 0.55$) or 0.056 mg/kg ($r = -0.56$, $p = 0.24$) was used for RQ6101. Similarly, the rank order correlation was not significant when 0.003 mg/kg ($r = -0.13$, $p=0.81$) or 0.056 mg/kg ($r=-0.56$, $p=0.25$) was used for RQ6101.

Finally, when the best dose of nicotine from the first nicotine dose-response curve re-determination was administered repeatedly (Fig. 4; data from 3 of 6 subjects shown), there was no statistically significant difference in accuracy following those best dose administrations compared to accuracy values obtained from vehicle and baseline performances during the same time period. Also, there were no statistically significant differences between individual subject average accuracy values during either of the two nicotine dose-response curve re-determinations and the 3-week non-drug period that preceded the first re-determination or during the repeated best dose administrations and the 3-week non-drug period.

Discussion

As noted above, the rationale of Experiment 1 was that significant best dose effects may be questionable if (a) significant effects are obtained for sessions in which no drug is administered, (b) significant effects are not obtained when all data are analyzed, (c) significant effects are not obtained when best dose performances are compared to best non-drug performances, (d) best dose identifications are not reliable, and (e) identified best doses do not produce reliable improvements. In Experiment 1, all five conditions were met. Collectively, these findings suggest that best dose effects can result from the selective nature of the analysis rather than from an actual effect of the drug.

The possibility that repeated administration of the best dose of nicotine produced either tolerance or sensitization to nicotine's effects on SOSS performance is unlikely because there were no clear changes in the nicotine dose-response curve in any direction. The absence of a significant correlation between the best doses identified in the first and second re-determinations of the nicotine dose-response curve further argues against the possibility of any systematic change in the effects of nicotine resulting from repeated administration of the initially-determined best dose of nicotine. Finally, the possibility that long-lasting effects of nicotine (Buccafusco et al. 2005), could have confounded comparisons of non-drug and drug administration sessions during the reliability assessment of nicotine's best dose effects appears unlikely due to the absence of any statistically significant differences between performances during nicotine administration periods and the non-drug sessions that preceded nicotine testing. The lack of reliable performance improvement calls into question the clinical significance of a significant best dose effect because a therapeutic effect must, of course, be reliable to be clinically important.

Experiment 2: Likelihood of Obtaining a Significant Effect When None Exists

Experiment 1 suggested that statistically significant effects can be obtained when a best dose analysis is applied to non-drug treatment performances and that neither best dose determinations nor best dose effects are reliable. Experiment 2 was conducted to evaluate the likelihood of obtaining statistically significant effects when a selective data analysis approach equivalent to a best dose analysis is applied to simulated data sets for which no actual difference exists. Simulated data sets were generated by randomly selecting numbers from a normal distribution of specified mean and standard deviation and classifying those numbers as either “vehicle” or “drug” and assigning them to simulated subjects. From those simulated data sets, the average vehicle values were compared statistically to the maximum (i.e., “best”) drug value (determined separately for each simulated subject).

Methods

Monte Carlo simulations were conducted using Visual Basic for Applications within Microsoft Excel 2007. Simulated data sets were generated to evaluate the likelihood of obtaining a statistically significant F-value according to a best dose analysis when all numbers in each data set were generated from the same underlying population. Each simulated data set was generated by randomly selecting accuracy values (0 – 100) from a normal distribution of specified mean and standard deviation. Within each data set, selected accuracy values were assigned to simulated subjects with half the values classified as “Drug” and half the values classified as “Vehicle.” The parameters of the simulated data sets were varied over the following values: Mean of the population = 80, 70, 60; standard deviation of the population = 0.5, 2.5, 5, and 10; number of simulated subjects = 4, 6, 8, 10, and 12; and number of observations per subject (per classification of “Drug” and “Vehicle”) = 4, 6, 8, 10, and 12. For each combination (3 means $\times$ 4 standard deviations $\times$ 5 subject numbers $\times$ 5 observation values = 300 combinations) of population mean, population standard deviation, number of subjects, and number of observations per subject, 100 simulated data sets were generated.

Data Analysis

Each simulated data set was analyzed according to a one-way ANOVA with factor of treatment (Best Dose vs. Vehicle) in which the maximum of the drug values (analogous to selecting the performance associated with the “best” dose) was identified for each simulated subject, and compared to the average vehicle value for each subject. As in Experiment 1, the maximum and average values were converted to proportions and arcsine square root transformed to increase normality for statistical analysis (McDonald 2009).

Results

The number of statistically significant F-values obtained according to a one-way ANOVA comparing the maximum “Drug” values to the average “Vehicle” values (procedurally identical to a best dose analysis) from the Monte Carlo simulated data sets are depicted in Figs. 5–7. When the mean of the population was 80, the number of statistically significant F-values increased as the number of subjects and the number of observations per subject increased (Fig. 5) from 27 – 36 (out of 100 simulated data sets) when the number of subjects and observations were both equal to 4 (Fig. 5) to 100 when the number of subjects and observations were both equal to 12. Similar functions relating the number of statistically significant outcomes were obtained when the population mean was 70 (Fig. 6) and 60 (Fig. 7). The number of statistically significant outcomes did not appear to depend on the standard deviation of the population (Figs. 5 – 7).

Discussion

The Monte Carlo simulations conducted revealed a high likelihood, far exceeding the typical alpha level of 0.05, of obtaining a significant statistical outcome even when no real difference existed. Further, the results of these Monte Carlo simulations suggest that even when there is very low variability in the percentage correct values, the likelihood of obtaining a statistically significant outcome is unacceptably high. Thus, it cannot be said that greater experimental control would solve the dilemma revealed by these results. In summary, the Monte Carlo simulation results underscore the experimental results from Experiment 1 and demonstrate that the selective nature of the best dose approach is subject to a high rate of false positives.

General Discussion

The current experiments demonstrated that: (1) a best dose analysis can yield statistically significant outcomes when applied to non-drug control sessions; (2) statistically significant best dose performances may not differ from best control session performances; (3) the identification of the doses that produce improvement may not be reliable; (4) the improvements identified by the best dose analysis may not be reliable; and (5) the best dose analysis is subject to a very high rate of false positive outcomes. It appears there is a high probability that best dose effects result from the selective nature of the analysis rather than from an actual drug effect. Although it remains possible that some significant best dose effects reported in the literature represent real drug effects, the question becomes one of distinguishing the real from the questionable outcomes. Thus, the best dose analysis does not appear to provide an acceptable approach to the analysis of drug effects where there are substantial individual differences in drug sensitivity.

In contrast to the procedures employed in Experiment 1, some previous studies utilizing best dose analysis have employed individualized procedural parameters (i.e., delays in DMTS) to ensure equivalent performances across monkeys (e.g., Buccafusco and Jackson 1991; Buccafusco et al. 1995; Buccafusco and Terry 2004). Such individualization of task parameters has the benefit of reducing inter-individual variability. This difference does not alter the conclusions of the current study for two reasons. First, the Monte Carlo simulations demonstrate that even with low variability, the best dose analysis yields a high rate of false positive outcomes. Second, Kangas and Branch (2012) demonstrated that questionable statistically significant enhancements were obtained with a best dose analysis of nicotine's effects on titrating DMTS performance despite the absence of any evidence of improvements in individual subjects, suggesting that individualizing task parameters does not alleviate problems posed by the best dose analysis.

The finding of Experiment 2 that statistically significant best dose effects did not replicate when best doses were repeated underscores an often overlooked fact that statistical significance does not imply replicability, a misconception documented by multiple authors (Branch 1999; Carver 1978; 1993; Falk and Greenbaum 1995; Sohn 1998). The disconnect between statistical significance and replication highlights the need for replication of drug effects to increase confidence in those effects. Such replications should be conducted within and across subjects as well as across studies. Furthermore, the fact that statistical significance does not imply replicability emphasizes the importance of failures to replicate in establishing the reliability and generality, or lack thereof, of reported drug effects.

The high likelihood of incorrectly concluding a drug effect that is inherent in a best dose analysis has important implications. Current success rates for central nervous system drugs entering phase I clinical development has been reported at 8% (Kola and Landis 2004) and difficulties translating preclinical findings into clinically effective treatments have been

noted (Hackam 2007; Hackam and Redelmeier 2006; Perel et al. 2007). There are many possible reasons for the discrepancies in preclinical and clinical outcomes including species differences in pharmacodynamics and pharmacokinetics (Geerts 2009). The results of the current experiments suggest that the high rate of false positives inherent in a best dose analysis also may hamper drug discovery efforts. Specifically, the results of these studies suggest that, to the degree that efforts to identify cognition-enhancing drugs rely on best dose approaches, a great deal of time, money, and effort could be wasted pursuing drugs that are without real beneficial effects.

One potential alternative to the best dose approach, particularly for study of drugs that increase cognitive performance would be to employ an approach that focuses on demonstration and replication of effects in individual subjects (Kazdin 2011; Sidman 1960). Single-subject research methods offer a range of strategies that are well suited to the problem of individual differences in drug effects (Sidman 1960). Notably, single subject or small n designs were employed in the original studies on the behavioral effects of drugs that are the foundation of the field of behavioral pharmacology (Brady 1956; Dews 1955a; b) and their value in biomedical research has been noted (Dallery et al. 2013; Madsen and Bytzer 2002; Morgan and Morgan 2001). The focus is on determining variables that reliably affect the behavior of the individual and, thus, effects must be replicated for each individual. Generality is determined through replication across subjects. Although Experiment 2 suggested that best dose effects of nicotine on SOSS performance were unreliable, this does not preclude the possibility that replicable best dose effects could be obtained in other experiments (e.g., Arnsten and Goldman-Rakic 1990; Bontempi et al. 2001; Terry et al. 1998) or that all studies on cognitive-enhancing drugs are flawed. Ideally, studies should also incorporate parametric analysis, even if it must be within a narrow range of drug doses. Graded dose-response curves can increase confidence in a drug's effects. If similar orderly dose-response relations are found for each individual, even if drug potency differs across individuals, we can be more confident in the reliability and generality of the obtained effects (Branch and Pennypacker 2013). Although a single-subject approach requires more time and effort, it is likely better suited to demonstrating drug effects in the face of inter-subject variability than a best dose approach.

Acknowledgments

We would like to thank Stacey Perry, Raymond Smith, and Virginia Bogdan for their expert technical assistance in conducting these studies. We thank Michael R. Weed for helpful discussions of the ideas set forth in this manuscript. This work was supported by NIA grant AG027798.

References

- Arnsten AF, Contant TA. Alpha-2 adrenergic agonists decrease distractibility in aged monkeys performing the delayed response task. *Psychopharmacology (Berl)*. 1992; 108:159–69. [PubMed: 1357704]
- Arnsten AF, Goldman-Rakic PS. Analysis of alpha-2 adrenergic agonist effects on the delayed nonmatch-to-sample performance of aged rhesus monkeys. *Neurobiol Aging*. 1990; 11:583–90. [PubMed: 1980719]
- Bain JN, Prendergast MA, Terry AV Jr, Arneric SP, Smith MA, Buccafusco JJ. Enhanced attention in rhesus monkeys as a common factor for the cognitive effects of drugs with abuse potential. *Psychopharmacology (Berl)*. 2003; 169:150–60. [PubMed: 12768267]
- Bontempi B, Whelan KT, Risbrough VB, Rao TS, Buccafusco JJ, Lloyd GK, Menzaghi F. SIB-1553A, (+/-)-4-[[2-(1-methyl-2-pyrrolidinyl)ethyl]thio]phenol hydrochloride, a subtype-selective ligand for nicotinic acetylcholine receptors with putative cognitive-enhancing properties: effects on working and reference memory performances in aged rodents and nonhuman primates. *J Pharmacol Exp Ther*. 2001; 299:297–306. [PubMed: 11561092]

- Brady JV. Assessment of drug effects on emotional behavior. *Science*. 1956; 123:1033–4. [PubMed: 13324142]
- Branch MN. Statistical inference in behavior analysis: Some things significance testing does and does not do. *Behav Anal*. 1999; 22:87–92. [PubMed: 22478324]
- Branch, MN.; Pennypacker, HS. Generality and generalization of research findings. In: Madden, GL.; Dube, WV.; Hackenberg, TD.; Hanley, GP.; Lattal, KA., editors. *APA Handbook of Behavior Analysis*. American Psychological Association; Washington, D.C: 2013. p. 151-175.
- Buccafusco JJ, Jackson WJ. Beneficial effects of nicotine administered prior to a delayed matching-to-sample task in young and aged monkeys. *Neurobiol Aging*. 1991; 12:233–8. [PubMed: 1876228]
- Buccafusco JJ, Jackson WJ, Jonnala RR, Terry AV Jr. Differential improvement in memory-related task performance with nicotine by aged male and female rhesus monkeys. *Behav Pharmacol*. 1999; 10:681–90. [PubMed: 10780510]
- Buccafusco JJ, Jackson WJ, Stone JD, Terry AV. Sex dimorphisms in the cognitive-enhancing action of the Alzheimer's drug donepezil in aged Rhesus monkeys. *Neuropharmacology*. 2003; 44:381–9. [PubMed: 12604096]
- Buccafusco JJ, Jackson WJ, Terry AV Jr, Marsh KC, Decker MW, Arneric SP. Improvement in performance of a delayed matching-to-sample task by monkeys following ABT-418: a novel cholinergic channel activator for memory enhancement. *Psychopharmacology (Berl)*. 1995; 120:256–66. [PubMed: 8524972]
- Buccafusco JJ, Letchworth SR, Bencherif M, Lippiello PM. Long-lasting cognitive improvement with nicotinic receptor agonists: mechanisms of pharmacokinetic-pharmacodynamic discordance. *Trends Pharmacol Sci*. 2005; 26:352–60. [PubMed: 15946748]
- Buccafusco JJ, Prendergast MA, Terry AV, Jackson WJ. Cognitive effects of nicotinic cholinergic receptor agonists in nonhuman primates. *Drug Dev Res*. 1996; 38:196–203.
- Buccafusco JJ, Terry AV. Donepezil-induced improvement in delayed matching accuracy by young and old rhesus monkeys. *J Mol Neurosci*. 2004; 24:85–91. [PubMed: 15314255]
- Buccafusco JJ, Terry AV Jr. Multiple central nervous system targets for eliciting beneficial effects on memory and cognition. *J Pharmacol Exp Ther*. 2000; 295:438–46. [PubMed: 11046074]
- Carver RP. The case against statistical significance testing. *Harvard Educational Review*. 1978; 48:378–399.
- Carver RP. The case against statistical significance testing, revisited. *Journal of Experimental Education*. 1993; 61:287–292.
- Collett BJ. Opioid tolerance: the clinical perspective. *British Journal of Anaesthesia*. 1988; 81:58–68. [PubMed: 9771273]
- Dallery J, Cassidy RN, Raiff BR. Single-case experimental designs to evaluate novel technology-based health interventions. *J Med Internet Res*. 2013; 15:e22. [PubMed: 23399668]
- Dews PB. Studies on behavior. I. Differential sensitivity to pentobarbital of pecking performance in pigeons depending on the schedule of reward. *J Pharmacol Exp Ther*. 1955a; 113:393–401. [PubMed: 14368508]
- Dews PB. Studies on behavior. II. The effects of pentobarbital, methamphetamine and scopolamine on performances in pigeons involving discriminations. *J Pharmacol Exp Ther*. 1955b; 115:380–9. [PubMed: 13272190]
- Elrod K, Buccafusco JJ, Jackson WJ. Nicotine enhances delayed matching-to-sample performance by primates. *Life Sci*. 1988; 43:277–87. [PubMed: 2899825]
- Falk R, Greenbaum CW. Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology*. 1995; 5:75–98.
- Franowicz JS, Arnsten AF. The alpha-2a noradrenergic agonist, guanfacine, improves delayed response performance in young adult rhesus monkeys. *Psychopharmacology (Berl)*. 1998; 136:8–14. [PubMed: 9537677]
- Gamo NJ, Wang M, Arnsten AF. Methylphenidate and atomoxetine enhance prefrontal function through alpha2-adrenergic and dopamine D1 receptors. *J Am Acad Child Adolesc Psychiatry*. 2010; 49:1011–23. [PubMed: 20855046]
- Geerts H. Of mice and men: bridging the translational disconnect in CNS drug discovery. *CNS Drugs*. 2009; 23:915–26. [PubMed: 19845413]

- Gould RW, Garg PK, Garg S, Nader MA. Effects of nicotinic acetylcholine receptor agonists on cognition in rhesus monkeys with a chronic cocaine self-administration history. *Neuropharmacology*. 2013; 64:479–88. [PubMed: 22921923]
- Hackam DG. Translating animal research into clinical benefit. *BMJ*. 2007; 334:163–4. [PubMed: 17255568]
- Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. *JAMA*. 2006; 296:1731–2. [PubMed: 17032985]
- Kaiyala KJ, Leroux BG, Watson CH, Prall CW, Coldwell SE, Woods SC, Ramsay DS. Reliability of individual differences in initial sensitivity and acute tolerance to nitrous oxide hypothermia. *Pharmacol Biochem Behav*. 2001; 68:691–9. [PubMed: 11526966]
- Kalant H, LeBlanc AE, Gibbins RJ. Tolerance to, and dependence on, some non-opiate psychotropic drugs. *Pharmacol Rev*. 1971; 23:135–91. [PubMed: 4398655]
- Kangas BD, Branch MN. Relations among acute and chronic nicotine administration, short-term memory, and tactics of data analysis. *J Exp Anal Behav*. 2012; 98:155–167. [PubMed: 23008520]
- Katner SN, Davis SA, Kirsten AJ, Taffe MA. Effects of nicotine and mecamylamine on cognition in rhesus monkeys. *Psychopharmacology (Berl)*. 2004; 175:225–40. [PubMed: 15112030]
- Kazdin, AE. Single-case research designs: Methods for clinical and applied settings. 2. Oxford University Press; New York, NY: 2011.
- Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*. 2004; 3:711–5. [PubMed: 15286737]
- Madsen LG, Bytzer P. Review article: Single subject trials as a research instrument in gastrointestinal pharmacology. *Aliment Pharmacol Ther*. 2002; 16:189–96. [PubMed: 11860401]
- McDonald, JH. Handbook of Biological Statistics. 2. Sparky House Publishing; Baltimore, MD: 2009.
- Morgan DL, Morgan RK. Single-participant research design. Bringing science to managed care. *Am Psychol*. 2001; 56:119–27. [PubMed: 11279805]
- Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, Macleod M, Mignini LE, Jayaram P, Khan KS. Comparison of treatment effects between animal experiments and clinical trials: systematic review. *BMJ*. 2007; 334:197. [PubMed: 17175568]
- Perkins KA, Lerman C, Coddington SB, Jetton C, Karelitz JL, Scott JA, Wilson AS. Initial nicotine sensitivity in humans as a function of impulsivity. *Psychopharmacology (Berl)*. 2008; 200:529–44. [PubMed: 18604520]
- Prendergast MA, Jackson WJ, Terry AV Jr, Decker MW, Arneric SP, Buccafusco JJ. Central nicotinic receptor agonists ABT-418, ABT-089, and (–)-nicotine reduce distractibility in adult monkeys. *Psychopharmacology (Berl)*. 1998; 136:50–8. [PubMed: 9537682]
- Prendergast MA, Terry AV Jr, Jackson WJ, Marsh KC, Decker MW, Arneric SP, Buccafusco JJ. Improvement in accuracy of delayed recall in aged and non-aged, mature monkeys after intramuscular or transdermal administration of the CNS nicotinic receptor agonist ABT-418. *Psychopharmacology (Berl)*. 1997; 130:276–84. [PubMed: 9151363]
- Sidman, M. Tactics of scientific research: Evaluating experimental data in psychology. Basic Books; New York, NY: 1960.
- Smolen A, Marks MJ, DeFries JC, Henderson ND. Individual differences in sensitivity to nicotine in mice: response to six generations of selective breeding. *Pharmacol Biochem Behav*. 1994; 49:531–40. [PubMed: 7862704]
- Sohn D. Statistical significance and replicability: Why the former does not presage the latter. *Theory & Psychology*. 1998; 8:291–311.
- Terry AV Jr, Buccafusco JJ, Bartoszyk GD. Selective serotonin 5-HT_{2A} receptor antagonist EMD 281014 improves delayed matching performance in young and aged rhesus monkeys. *Psychopharmacology (Berl)*. 2005; 179:725–32. [PubMed: 15619109]
- Terry AV Jr, Buccafusco JJ, Jackson WJ. Scopolamine reversal of nicotine enhanced delayed matching-to-sample performance in monkeys. *Pharmacol Biochem Behav*. 1993; 45:925–9. [PubMed: 8415832]
- Terry AV Jr, Buccafusco JJ, Jackson WJ, Prendergast MA, Fontana DJ, Wong EH, Bonhaus DW, Weller P, Eglen RM. Enhanced delayed matching performance in younger and older macaques

administered the 5-HT₄ receptor agonist, RS 17017. *Psychopharmacology (Berl)*. 1998; 135:407–15. [PubMed: 9539266]

Weed MR, Bryant R, Perry S. Cognitive development in macaques: attentional set-shifting in juvenile and adult rhesus monkeys. *Neuroscience*. 2008; 157:22–8. [PubMed: 18805462]

Fig. 1.

Percentage correct trials in the DMTS procedure on trials with 2-, 30-, and 300-s delays. The graphs in the top row depict percentage of correct trials in sessions that occurred during determination of a range of doses of nicotine (0.001 – 0.056 mg/kg): nicotine administration sessions and vehicle sessions (leftmost graph), individually-determined best dose sessions and vehicle sessions (second graph from left); best vehicle session and other vehicle sessions (third graph from the left); best baseline session and other baseline sessions (fourth graph from the left); and best dose, best vehicle, and best baseline sessions (last graph from the left). The graphs in the bottom row depict percentage correct trials in sessions that occurred during determination of a range of doses of methylphenidate (0.01 – 0.3 mg/kg). Each data point represents the average of 7 or 8 monkeys. Error bars represent $\pm$ SEM. * indicates a statistically significant difference from other data points at same parameter value

Fig. 2. Percentage correct trials observed in the SOSS procedure on trials with 2, 3, and 4 boxes. Each data point represents the average of 6 monkeys. All other details are as in Fig 1

Fig. 3.

Percentage correct trials, overall, observed in the SOSS procedure after vehicle administration (points above "V") and nicotine administration (0.001 – 0.056 mg/kg) during the first and second nicotine dose-response curve re-determinations for individual monkeys. Points marked with a "+" symbol are the individually-determined best performances

Fig. 4. Percentage correct trials for three monkeys during SOSS sessions following administration of each monkey's individually-determined best dose of nicotine and interspersed baseline and vehicle sessions

Fig. 5.

The percentage of statistically significant outcomes obtained from Monte Carlo simulations in which a best dose analysis was applied to sets of randomly generated data selected from a normal distribution with a mean of 80 and a standard deviation of 0.5 (top left graph), 2.5 (top right graph), 5 (bottom left graph), and 10 (bottom right graph)

Fig. 6.

The percentage of statistically significant outcomes obtained from Monte Carlo simulations in which a best dose analysis was applied to sets of randomly generated data selected from a normal distribution with a mean of 70 and a standard deviation of 0.5 (top left graph), 2.5 (top right graph), 5 (bottom left graph), and 10 (bottom right graph)

Fig. 7. The percentage of statistically significant outcomes obtained from Monte Carlo simulations in which a best dose analysis was applied to sets of randomly generated data selected from a normal distribution with a mean of 60 and a standard deviation of 0.5 (top left graph), 2.5 (top right graph), 5 (bottom left graph), and 10 (bottom right graph)

Table 1

Statistical outcomes of analyses comparing performances associated with best doses of nicotine, and methylphenidate to performances associated with all vehicle sessions (“Best Dose”), performance associated with best vehicle sessions compared to remaining vehicle sessions (“Best Vehicle”), performances associated with best baseline sessions compared to remaining baseline sessions (“Best Baseline”), and performances associated with best dose, best vehicle, and best baseline sessions (“Best Dose, Best Vehicle, Best Baseline”).

	Nicotine	Methylphenidate
Delayed Matching-to-Sample		
Best Dose (vs. All Vehicle)	$F_{1,6}=71.376, p<0.001$	$F_{1,6}=21.987, p=0.003$
Best Vehicle (vs. Other Vehicle)	$F_{1,6}=59.823, p<0.001$	$F_{1,6}=28.727, p=0.002$
Best Baseline (vs. Other Baseline)	$F_{1,6}=23.455, p=0.003$	$F_{1,6}=42.273, p<0.001$
Best Dose, Best Vehicle, Best Baseline	$F_{2,12}=0.365, p=0.702$	$F_{2,14}=1.708, p=0.217$
Self-Ordered Spatial Search		
Best Dose (vs. All Vehicle)	$F_{1,5}=11.539, p=0.003$	$F_{1,5}=8.468, p=0.033$
Best Vehicle (vs. Other Vehicle)	$F_{1,5}=20.088, p=0.007$	$F_{1,5}=8.468, p=0.033$
Best Baseline (vs. Other Baseline)	$F_{1,5}=35.923, p=0.002$	$F_{1,5}=7.288, p=0.043$
Best Dose, Best Vehicle, Best Baseline	$F_{2,10}=1.065, p=0.381$	$F_{2,10}=3.129, p=0.088$

Table 2

Individually-determined best doses for the SOSS group from the two re-determinations of the nicotine dose-response curve conducted 3 weeks apart.

Monkey	First Re-determination	Second Re-determination
98003	0.01	0.003
98007	0.056	0.01
RQ6101	0.001	0.003/0.056 ^a
RQ6178	0.01	0.056
RQ6118	0.001	0.03
14P	0.001	0.056

^aTie for overall accuracy between 0.003 and 0.056 mg/kg