Published in final edited form as:

Drug Alcohol Depend. 2015 September 1; 154: 300–303. doi:10.1016/j.drugalcdep.2015.07.011.

Examination of a Recommended Algorithm for Eliminating Nonsystematic Delay Discounting Response Sets

Thomas J. White^{1,2}, Ryan Redner^{1,2}, Joan M. Skelly³, and Stephen T. Higgins^{1,2,4}

¹Vermont Center on Behavior and Health, University of Vermont

²Departments of Psychiatry, University of Vermont

³Medical Biostatistics, University of Vermont

⁴Psychology, University of Vermont

Abstract

Purpose—To examine (1) whether use of a recommended algorithm (Johnson and Bickel, 2008) improves upon conventional statistical model fit (R2) for identifying nonsystematic response sets in delay discounting (DD) data, (2) whether removing such data meaningfully effects research outcomes, and (3) to identify participant characteristics associated with nonsystematic response sets.

Methods—Discounting of hypothetical monetary rewards was assessed among 349 pregnant women (231 smokers and 118 recent quitters) via a computerized task comparing \$1000 at seven future time points with smaller values available immediately. Nonsystematic response sets were identified using the algorithm and conventional statistical model fit (R2). The association between DD and quitting was analyzed with and without nonsystematic response sets to examine whether the inclusion or exclusion impacts this relationship. Logistic regression was used to examine whether participant sociodemographics were associated with nonsystematic response sets.

Results—The algorithm excluded fewer cases than the R2 method (14% vs. 16%), and was not correlated with log k as is R2. The relationship between log k and the clinical outcome (spontaneous quitting) was unaffected by exclusion methods; however, other variables in the model were affected. Lower educational attainment and younger age were associated with nonsystematic response sets.

Correspondent: Thomas J. White, Vermont Center on Behavior and Health, University of Vermont, UHC Campus, Rm 1415D, 1 So. Prospect St. Burlington, VT 05401, tjwhite@uvm.edu, Phone: 802-656-8714, Fax: 802-656-5793.

Contributors: Dr. White and Dr. Redner devised the concept for the paper and proposed the analyses; Dr. White wrote the manuscript. Ms. Skelly conducted the data analyses and wrote the statistical methods section. She approved the statistical language in the results section and tables, and provided feedback on the manuscript. Dr. Higgins was the lab director and principal investigator for the pregnant smokers project from which this data set derived. He was also the postdoctoral mentor to Dr. White and Dr. Redner. All authors have read and approved the submission of this manuscript to *Drug and Alcohol Dependence*.

Conflicts of interest: none.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the Food and Drug Administration.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—The algorithm eliminated data that were inconsistent with the nature of discounting and retained data that were orderly. Neither method impacted the smoking/DD relationship in this data set. Nonsystematic response sets are more likely among younger and less educated participants, who may need extra training or support in DD studies.

Keywords

delay discounting; temporal discounting; hyperbolic discounting; R²; outliers; behavioral economics

1. INTRODUCTION

Delay discounting (DD) is a behavioral-economic concept that measures reductions in the subjective value of consequences as a function of temporal delays to their delivery, with decrements in subjective value occurring relatively rapidly at shorter delays, and then less rapidly as delays become longer (Bickel and Marsch, 2001; Critchfield and Kollins, 2001). Some questions remain whether DD is a stable trait variable representing a type of impulsivity (Lowenstein et al., 2007) or personality characteristic (Odum, 2011) or whether the rate of discounting can be manipulated (Bickel et al., 2012). Because a number of interventions have been shown to alter discounting rate (Koffarnus et al., 2013), DD may prove to be an important target for treatment of the myriad of health-related risk behaviors and associated problems with which it is associated, including smoking and other substance abuse, HIV and other sexual risk behavior, pathological gambling, eating disorders, bipolar disorder, and adherence with disease prevention regimens (Bickel et al., 2012; Bickel and Marsch, 2001; Bradford et al., 2010; Davis et al., 2010; Herrmann et al., 2014; MacKillop et al., 2011; Rogers et al., 2010).

One practical challenge when attempting to predict real-world behavior using a self-report laboratory task is determining whether a subject's responses are conceptually interpretable (Critchfield and Kollins, 2001). DD research has incorporated the statistic R^2 from linear regression to measure deviation from the hyperbolic curve; this statistic is typically described as measuring goodness-of-fit in delay discounting research (e.g., Green and Myerson, 1995; Myerson and Green, 1995; Ohmura et al., 2006; Rachlin et al., 1991). However, the use of R^2 has significant problems. In general, the use of model fit as a way to assess data orderliness is itself limited because it imposes the presumption of a specific model onto data. Additionally and independently, R^2 is systematically confounded because it uses a flat function as the comparator model, which means that R^2 becomes systematically more stringent at lower discounting rates because the discounting function itself starts to resemble the comparator model. This means that R^2 is biased towards labeling steeper (more impulsive) discounting curves as inherently better fitting than shallower discounting curves, which introduces a systematic bias against shallow discounters (Johnson and Bickel, 2008).

Johnson and Bickel (2008) strongly recommended the use of algorithms that test basic assumptions of discounting data (namely, that the immediate value of rewards should diminish as the delay to receiving those rewards increases, and that this process is unidirectional, i.e. that the immediate value of delayed rewards only diminishes as delay increases). In this (2008) paper, the authors proposed and tested two criteria that were

examples of the algorithmic approach. These example criteria (called *Rule 1* and *Rule 2* for the rest of this paper) were defined as follows: in Rule 1, cases in which any indifference point was greater than its predecessor by a magnitude exceeding 20% of the larger later reward (e.g., an increase of over \$200 between two sequential time points on a task in which the large later reward is \$1000); in Rule 2, if the indifference point at the final time point was not lower than the first time point by at least 10% of the large later reward (e.g., when the large later reward is \$1000, the indifference point at the farthest time point is less than \$100 smaller than the indifference point at the closest time point). The present study examines the relative merits of Johnson and Bickel's algorithm and conventional statistical model fit for characterizing DD data. Although data elimination was not the primary focus of the 2008 article, the current report examines how a DD data set is affected by eliminating data via the use of algorithms and statistical model fit, which were operationalized here as *Rule 1* and *Rule 2* vs. multiple cutoffs of R^2 (0, .5 and .9). Like the algorithms, these cutoffs are not set in stone; various cutoffs can be found in the literature (e.g., .3, .4, etc.). In the current paper, three questions are asked: First, what percentage of a data set is eliminated when using these methods? Second, do these methods of eliminating data lead to meaningful outcome changes when DD is used as a predictor of smoking status? Third, do basic sociodemographic characteristics predict nonsystematic response sets on a computerized DD task?

2. METHOD

2.1. Participants

Participants were 349 pregnant women who were regular smokers upon learning of their pregnancy: 118 of them quit smoking prior to initiating prenatal care, and 231 were still smoking at the start of prenatal care. Smoking status for all participants was biochemically verified using urine cotinine testing. Participants were recruited from obstetrical clinics located in the greater Burlington, Vermont area. The University of Vermont College of Medicine's Institutional Review Board approved these studies, and all participants provided written informed consent.

2.2. Measures

All participants completed a computerized DD task, which has been described previously (Johnson and Bickel, 2002; White et al., 2014). Briefly, participants were seated in front of the computer screen, which displayed the following message:

Imagine that you have a choice between waiting [length of time] and then receiving \$1,000 and receiving a smaller amount of money right away. Please choose between the two options.

Seven delays were examined (1 day, 1 week, 1 month, 6 months, 1 year, 5 years, or 25 years); participants always chose between \$1,000 at one of these seven delays and a smaller amount available immediately. The program presented different values of the immediate reward until an indifference point was identified, where the value of the immediate amount was subjectively equivalent to the delayed \$1,000 reward (Johnson and Bickel, 2002). This process was repeated for all time delays. The baseline DD assessment was not available for

27 women (9% of the overall sample); for these women, their DD data from the next available follow-up visit was used instead.

2.3. Statistical Methods

To analyze DD results, Mazur's (1987) hyperbolic equation, $V = A/(1+kD)$, was fitted to each subject's DD data across the seven temporal delays (1 day, 1 week, 1 month, 6 months, 1 year, 5 years, and 25 years) using nonlinear regression (SAS PROC NLIN), which generated both R^2 and the discounting parameter k . Because the distribution of estimated k values was skewed, analyses were performed using a log transformation of k . Each of the 349 cases were then assessed for meeting Rules 1 and 2 and dummy-coded for exclusion uniquely by criteria (e.g., we noted which cases should be removed for each of the rules so that we could address their use separately and in combination). Participants' data sets were coded as systematic or nonsystematic based on the criteria set by the algorithm and by R^2 , and categorized by the number and percentage falling into various ranges of R^2 representing typical but arbitrary cutoffs (0.0, 0.5, 0.9). Therefore, the differences in data classification between the algorithm and R^2 could be examined.

Next, simple logistic regression was used to examine the relationship between log k and spontaneous quitting with no data excluded and with cases excluded based on the algorithm and statistical model fit (i.e., excluded based on Rule 1, Rule 2, both rules, and $R^2 < 0$ or missing). Then, in order to determine whether the exclusion methods may have affected the outcome of our study, we repeated the final logistic regression model from our previous paper (White et al., 2014) for each of the following exclusion rules: by Rule 1, Rule 2, both rules, or $R^2 < 0$ or missing.

Finally, a series of logistic regression analyses were performed in order to examine whether provision of nonsystematic response sets could be predicted by baseline measures including sociodemographic variables, smoking history and psychiatric symptoms. First, simple logistic regression was run to predict provision of nonsystematic response sets for Rule 1, Rule 2, and either Rule 1 or Rule 2, using each of the baseline measures. Significant predictors from the simple logistic regressions were then combined in a backwards elimination logistic regression procedure which removed nonsignificant variables in a stepwise manner to create final models. All analyses were performed with SAS Version 9.1 statistical software (SAS Institute, Cary, NC, USA). Statistical significance was defined as $p < .05$.

3. RESULTS

The use of the algorithm to exclude data sets resulted in 14% (49/349) of cases being excluded (Table 1): Rule 1 eliminated 25 cases, Rule 2 eliminated 32 cases, and the use of both criteria eliminated 49 cases (i.e., 8 participants met exclusion criteria under both rules). The use of R^2 excluded 16% (55/349) of cases; 45 were excluded because $R^2 < 0$, and 10 were removed because indifference points were identical across all seven time points (flat response set) and so a R^2 statistic could not be generated. In many instances, the data excluded by the algorithm were also excluded using R^2 : 35 of the 49 cases (71%) excluded by the algorithm also met exclusion criteria by R^2 , and conversely, 35 of the 55 cases (64%)

excluded by R^2 met exclusion criteria by the algorithm. Importantly, R^2 excluded 20 cases that were classified as orderly by the algorithm, and the algorithm excluded 14 cases that were classified as orderly by R^2 .

The relationship between log k and smoking status remained significant regardless of whether cases were excluded using algorithmic or model fit methods (Table 2). Briefly, the model predicted spontaneous quitting in this population with DD, educational attainment, DD x smoking rate, belief that smoking will harm the baby, age of smoking initiation, stress, marital status, and number of pre-pregnancy quit attempts (White et al., 2014). When Rule 1 was applied, the model remained intact. When Rule 2 was applied, age of smoking initiation and marital status were no longer significant. When both rules were applied, age of smoking initiation and stress were no longer significant.

Nonsystematic response sets as defined by Rule 1 (i.e. one or more indifference points were >20% greater than at the previous time delay) were predicted by younger age and lower educational attainment in simple logistic regressions. When the two predictors were combined in a backward elimination logistic regression procedure, younger age predicted nonsystematic response sets based on Rule 1 (OR and 95% CI: 0.85 (0.76 – 0.96); $p = .01$). Nonsystematic response sets for Rule 2 (i.e., final indifference point is > 90% of the value of the initial indifference point) were not significantly associated with any of the baseline measures. When predicting provision of nonsystematic response sets based on either/both criteria, both younger age and lower educational attainment predicted nonsystematic data in simple logistic regressions. When the two predictors were combined in a backward elimination procedure, only education remained in the model. Women with more than a high school education had 0.2 times the odds of nonsystematic response sets as defined by either/both criteria compared to women with less than a high school education (OR and 95% CI: 0.76 (0.38 – 1.53) for high school vs < high school, 0.20 (0.07 – 0.58) for > high school vs < high school; $p = .01$).

4. DISCUSSION

This study aimed to systematically examine the differential impact of two methods for characterizing the orderliness of DD data: the algorithmic method and conventional statistical model fit. DD data were systematically eliminated from a large data set of data from pregnant smokers, using various combinations of the methods, to see the effects on the data set. We found that in our data set (349 pregnant women who were smoking when they learned of pregnancy), a large percentage of subjects (14%) provided data that were not systematic in keeping with the nature of discounting. The algorithm established by Johnson and Bickel (2008) identified fewer cases as non-systematic than did the conventional R^2 statistic. Overall, the algorithm was superior in its discriminative ability to target cases that did not keep with the assumptions of discounting (i.e., that it occurs and that it is unidirectional) when compared to the conventional model fit statistics, which had poorer selectivity and a systematic bias (e.g., in the present report 20 cases were excluded by R^2 that would be considered orderly by the algorithm). The number of cases excluded by the algorithm were slightly fewer (14% vs. 16%), but much more importantly, the selectivity of

the algorithm to identify data that are not systematic and retain data that are systematic is the essence of its importance.

Excluding cases based on nonsystematic data using either method did not meaningfully affect the overall relationship between DD and spontaneous quitting in this specific data set with pregnant smokers. This may be true in part because this data set is large. However, it is possible that in other data sets, the inclusion or exclusion of nonsystematic data may significantly change the relationship observed between log k and clinical outcomes. To our knowledge, ours is the first analysis to treat nonsystematic discounting data as a dependent variable. This practice may prove helpful to researchers further seeking to characterize participants who provide problematic data on behavioral economic tasks.

Algorithmic approaches, including but not limited to the one created by Johnson and Bickel (2008), evidence significant benefits over conventional model fit statistics such as R^2 : they are not correlated with k , they are not systematically biased against shallower discounting curves as is R^2 , and (as seen in the present study) when used for data elimination, they remove fewer cases than does R^2 . In the present report we attempted to identify participant characteristics that predicted non-systematic responding. Younger participants and those with lower levels of educational attainment had an increased likelihood of providing nonsystematic response sets on this DD task. This suggests that participants with these characteristics may benefit from increased training, explanations, or other support on DD tasks.

Acknowledgments

Role of Funding Source: nothing declared.

This research was supported by National Institutes of Health Center of Biomedical Research Excellence award P20GM103644 from the National Institute of General Medical Sciences, Tobacco Centers of Regulatory Science award P50DA036114 from the National Institute on Drug Abuse and U.S. Food and Drug Administration, research grants R01DA14028 and R01HD075669 from the National Institute on Drug Abuse and National Institute of Child Health and Human Development, respectively, and Institutional Training grant T32DA07242 from the National Institute on Drug Abuse.

References

- Bickel WK, Jarmolowicz DP, Mueller ET, Koffarnus MN, Gatchalian KM. Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: emerging evidence. *Pharmacol Therapeut.* 2012; 134:287–297.
- Bickel WK, Marsch LA. Toward a behavioral economic understanding of drug dependence: delay discounting processes. *Addiction.* 2001; 96:73–86. [PubMed: 11177521]
- Bradford WD. The association between individual time preferences and health maintenance habits. *Med Decis Making.* 2010; 30:99–112. [PubMed: 19675322]
- Critchfield TS, Kollins SH. Temporal discounting: basic research and the analysis of socially important behavior. *J Appl Behav Anal.* 2001; 34:101–122. [PubMed: 11317983]
- Davis C, Patte K, Curtis C, Reid C. Immediate pleasures and future consequences. A neuropsychological study of binge eating and obesity. *Appetite.* 2010; 54:208–213. [PubMed: 19896515]
- Green L, Myerson J. Exponential versus hyperbolic discounting of delayed outcomes: risk and waiting time. *Am Zool.* 1996; 36:496–505.

- Herrmann ES, Hand DJ, Johnson MW, Badger GJ, Heil SH. Examining delay discounting of condom-protected sex among opioid-dependent women and non-drug-using control women. *Drug Alcohol Depend.* 2014; 144:53–60. [PubMed: 25190049]
- Johnson MW, Bickel WK. Within-subject comparison of real and hypothetical money rewards in delay discounting. *J Exp Anal Behav.* 2002; 77:129–146. [PubMed: 11936247]
- Johnson MW, Bickel WK. An algorithm for identifying nonsystematic delay-discounting data. *Exp Clin Psychopharm.* 2008; 16:264.
- Koffarnus MN, Jarmolowicz DP, Mueller ET, Bickel WK. Changing delay discounting in the light of the competing neurobehavioral decision systems theory: a review. *J Exp Anal Behav.* 2014; 99:32–57. [PubMed: 23344987]
- Loewenstein G, Brennan T, Volpp KG. Asymmetric paternalism to improve health behaviors. *JAMA.* 2007; 298:2415–2417. [PubMed: 18042920]
- MacKillop J, Amlung MT, Few LR, Ray LA, Sweet LH, Munafò MR. Delayed reward discounting and addictive behavior: a meta-analysis. *Psychopharmacology.* 2011; 216:305–321. [PubMed: 21373791]
- Mazur, JE. An adjustment procedure for studying delayed reinforcement. In: Commons, ML.; Nevin, JA., editors. *The Effect Of Delay And Of Intervening Events On Reinforcement Value.* Lawrence Erlbaum Associates, Inc; Hillsdale, NJ: 1987. p. 55-73.
- Myerson J, Green L. Discounting of delayed rewards: models of individual choice. *J Exp Anal Behav.* 1995; 64:263–276. [PubMed: 16812772]
- Odum AL. Delay discounting: I'm a k, you're a k. *J Exp Anal Behav.* 2011; 96:427–439. [PubMed: 22084499]
- Ohmura Y, Takahashi T, Kitamura N, Wehr P. Three-month stability of delay and probability discounting measures. *Exp Clin Psychopharmacol.* 2006; 14:318–328. [PubMed: 16893275]
- Rachlin H, Raineri A, Cross D. Subjective probability and delay. *J Exp Anal Behav.* 1991; 55:233–244. [PubMed: 2037827]
- Ratkowsky, DA. *Handbook Of Nonlinear Regression Models.* Marcel Dekker, Inc; New York: 1990.
- Rogers RD, Moeller FG, Swann AC, Clark L. Recent research on impulsivity in individuals with drug use and mental health disorders: implications for alcoholism. *Alcohol Clin Exp Res.* 2010; 34:1319–1333. [PubMed: 20528825]
- White TJ, Redner R, Skelly JM, Higgins ST. Examining educational attainment, pre-pregnancy smoking rate, and delay discounting as predictors of spontaneous quitting among pregnant smokers. *Exp Clin Psychopharmacol.* 2014; 22:384–391. [PubMed: 25069014]

Highlights

- Nonsystematic delay discounting response sets were identified using an algorithm.
- The algorithm removes fewer cases than the conventional R^2 statistic.
- Removing nonsystematic response sets did not change the relationship between discounting and quitting smoking in this data set.
- Younger age and lower educational attainment both predict provision of nonsystematic data sets.

Table 1

Comparison of exclusion across rules, organized by R^2 value.

R^2 range	Number of cases in range	Rule 1 exclusions in range	Rule 2 exclusions in range	Total exclusions in range	Percentage excluded in range
Cannot be calculated	10	0	10	10	100%
< 0.0	45	11	20	25 ^a	56%
0.0 – < 0.5	32	7	2	7 ^b	22%
0.5 – < 0.9	99	7	0	7	7%
0.9	163	0	0	0	0%

^aIn this range, there were 6 cases who would be excluded by both rules 1 and 2.

^bIn this range, there were 2 cases who would be excluded by both rules 1 and 2.

Table 2Associations between discounting ($\log k$) and quitting across delay discounting exclusion criteria.

Exclusion rule	<i>N</i>	<i>OR</i>	95% CI	<i>p</i>
No exclusions	349	0.87	(0.79, 0.96)	0.004
Use Rule 1 only	324	0.87	(0.79, 0.96)	0.006
Use Rule 2 only	317	0.82	(0.72, 0.93)	0.003
Use both rules	300	0.82	(0.72, 0.94)	0.003
Exclude if $R^2 < 0$ or missing	294	0.79	(0.69, 0.91)	0.001