Published in final edited form as:

*J Exp Anal Behav*. 2019 March ; 111(2): 207–224. doi:10.1002/jeab.497.

# A Monte Carlo Method for Comparing Generalized Estimating Equations to Conventional Statistical Techniques for Discounting Data

**Jonathan E. Friedel**[1], **William B. DeHart**[2], **Anne M. Foreman**[1], and **Michael E. Andrew**[1]

[1]National Institute for Occupational Safety and Health

[2]Virginia Tech Carillion Research Institute

## Abstract

Discounting is the process by which outcomes lose value. Much of discounting research has focused on differences in the degree of discounting across various groups. This research has relied heavily on conventional null hypothesis significance testing that is familiar to psychology in general such as *t*-tests and ANOVAs. As discounting research questions have become more complex by simultaneously focusing on within-subject and between-group differences conventional statistical testing is often not appropriate for the obtained data. Generalized estimating equations (GEE) are one type of mixed-effects model that are designed to handle autocorrelated data, such as within-subject repeated-measures data, and are therefore more appropriate for discounting data. To determine if GEE provides similar results as conventional statistical tests, were compared the techniques across 2,000 simulated data sets. The data sets were created using a Monte Carlo method based of an existing data set. Across the simulated data sets, the GEE and the conventional statistical tests generally provided similar patterns of results. As the GEE and more conventional statistical tests provide the same pattern of result, we suggest researchers use the GEE because it was designed to handle data that has the structure that is typical of discounting data.

### Keywords

discounting; generalized estimating equations; mixed-effects models; Monte Carlo

Delay discounting describes the process by which delayed outcomes lose value (Mazur, 1987; Odum, 2011a). For a given individual, as the degree of delay discounting for an outcome increases the effective value of that outcome decreases when it is delayed. Across individuals, someone with a higher degree of delay discounting will value a delayed outcome less than a person with a lower degree of delay discounting. A wide variety of research assessing the degree of discounting of people who exhibit maladaptive behaviors has been conducted. People who exhibit maladaptive behaviors tend to more steeply discount delayed outcomes than those who do not exhibit the behaviors. For example, people who smoke cigarettes (Bickel, Odum, & Madden, 1999; Friedel, DeHart, Madden, & Odum, 2014; Mitchell, 1999), people who are obese or overeat (Rasmussen, Lawyer, & Reilly, 2010), people who engage in problematic gambling (Dixon, Marley, & Jacobs, 2003), and people diagnosed with attention deficit hyperactivity disorder (Barkley, Edwards, Laneri, Fletcher, & Metevia, 2001) all exhibit higher degrees of discounting than their respective matched controls. Additionally, within individuals the degree of discounting exhibited for one outcome is related to the degree of discounting for other outcomes (Charlton & Fantino, 2008; Friedel, DeHart, Frye, Rung, & Odum, 2016; Friedel et al., 2014; Odum, 2011b; Odum, Baumann, & Rimington, 2006; Rasmussen et al., 2010). Together, these results have drawn a wide audience, in part because overly steep delay discounting appears to be a trans-disease process (Bickel, Jarmolowicz, Mueller, Koffarnus, & Gatchalian, 2012).

There are several commonly used direct measures and several derived measures of the degree of discounting. The most common direct measure of discounting are indifference points. Indifference points are the immediate amount of an outcome that is equal to a larger but delayed amount of that outcome. As the delayed outcome has a discounted value, this smaller but immediate amount has the same reinforcing efficacy as the delayed outcome. A popular derived measure of discounting is area under the curve (AUC; Myerson, Green, & Warusawitharana, 2001). Area under the curve is the integral of indifference points standardized by the longest delay to receiving the outcome and the amount of the delayed outcome (Myerson et al., 2001). The measure is bound between 0 and 1, with higher AUC values indicating lower degrees of delay discounting and lower AUC values indicating higher degrees of delay discounting.

Another technique to describe discounting data is to quantify the relation between delay and the value of an outcome by fitting theoretical models of delay discounting (Mazur, 1987; Myerson & Green, 1995; Rachlin, 2006) to the indifference points. That is, models of discounting are fit to the obtained indifference points and the fitted models quantify the relation between delay and the indifference points. All models of discounting take the shape of a sloped, decreasing function. Another measure of discounting that is derived from the fitted models has seen recent increased use: the effective-delay 50 (ED-50; Yoon & Higgins, 2008). The ED-50 is the delay at which the larger, delayed outcome has lost half of its value. The measure is derived after fitting a theoretical model to the indifference points (Franck, Koffarnus, House, & Bickel, 2015). Effective-delay 50 is usually obtained through interpolation using the fitted model.

Often the goal of obtaining these measures of discounting is to compare the degree of discounting based on some other question of interest (e.g., group differences). These research questions about similarities or differences in the degree of discounting typically rely on the derived measures of discounting described above. For these sorts of questions, statistical comparisons are made on AUC, free parameters obtained from the theoretical models of discounting, or ED-50.

The search for differences in delay discounting between groups has relied heavily on traditional inferential null-hypothesis statistical testing (NHST). With NHST, researchers have generally have attempted to determine how compatible measured group differences in the degree of discounting are with the null hypothesis that there are no differences (e.g., Wasserstein & Lazar, 2016). Often, researchers have relied on parametric NHSTs in studies of discounting. For example, in an early study examining differences in the degree of delay discounting Green, Fry, and Myerson (1994) used a series of *t*-tests to compare the degree of discounting between children, young adults, and older adults. In another early study, Petry (2001) used a combination of *t*-tests and repeated measures ANOVAs to examine the differences in the degree of discounting for alcohol and money in actively using alcoholics, currently abstinent alcoholics, and controls. It is also common for research to rely on non-parametric NHSTs when the delay discounting data structure does not meet the assumptions of parametric NHSTs. In the seminal paper, Bickel et al. (1999) used a Kruskal-Wallis test, Mann-Whitney U tests, and Wilcoxon rank-sum tests to compare the degree of discounting for cigarettes and money in current smokers, ex-smokers, and never smokers.

As interest in discounting has grown, researchers are asking more complex research questions (e.g., simultaneous between-group and within-subjects differences in discounting, changes in the degree of discounting over time) that often rely on more complex study designs (e.g., clinical controlled trials, longitudinal designs). The conventional statistical tests used to compare delay discounting either between or within groups may no longer be adequate in these circumstances. For example, (Friedel et al., 2016) examined differences in the degree of delay discounting for various outcomes (e.g., monetary gains, monetary losses, temporary increases in health, and temporary relief from a debilitating disease) in cigarette smokers and non-smokers. A mixed-effects ANOVA would be the conventional statistical test for such data because there were group comparisons (smokers vs. non-smokers) and within-subject comparisons (qualitatively different outcomes). However, the measures of delay discounting were not normally distributed making a mixed-effects ANOVA inappropriate. In addition, many of the assumptions of typical inferential tests including homogeneity of variance, normal distribution of the dependent variable, and normal distribution of the residuals were not met. The parametric tests could have been used without regard to violating the assumptions of the test, but when the assumptions of a statistical test are violated then there is an increase in the likelihood of a Type I error (i.e., false positive). One solution to analyze such data is to conduct a non-parametric statistical tests. However, a non-parametric approach was not possible because there is no widely available non-parametric version of a mixed-effects ANOVA. Using a battery of non-parametric tests to approximate a mixed-effects ANOVA is also undesirable because using multiple tests (e.g., multiple *t*-tests, Mann-Whitney U tests, ANOVAs, etc) within the same analyses will also increase the Type I error rate. This complex data structure highlights the need for discounting researchers to adopt more advanced statistical techniques.

One test—the Generalized Estimating Equation (GEE)— can address many of the above mentioned shortcomings of conventional tests. A GEE is a semi-parametric test that accounts for the inter-correlation of repeated measures, similar to the random intercept of a mixed-effect model (Hanley, Negassa, Edwardes, & Forrester, 2003). GEE is different from a mixed-effect model (e.g., multi-level model) in that it accounts for clustering but the analytical focus is on group differences. GEE is also relatively robust against violations of the assumptions of alternative tests when compared to ANOVA. A GEE can potentially allow for better between group comparisons by controlling the variance introduced into the data by repeated measures (i.e., clustering) from a single participant in a data set. This means that a GEE can be conducted directly on indifference points instead of another derived measure. Additionally, the output of a GEE can be interpreted similarly to the output of a multiple regression or an ANOVA. For example, GEE results include a beta value (e.g., regression) and a Robust Z score (e.g., t-test-like value adjusted for inter-correlations). Furthermore, pairwise comparisons (with the appropriate adjustment for multiple comparisons) can easily be conducted to compare specific between- and within-group means as with other typical post-hoc tests. A single GEE with several post-hoc tests was used to analyze the complex data structure described above for Friedel et al. (2016). Although the results were similar to an ANOVA, as the GEE is more robust to violations the analyses provided a more accurate description of the results and provided greater confidence in the conclusions drawn.

The goal of this study is to compare the analyses of discounting data with conventional statistical methods—specifically a mixed-effects ANOVA—and GEE methods. Generalized Estimating Equations have not been widely used in delay discounting research. Although they are similar to mixed-effects ANOVAs, it is possible that the output of a GEE could be drastically different than a mixed-effects ANOVA output for discounting data. A Monte Carlo method was used to facilitate the comparisons between the statistical techniques. Monte Carlo methods use various forms of random sampling to simulate data (Kroese, Brereton, Taimre, & Botev, 2014). Here, a Monte Carlo method was used to randomly select data with replacement from a previously published study (Hayashi, Fessler, Friedel, Foreman, & Wirth, 2018) to create new simulated data sets. In total, 2,000 simulated data sets were created based on the experimentally obtained data. For 1,000 of the simulated data sets, participants were randomly selected with replacement from within their respective group. With the other 1,000 simulated data sets, participants were randomly selected from across the groups and then were randomly assigned to a group. The goal of the second set of simulations was to assess whether the GEE was less likely than the conventional statistical tests to detect a false-positive in regards to the effect of group on the degree of discounting when group membership was randomly determined (i.e., a null effect). If there was a real difference in discounting between the groups it should be reliably detected across the simulations in which participants were drawn from within their groups. Additionally, in the simulations in which participants were randomly assigned to groups we would expect only a small number of simulations to detect differences between the groups (i.e., a Type I error; a significant effect when none exists). For all 2,000 simulated samples, all applicable statistical comparisons were made with conventional parametric statistical tests, non-parametric statistical tests, and a GEE.

For this study, we used the discounting data from Hayashi et al. (2018). In this study about texting while driving, participants were asked to report their likelihood of waiting until they arrived at their destination to reply to a text message in a variety of scenarios. In the scenarios, the delay to arriving at their destination and the probability of being in a car crash were systematically varied. Participants were divided into two groups based on their self-reported likelihood of texting while driving. This is an interesting data set from an analytical standpoint because this data set has a grouping factor and two additional independent variables that were always presented together in each pairwise combination of those variables.

We made two predictions with this study. The first prediction was that the conventional statistical tests and the GEE will produce similar patterns of results. The conventional statistical tests and the GEE rely on similar assumptions and statistical models. Therefore, it is reasonable to assume that the different approaches will provide a similar pattern of results. As described above, the GEE is more appropriate for discounting data because it relatively more robust to deviations in the standard assumptions of the tests and it was designed to account for clustering within the data (i.e., data from a single participant). In other words, the GEE will produce a similar pattern of results but is a more appropriate statistical test for discounting data. We also predicted that when the simulation ignored the experimental group membership (i.e., simulating a null-condition by random assignment of participant to the groups), the GEE would be less likely to find significant group differences. In other words, when both groups are sampled from the same population, then the null hypothesis (i.e., there are no differences between the groups) is true. Therefore, any significant group differences when the simulation randomly assigns participants to a group will be a Type I error (false-positive).

# Method

## Data

The data for this study are from Hayashi et al. (2018). In that study, scenarios were posed to participants in which they were driving and they received a text message. Participants were asked to indicate with a visual analogue scale how likely they were to wait until they reached their destination to respond to the text message. Across the choice scenarios, the time until the participant reached their destination and the probability of being in a car crash were systematically varied. The social reinforcing component of replying to a text message and the probabilistic negative consequences of a car crash were the outcomes that were discounted. The likelihoods of responding to a text message that each participant reported, across each combination of delay and odds against a car crash, were treated as the indifference points. The authors found that a theoretical model of discounting (e.g., Rachlin, 2006) provided a good description of the data.

In this data set, the two outcomes that were discounted were the delay to arriving at the destination and the probability of being in a car crash. The delays to the destination were 30 seconds, 3 minutes, 15 minutes, 1 hour, and 3 hours. The probabilities shown to the participants of a car crash were 10%, 1%, 0.3%, 0.1%, and 0.03%. As is convention, the probability factor was analyzed as odds against the car crash (e.g., the likelihood of arriving at the destination safely). Participants rated their likelihood of responding to a text message for each of the 25 pairwise combinations of the delay to the destination and the probability of a car crash. Figure 1 displays set of indifference points with the associated delays to the destination and odds against a car crash.

For the purpose of this study, participants from Hayashi et al. (2018) were dichotomized into likely to text while driving (high-TWD) and less likely to text while driving (low-TWD) with a 4-question survey that assessed how often they engaged in texting while driving behaviors. Although it is problematic to dichotomize data based on a variable that is by nature on an interval or ratio scale (Young, 2016), our goal is not to extend the claims nor make new claims based on that data in Hayashi et al. (2018) but to compare statistical techniques. We chose to dichotomize the data from Hayashi et al. (2018) because the data set provides interesting analytic challenges when there are discrete groups. Specifically, by creating groups through dichotomization the data set has two independent variables that always interact (delay and odds against a car crash) across two separate groups. Therefore, the dichotomization was made to highlight the strengths and weaknesses of both the conventional statistical tests and the GEE.

## Analyses

Conventional parametric statistical tests, conventional non-parametric statistical tests, and a GEE were used to compare differences in the degree of discounting. With each statistical analysis, only certain comparisons were made with the Hayashi et al. (2018). The comparisons were selected based on our presumption that other researchers would be interested in these questions if they had discounting data of a similar structure. Table 1 includes the list of comparisons in the analyses of the simulated samples in this study. For this specific data set the questions of interest would be: differences in the degree of discounting based on grouping in high- or low-TWD, differences in delay discounting based on the odds against a car crash, differences in probability discounting based on the delay to the destination, differences in delay discounting across the odds against a car crash by group, and differences in probability discounting across the delays to the destination by group. These represent all of the main effects and the two-way interactions between group membership and delay to the destination as well as odds against a car crash. For each statistical test, the comparison were only included in the analysis when the statistical test(s) were capable of making that comparison. For example, group by delay and group by probability interactions were not possible with the non-parametric analyses because there is no widely available non-parametric equivalent to a mixed-effects ANOVA.

To conduct the parametric and the non-parametric statistical analyses, we used AUC as our measure of discounting. For each participant, AUC for probability discounting was calculated for all five of the delays to the destination. Additionally, the AUC for delay discounting was calculated for all five of the odds against a car crash. These AUC values are displayed as the marginal figures Figure 1. For the GEE analysis, the untransformed indifference points were the data of interest. All analyses were conducted with R (R Core Team, 2018) with the addition of the RStudio integrated development environment (RStudio Team, 2016). The experimental data did not meet the assumptions for parametric analyses (Hayashi et al., 2018) nor did we check whether the simulated data sets met the assumptions for parametric analyses. Additionally for all of the analyses, the $\alpha$ was .05 and no correction was made to control for family-wise Type I error rate. Typically, a correction to the $\alpha$ is made to ensure that the family-wise Type I rate is not inflated by multiple testing. The $\alpha$ was left uncorrected to informally assess the relative robustness of each type of analysis violations of the assumptions of those tests.

**Parametric analyses.—**The goal of the parametric analyses was to duplicate a mixed-effects ANOVA with the relevant post-hoc tests. In lieu of ANOVAs on AUC, we conducted linear mixed-effects regression models on AUC (lme function; no random effects specified). The mixed-effects models were selected because the post-hoc tests of interest were burdensome to conduct when using the built in R functions for a mixed-effects ANOVA. The key difference between a mixed-effect regression model and a mixed-effects ANOVA is that the regression model provided results in terms of differences in AUC from an arbitrary reference (i.e., an intercept parameter) and the ANOVA provided results in terms of differences from the mean of AUC. As a supplemental step, we used the R “anova” function to convert the linear mixed-effects model results into mixed-effects ANOVA results.

Two separate models were specified for the parametric tests: one model for AUC of delay discounting and one model for AUC of probability discounting. The model for AUC of delay discounting was specified to include the main effects of group, probability, and the interaction of group and probability. The model for AUC of probability discounting was specified to include the main effects of group, delay, and the interaction of group and delay. For each mixed-effects model, post-hoc comparisons were completed with the “emmeans” and “contrast” functions (emmeans package; Lenth, 2018).

**Non-parametric analyses.—**The non-parametric statistical testing was relatively limited compared to the parametric tests and the GEE, because there is no non-parametric equivalent for mixed-effects models. For this reason, only the results of non-parametric post-hoc tests can be directly compared to the results of the parametric and GEE post-hoc tests. The post-hoc tests for the non-parametric analyses were Mann-Whitney U tests (“wilcox.test” function with the paired parameter set to false). The non-parametric statistical tests were included in this study because the data from Hayashi et al. (2018) were not normally distributed and not amenable to transformation. Due to that fact, the authors reported non-parametric statistical tests for their statistical analyses.

We also conducted four Friedman’s tests to provide information that is related to the parametric tests’ and GEE’s main effects of delay on probability discounting and probability on delay discounting. A Friedman’s test (“friedman.test” function) is the non-parametric equivalent of a repeated-measures ANOVA. The Friedman’s tests can only provide the main effects within each group. Whereas, the main effects in the parametric tests and the GEE collapse the data across the groups. Therefore the results of the Friedman’s provide a limited amount of similar—but not identical—information as the main effects in the parametric tests and the GEE. One Friedman’s test per group was used to determine in if there were differences in delay discounting across the odds against a car crash. Two additional Friedman’s tests were used to determine there were differences in probability discounting across the delays to the destination.

**General Estimating Equation.—**The GEE analysis was conducted with the “geeglm” function (geepack R package; Halekoh, Højsgaard, & Yan, 2006). The GEE model was specified to predict indifference points with the main effects of group membership, delay to the destination, and odds against a car crash as well as the interactions of group membership and delay to the destination and group membership and odds against a car crash. For parameters supplied to the “geeglm” function: a Gaussian link function was selected to be in-line with the mixed-effects model, an autoregressive correlation matrix was specified, and participants were identified as clusters. Any comparisons that were not direct outputs of the model (such as the post-hoc tests) were conducted with the “emmeans” and “contrast” functions.

### Monte Carlo Simulations

The Monte Carlo simulations relied on randomly selecting participants to create a simulated sample and then conducting the relevant conventional parametric tests, non-parametric tests, and the GEE on that simulated sample. Participants were randomly selected, with replacement, via the “sample” function in R. When a participant was selected for inclusion, all of their experimentally obtained data were copied into the current simulated sample. After being selected for inclusion in a simulated sample, a new participant number was assigned to the currently selected data so that if a participant was randomly selected twice for a single simulated sample, then the GEE would treat each selection as a separate participant cluster.

Two separate Monte Carlo simulations were conducted. The first Monte Carlo simulation maintained the experimental group memberships. For these simulations, participants were randomly selected from within each group. For the second Monte Carlo simulation, participants were randomly selected with replacement and then assigned to either the high-TWD group andor the low-TWD group. After the total sample had been created by randomly selecting participants, each analysis was conducted on the simulated sample.

After the analyses on all of the simulated data sets were completed, we calculated two main measures to describe the pattern of results across the simulated data sets. These measures were calculated separately for the simulations in which group membership was maintained and in which group membership was randomly assigned. The first measure was the number of simulations in which a statistically significant effect was found. This measure was calculated for each comparison in the parametric tests, the non-parametric tests, and the GEE. For example, the number of data sets in which the GEE determined there was a statistically significant effect of delay on probability discounting was calculated. A second measure was calculated to assess the degree to which the results of the parametric statistical tests agreed with the results of the GEE. For each simulated data set, the results of the parametric analyses were compared to the resultsof the GEE. For each comparison of interest (e.g., main effects, interactions, post-hoc tests), the statistical significance of the parametric test was compared to the statistical significance of the GEE. If the statistical significance was in agreement across both the parametric test and the GEE, then that test was counted as in agreement. If one test analysis indicated a statistically significant effect for that simulated sample and the other analysis indicated a non-significant effect than that test was not counted as being in agreement. The agreement scores were then counted across all of the simulated samples for each specific comparison of interest. This measure of agreement was calculated only when participants were resampled from within the existing groups. Agreement scores were not calculated for the non-parametric tests because only the post-hoc Mann-Whitney U tests could be compared to the conventional parametric and GEE post-hoc tests.

## Results

Two sets of 1,000 samples were simulated via a Monte Carlo method resampling procedure. In the first set, samples were simulated by randomly selecting participants and including their data with the membership of each participant in either the high- or low-TWD being maintained. In the second set of simulated samples, participants were randomly selected and then assigned to either the high- or low-TWD based on the resampling procedure. To determine if the random sampling was adequate, Kullback-Leibler divergence values were calculated for both groups in each set of simulated samples. Kullback-Leibler divergence provides a metric for the difference between two distributions. The measure is bound between 0 and 1, lower values indicate more similar distributions and higher values indicate more dissimilar distributions. The obtained frequency of each participant in the simulated samples was compared to the expected uniform distribution. When group memberships were maintained, the divergence of the high-TWD samples was 0.0004 and the divergence of the low-TWD samples was 0.0006. When participants were randomly assigned to groups, the divergence for the high-TWD samples was 0.0009 and the divergence of the low-TWD samples was 0.0014. Overall, these low divergence values indicate that participants were almost uniformly distributed across the groups in the simulated samples.

## Effect of group membership on degree of discounting

Only the GEE was capable of providing a single statistical comparison to determine if there was an overall difference in discounting based on whether a person was more likely to text while driving (high-TWD) or less likely (low-TWD). In the experimentally obtained data set, there were significant differences in indifference points based on group membership ($\chi^2 = 276.2$, $p < .001$). For the specific differences across the groups, we compared the marginal indifference points. Based on the 0–100% likelihood of responding to a text message, the scale that was used in the study, the low-TWD group was 27.3% more likely to wait until arriving at the destination to respond to a text message than the high-TWD group across all the delays to the destination and likelihoods against a car crash (Z = 17. 5 $p < .001$). In other words, the average indifference point for the low-TWD group was 27.3 lower than the average indifference point for the high-TWD group.

The Monte Carlo simulations allow for the assessment of how likely the results obtained from the experimental data sets were due to a chance sampling. When participants were resampled from within their respective experimental groups, there was a significant difference in the average indifference points between the groups 988 times. When participants were randomly assigned to the experimental groups, there was a significant difference in the average indifference points between the groups 54 times.

## Effect of delay on probability discounting

The effect of delay on probability discounting across both groups combined were assessed statistically with the GEE and the conventional parametric tests. The non-parametric Friedman's tests were also conducted with each group, which—as described above—do not provide the exact same information as the conventional parametric and GEE tests which collapse the data across the groups. For the experimentally obtained data, with the GEE there was a significant effect of delay on indifference points for probability discounting ($\chi^2 = 131.2$ $p < .001$). For the conventional parametric tests, there were significant differences in AUC for probability discounting across the delays to the destination (F = 25. 8, $p < .001$). For the non-parametric tests, there were also significant differences in AUC for probability discounting across the delays to the destination for both the high-TWD group (Q = 131.1, *p*

< .001) and the low-TWD group (Q = 126.9, $p < .001$). In summary, all three statistical comparisons indicated that delay to the destination significantly affected the degree of probability disocunting.

The Monte Carlo simulation provided measures of how likely each statistical comparison would lead to a significant effect as well as the how often the GEE and the conventional parametric statistical tests agreed. When participants were selected from within groups, there was a significant effect of delay to the destination on the degree of probability discounting 992 times for the GEE, 1,000 times for the conventional parametric test, as well as 1,000 times for non-parametric tests for both the high- and low-TWD groups. The pattern of results obtained from the GEE was the same as the pattern obtained with the conventional parametric tests 992 times, indicating that the GEE and conventional parametric test were providing similar results. When participants were randomly assigned to groups, there was a significant effect of delay to the destination on the degree of probability discounting 994 times for the GEE, 1,000 times for the conventional statistical test, as well as 1,000 times for both the non-parametric tests for both the high- and low- TWD group. Across both sets of simulations, all three statistical techniques produced similar results.

## Effect of probability on delay discounting

The effect of odds against a car crash on the degree of delay discounting shared identical features (with different variables) to the analyses described above. Therefore, the nature of the statistical comparisons was the same. In the experimentally obtained data, in the GEE there was a significant effect of odds against a car crash on indifference points for delay discounting ($\chi^2 = 4935$, $p < .001$). For the conventional parametric tests, there were significant differences in AUC for delay discounting across the odds against a car crash (F = 30. 2, $p < .001$). For the non-parametric tests, there were also significant differences in the AUC for delay discounting across the odds against a car crash for both the high-TWD group (Q = 144. 6, $p < .001$) and the low-TWD group (Q = 119.7, $p < .001$). In summary, all three statistical comparisons indicated that odds against a car crash significanly affected the degree of delay disocunting.

The Monte Carlo simulation provided measures of how likely each statistical comparison would lead to a significant effect as well as the how often the GEE and the conventional parametric statistical tests agreed. When participants were selected from within the groups, all four statistical tests found significant effects for all 1,000 simulated samples. The pattern of results obtained from the GEE was the same as the pattern of results obtained with the conventional parametric for each of the 1,000 simulated samples. When participants were randomly assigned to groups the comparisons were statistically significant for the GEE, the conventional parametric test, and the non-parametric tests for all 1,000 simulated samples.

## Interaction of group and delay on probability discounting

The interaction of whether a person was more likely to text while driving or less likely to text while driving (i.e., group membership) and the effect of delay to the destination on probability discounting could be assessed statistically with the GEE and the conventional parametric statistics. For the experimental data set, with the GEE there were differences across the groups in the effect that delay to the destination had on indifference points (i.e., a significant interaction; $\chi^2 = 26.4$ $p < .001$). For the conventional paramteric tests, the interaction of group membership and delay to the destination on AUC for probability discounting was statistically signiciant ($F = 3.58$, $p = .008$).

The Monte Carlo was again used to determine the likelihood of a significant interaction of group membership and the delay to the destination as well as the degree of agreement between the GEE and the conventional parametric tests. When participants were selected from within groups, there was a statistically significant interaction detected with the GEE 369 times and a significant interaction detected with the conventional parametric test 708 times. The Monte Carlo simulation revealed that the GEE and the conventional statistical test only agreed 383 times. When particpants were randomly assigned to the groups, there was a signiciant interaction detected only 66 times with the GEE and only 93 times with the conventional statistcal test.

## Interaction of group and probaiblity on delay discounting

The interaction of whether a person was more likely to text while driving or less likely to text while driving (i.e., group membership) and the effect of odds against a car crash on delay discounting could be assessed statistically with the GEE and the conventional parametric statistics. For the experimentally obtained data, with the GEE there were significant differences across the groups in the effect that odds against a car crash had on indifference points (i.e., a significant interaction; $\chi^2 = 26.4$, $p =.002$). With the conventional parametric tests, the interaction of group membership and the odds against a car crash on AUC for degree of delay discounting was statistically significant ($F = 4.02$, $p = .003$).

The Monte Carlo simulations were used to derived the same measures described above. When participants were selected from within groups, there was a statistically significant interaction detected 971 times with the GEE and 724 imes with the convetional statistical tests. The GEE and conventional parametric tests agreed for 745 of the 1,000 simulated samples. When participants were randomly assigned to the groups, there was a significant interaction detected only 78 times with the GEE and only 88 times with the conventional statistical test.

## Post-hoc: Differences in probability discounting between groups at each level of delay

The post-hoc comparisons of the degree of probability discounting between the groups could be examined with the GEE, the conventional statistical tests, and the non-parametric tests. Table 2 has the number of simulations in which a significant difference was found, for the specified between group comparison, and for all three of the analyses. In the experimentally obtained data set, there were significant differences between the groups in the degree of probability discounting across all 5 of the delays to the destination across each analysis. When participants were selected from within groups, the simulated samples frequently had statistically significant differences in line with the experimentally obtained data set. Only at the shortest delay to the destination (30 seconds) did the Monte Carlo procedure fail to reliably to produce samples in which a statistically significant difference between the groups existed. For all 5,000 post-hoc comparisons across the 1,000 simulated samples, the GEE and the conventional parametric tests agreed 4,640 times. The fewest number of agreements for a single post-hoc test was 782 and those agreements were associated with the degree of probability discounting at the shortest delay to the destination. In the simulations in which the participants were randomly assigned to groups, none of the analyses could reliably find significant differences between the groups in the post-hoc tests.

### Post-hoc: Differences in delay discounting between groups at each level of probability

Table 3 has the results Monte Carlo simulations for the post-hoc tests to determine if there were differences in the degree of delay discuonting between groups at each level of odds against a car crash. In the experimentally obtained data set, there were significant differences in delay discounting between the groups at each level of odds against a car crash. When participants were resampled from within groups, the simulated samples frequently had statistically significant differences in line with the experimentally obtained data set. In the 15 possible post-hoc comparions conducted, 14 of the comparisons had 900 or more simulated samples in which a statistically significant difference in the degree of delay discounting was detected. The post-hoc comparison of delay discounting when the odds against a car crash were 9 had the lowest number of samples in which a statistically signicant result was found with only 754 samples. For all 5,000 post-hoc comparisons across the 1,000 simulated samples, the GEE and the conventional parametric tests agreed 4,767 times. The fewest number of aggrements was 816 and those agreements were associated with the degree of probability discounting at the lowest odds against a car crash. In the simulations in which the participants were randomly assigned to the groups, none of the analyses could reliably find significant differences between the groups in the post-hoc tests.

## Discussion

In this study, we compared the effectiveness of three different statistical approaches in describing discounting data using Monte Carlo simulations. The Monte Carlo simulations were used to create data sets from an experimentally obtained data set. Participants were randomly selected with replacement and placed into simulated data sets. Statistical analyses were then conducted on the simulated data sets. In one set of simulations, participants were only randomly selected for a group if they were members of that group in the experimental data set. In the second set of simulations, participants were randomly assigned to a group. Across the three analysis types, the Monte Carlo simulations found a pattern of results that was generally similar to what was obtained with the experimental data set. Where it was possible to assess the agreement between significant effects in the GEE analysis and significant effects in the conventional parametric tests, the two types of analysis were largely in agreement. Finally, when participants were randomly assigned to groups (i.e., an imposed null-effect of group membership) the analyses found significant effects in only a small number of the simulated samples which was inline with the expected Type I error rate of .05.

The GEE provided a similar pattern of results as conventional mixed-effects ANOVAs. Across the 1,000 simulations in which participants were resampled from within their respective experimental groups, the GEE and the conventional parametric statistical tests had a high degree of agreement in terms of whether or not statistically significant effects were detected in the simulated samples. The GEE and conventional tests only had a low degree of agreement for the comparison that was testing for an interaction between group membership and delay to the destination on the degree of probability discounting. Additionally, the experimental data were not normally distributed and it is reasonable to assume that the majority of the simulated samples did not have normally distributed data. Despite this violation, both the GEE and the conventional parametric tests reliably found statistical differences across the simulated samples and there was a high degree of agreement in the pattern of results found by the tests. It is therefore reasonable to conclude that the GEE is in line with more conventionally used statistical tests and can be a useful and reliable tool for NHST with discounting data (e.g., Friedel et al., 2016).

The results of the GEE and the conventional statistical tests diverged when an interaction between group membership and delay to the destination on the degree of probabilty was tested. There are several possible reasons for this pattern of disagreement. It is possible that for this specific comparison, the true difference between the groups was very small so the simulation procedure could not reliably produce samples in which a difference existed. If such a scenario occurred, the GEE would likely be more conservative than the conventional mixed-effects ANOVA because the GEE is accounting for the entirety of the simulated data set all at once while the mixed-effects ANOVA only accounts for some features of the data.

An alternative explanation for this discrepancy might lie in the distributions of the obtained indifference points and the derived AUC. Hayashi et al. (2018) used non-parametric statistical analyses because the data were not amenable to parametric statistical tests. However, in this re-analysis we ignored whether the data met the relevant assumptions of the statistical test we were using. It is possible that the extreme deviations from normality in the data (see Appendix) were the cause of this discrepancy between the findings of the GEE and the conventional parametric tests. That is, with the other comparisons there were enough data points that both the GEE and conventional tests were relatively robust to the violation of normality but for this specific test the deviations from normality were too great and the tests broke down. A future study could use a Monte Carlo method to create delay discounting data with varying degrees of divergence from normality to determine the limits of when these statistical techniques are and are not apporpriate. Whatever the cause of the discrepancy for this interaction, the results of the Monte Carlo simulations largely indicate that the GEE and the parametric statistical provide a similar pattern of results.

Condsidering that GEE appears to produce patterns of results that are generally similar to more conventional parametric tests it is important to note that the GEE has several properties that make it potentially superior to the conventional mixed-effects ANOVA and non-parametric statistics. First, a GEE can be easily be conducted on the obtained indifference points instead of a derived measure like AUC or an output from model fitting. This allows the GEE to compare differences across measures of interest (e.g., group membership) without making any underlying assumptions about the nature of the data. There are two key benefits of conducting an analysis on the indifference points instead of on a summary measure or the output from model fitting. The first benefit is that the indifference points provide a richer data source. When calculating AUC or a model output, several indifference points are collapsed into a single measure. By collapsing the data, both degrees of freedom for later analyses and variability within that collapsed data is lost. Any time degrees of freedom or data richness are lost future comparisons become more difficult. For example, imagine a scenario in which a researcher had two subjects and obtained seven indifference points per subject. It would not be possible to compare the degree of discounting across the subjects using AUC or an output from a model because the seven indifference points have been collapsed into a single data point per subject. As the indifference points are a richer data source it would be possible to use a GEE to compare the degree of discounting between those subjects, although in this extreme case researchers would not typically compare the differences between two subjects. Relying on a richer data source is always a better analytic strategy.

A second benefit of GEE is that it can, potentially, abate a common problem in the characteristics of obtained AUC values or model outputs. There are serveral well known limitations of using AUC or the results of model fits to analyze delay discounting data (Borges, Kuang, Milhorn, & Yi, 2016; Gilroy & Hantula, 2018; Yoon et al., 2018). In relation to conventional parametric testing, one of the largest concerns is that often times delay discounting data are not normally distributed, which is often a requirement for many of the conventional parametric tests that we rely on. To address the non-normal data, it is popular to log transform the parameters from the model (i.e., ln(k); Petry, 2001) or use a version of AUC in which delays are scaled logarithmically instead of arithmetically (AUClog; Borges et al., 2016). Frequently, log transformed discounting data has lead to more normally distributed data which facilitates the usage of conventional parametric statistics. However, in our personal experiences frequently ln(k) and AUClog do not lead to normally distributed data. With multivariate measures of discounting, often times data are so skewed or distributed so abnormally that a simple log transformation is not sufficient to normalize the data. For example, in both Friedel et al. (2014) and Friedel et al. (2016) neither ED-50 nor AUClog could sufficiently adjust the distribution of the obtained data to make conventional mixed-effects ANOVAs appropriate. As the GEE fits into the broad class of generalized linear models, it can often be robust to errors introduced by deviations from the expected distribution of the data (e.g., non-normal data). Additionally, as a generalized linear model, the researcher can specify error distributions beyond just the simple Gaussian which may be more in line with the structure of the underlying data. It should be noted the error distribution can be specified with any generalized linear model and that it is not a special feature of GEE-based analyses.

Finally, one limitation of analyses that rely on categorical comparisons (i.e., *t*-tests, ANOVAs, etc) is that often times researchers must create discrete categories where none naturally exist (Young, 2016). Any time a researcher imposes order on data they are creating fractures instead of finding the natural lines of fracture (e.g., Skinner, 1935). For this study, we created experimental groups by taking data from a Likert scale for how frequently a participant responded to text messages while driving and dichotimized the data based on a median split. While this is defensible for creating a complex data set with which to challenge different analytic techniques, this sort of median split is forcing the data to fit a certain pattern. Regression based techniques, such as the GEE, have a distinct advantage because they can be used to analyze both categorical and interval/ratio data. This means that instead of creating groups via a median-split the raw Likert scores of frequency of texting while driving could have been included in the statistical model. As the Likert scales carry more information than a group categorization, a more complex or subtle analysis could be conducted. In other words, often times creating categories allows us to anwer whether or not a difference exists based on some variable whereas using data that naturally has a range of values allows us to potentially determine the nature of the relation between variables. Generalized estimating equations, as well as other regression techniques, allow us to ask these sorts of questions when our data allow.

This study relied on a Monte Carlo method to simulate data sets so that many “aritificial experiments” could be conducted to compare different types of analyses. This study provides further evidence that Monte Carlo methods are a tool that should be more commonly used by behavior analysts. Monte Carlo methods can be used to answer a wide variety of quantitative questions. For example Friedel, Galizio, Berry, Sweeney, and Odum (under review), used a Monte Carlo method to establish whether organisms across multiple experiments reliably demonstrated relapse behaviors or whether the organisms happen to engage in higher rates of behavior during relapse testing. In an unrelated study, Friedel, DeHart, and Odum (2017) used Monte Carlo methods to determine whether relatively small but consistent decreases in the rate of response during probe sessions were reliably caused the response dependent 100 dB tones during those sessions. As Monte Carlo methods rely on resampling of data, the exact nature of the question researchers ask is open ended. These methods should be considered a tool that does not rely on underlying assumptions that, when properly used, can provide some measure of confidence about different research questions.

In conclusion, there are many potential benefits to analyzing discounting data with a GEE. Often times, with discounting data more conventional and simpler statistical tests are appropriate. However, as behavior analysts begin to ask more complex questions which lead to extremly complex data structures it is important that we use the proper statistical tools for the data at hand. The GEE might not be the best or most appropriate statistical test in every situation, but we encourage authors to be as careful with their decision-making about statistical analyses as they are with decision-making about experimental designs.

## Acknowledgments

Authors’ note: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention. The authors would like to thank Amy Odum and Oliver Wirth for providing comments on an earlier version of this manuscript and Yusuke Hayashi for thoughtful conversations in developing this study. Correspondence concerning this article should be addressed to Jonathan Friedel.

## Appendix

Figure A1 displays the distributions of all indifference points, AUC for delay discounting, and AUC for probability discounting. The distributions are noticeable not normally distributed, which was one of the reasons Hayashi et al. (2018) chose to conduct non-parametric statistics. The data are most likely best described by a beta distribution with both $\alpha$ and $\beta$ being less than 1. As determining the correct distribution of the data was not necessary for this study, no formal attempts were made to identify the correct distribution.

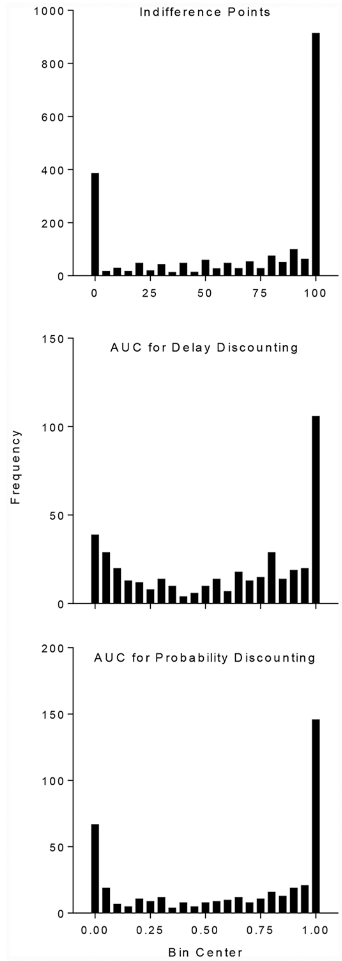


**Figure A1.**
Histograms of all indifference points, AUC for delay discounting, and AUC for probability discounting from Hayashi et al. (2018).

## References

Barkley RA, Edwards G, Laneri M, Fletcher K, & Metevia L (2001). Executive functioning, temporal discounting, and sense of time in adolescents with attention deficit hyperactivity disorder (adhd) and oppositional defiant disorder (odd). Journal of Abnormal Child Psychology, 29(6), 541–556. doi: 10.1023/A:1012233310098 [PubMed: 11761287]

Bickel WK, Jarmolowicz DP, Mueller ET, Koffarnus MN, & Gatchalian KM (2012). Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: Emerging evidence. Pharmacology and Therapeutics, 134(3), 287–297. doi:10.1016/j.pharmthera.2012.02.004 [PubMed: 22387232]

Bickel WK, Odum AL, & Madden GJ (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. Psychopharmacology, 146(4), 447–454. doi: 10.1007/PL00005490 [PubMed: 10550495]

Borges AM, Kuang J, Milhorn H, & Yi R (2016). An alternative approach to calculating area-under-the-curve (auc) in delay discounting research. Journal of the Experimental Analysis of Behavior, 106(2), 145–155. doi:10.1002/jeab.219 [PubMed: 27566660]

Charlton SR, & Fantino E (2008). Commodity specific rates of temporal discounting: Does metabolic function underlie differences in rates of discounting? Behavioural Processes, 77(3), 334–342. doi: 10.1016/j.beproc.2007.08.002 [PubMed: 17919848]

Dixon MR, Marley J, & Jacobs EA (2003). Delay discounting by pathological gamblers. Journal of Applied Behavior Analysis, 36(4), 449–458. doi:10.1901/jaba.2003.36-449 [PubMed: 14768665]

Franck CT, Koffarnus MN, House LL, & Bickel WK (2015). Accurate characterization of delay discounting: A multiple model approach using approximate bayesian model selection and a unified discounting measure. Journal of the Experimental Analysis of Behavior, 103(1), 218–233. doi: 10.1002/jeab.128 [PubMed: 25556903]

Friedel JE, DeHart WB, Frye CC, Rung JM, & Odum AL (2016). Discounting of qualitatively different delayed health outcomes in current and never smokers. Experimental and Clinical Psychopharmacology, 24(1), 18–29. doi:10.1037/pha0000062 [PubMed: 26691848]

Friedel JE, DeHart WB, Madden GJ, & Odum AL (2014). Impulsivity and cigarette smoking: Discounting of monetary and consumable outcomes in current and non-smokers. Psychopharmacology, 231(23), 4517–4526. doi:10.1007/s00213-014-3597-z [PubMed: 24819731]

Friedel JE, DeHart WB, & Odum AL (2017). The effects of 100 dB 1-kHz and 22-kHz tones as punishers on lever pressing in rats. Journal of the Experimental Analysis of Behavior, 107(3), 354–368. doi:10.1002/jeab.254 [PubMed: 28453234]

Friedel JE, Galizio A, Berry MS, Sweeney MM, & Odum AL (under review). An alternative approach to relapse analysis: Using Monte Carlo methods and proportional rates of response. Journal of the Experimental Analysis of Behavior.

Gilroy SP, & Hantula DA (2018). Discounting model selection with area-based measures: A case for numerical integration. Journal of the Experimental Analysis of Behavior, 109(2), 433–449. doi: 10.1002/jeab.318 [PubMed: 29498424]

Green L, Fry AF, & Myerson J (1994). Discounting of delayed rewards: A life-span comparison. Psychological Science, 5(1), 33.

Halekoh U, Højsgaard S, & Yan J (2006). The R package geepack for generalized estimating equations. Journal of Statistical Software, 15(2). doi:10.18637/jss.v015.i02

Hanley JA, Negassa A, Edwardes M. D. d., & Forrester JE (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. American Journal of Epidemiology, 157(4), 364–375. doi:10.1093/aje/kwf215 [PubMed: 12578807]

Hayashi Y, Fessler HJ, Friedel JE, Foreman AM, & Wirth O (2018). The roles of delay and probability discounting in texting while driving: Toward the development of a translational scientific program. Journal of the Experimental Analysis of Behavior. doi:10.1002/jeab.460

Kroese DP, Brereton T, Taimre T, & Botev ZI (2014). Why the Monte Carlo method is so important today. Wiley Interdisciplinary Reviews: Computational Statistics, 6(6), 386–392. doi:10.1002/wics.1314

Lenth R (2018). emmean: Estimated marginal means, aka least-squares means. Retrieved from https://CRAN.R-project.org/package=emmeans

Mazur JE (1987). An adjusting procedure for studying delayed reinforcement In Commons ML, Mazur JE, Nevin JA, & Rachlin H (Eds.), Quantitative analysis of behavior: Vol. 5. The effect of delay and intervening events on reinforcement value. (pp. 55–73). Hillsdale: Earlbaum.

Mitchell SH (1999). Measures of impulsivity in cigarette smokers and non-smokers. Psychopharmacology, 146(4), 455–464. doi:10.1007/PL00005491 [PubMed: 10550496]

Myerson J, & Green L (1995). Discounting of delayed rewards: Models of individual choice. Journal of the Experimental Analysis of Behavior, 64(3), 263–276. doi:10.1901/jeab.1995.64-263 [PubMed: 16812772]

Myerson J, Green L, & Warusawitharana M (2001). Area under the curve as a measure of discounting. Journal of the Experimental Analysis of Behavior, 76(2), 235–243. doi:10.1901/jeab.2001.76-235 [PubMed: 11599641]

Odum AL (2011a). Delay discounting: I'm a k, you're a k. Journal of the Experimental Analysis of Behavior, 96(3), 427–439. doi:10.1901/jeab.2011.96-423 [PubMed: 22084499]

Odum AL (2011b). Delay discounting: Trait variable? Behavioural Processes, 87(1), 1–9. doi:10.1016/j.beproc.2011.02.007 [PubMed: 21385637]

Odum AL, Baumann AA, & Rimington DD (2006). Discounting of delayed hypothetical money and food: Effects of amount. Behavioural Processes, 73(3), 278–284. doi:10.1016/j.beproc.2006.06.008 [PubMed: 16926071]

Petry NM (2001). Delay discounting of money and alcohol in actively using alcoholics, currently abstinent alcoholics, and controls. Psychopharmacology, 154(3), 243–250. doi:10.1007/s002130000638 [PubMed: 11351931]

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foudnation for Statistical Computing Retrieved from https://www.R-project.org/

Rachlin H (2006). Notes on discounting. Journal of the Experimental Analysis of Behavior, 85(3), 425–435. doi:10.1901/jeab.2006.85-05 [PubMed: 16776060]

Rasmussen EB, Lawyer SR, & Reilly W (2010). Percent body fat is related to delay and probability discounting for food in humans. Behavioural Processes, 83(1), 23–30. doi:10.1016/j.beproc.2009.09.001 [PubMed: 19744547]

Team RStudio. (2016). Rstudio: Integrated development environment for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Skinner BF (1935). The generic nature of the concepts of stimulus and response. The Journal of General Psychology, 12(1), 40–65. doi:10.1080/00221309.1935.9920087

Wasserstein RL, & Lazar NA (2016). The asa's statement on p-values: Context, process, and purpose. The American Statistician, 70(2), 129–133. doi:10.1080/00031305.2016.1154108

Yoon JH, & Higgins ST (2008). Turning k on its head: Comments on use of an ED50 in delay discounting research. Drug and Alcohol Dependence, 95(1–2), 169–172. doi:10.1016/j.drugalcdep.2007.12.011 [PubMed: 18243583]

Yoon JH, Weaver MT, De La Garza R, 2nd, Suchting R, Nerumalla CS, Omar Y, … Newton TF (2018). Comparison of three measurement models of discounting among individuals with methamphetamine use disorder. American Journal on Addictions. doi:10.1111/ajad.12761

Young ME (2016). The problem with categorical thinking by psychologists. Behavioural Processes, 123, 43–53. doi:10.1016/j.beproc.2015.09.009 [PubMed: 26424490]

| | | Odds against a car crash | | | | | Probability |
|---|---|---|---|---|---|---|---|
| | | 9 | 99 | 299 | 999 | 2999 | discounting AUC |
| Delay to destination (minutes) | 0.5 | $V_{0.5,9}$ | $V_{0.5,99}$ | $V_{0.5,299}$ | $V_{0.5,999}$ | $V_{0.5,2999}$ | $AUC_{0.5}$ |
| | 3 | $V_{3,9}$ | $V_{3,99}$ | $V_{3,299}$ | $V_{3,999}$ | $V_{3,2999}$ | $AUC_{3}$ |
| | 15 | $V_{15,9}$ | $V_{15,99}$ | $V_{15,299}$ | $V_{15,999}$ | $V_{15,2999}$ | $AUC_{15}$ |
| | 60 | $V_{60,9}$ | $V_{60,99}$ | $V_{60,299}$ | $V_{60,999}$ | $V_{60,2999}$ | $AUC_{60}$ |
| | 180 | $V_{180,9}$ | $V_{180,99}$ | $V_{180,299}$ | $V_{180,999}$ | $V_{180,2999}$ | $AUC_{180}$ |
| Delay discounting AUC | | $AUC_{9}$ | $AUC_{99}$ | $AUC_{299}$ | $AUC_{999}$ | $AUC_{2999}$ | |

**Figure 1.**

Indifference points and AUC values from Hayashi et al. (2018). The marginal values are the AUC values that would be calculated from each row and column of indifference points.

**Table 1.**

List of statistical comparisons included in the Monte Carlo Simulations

| Analysis | Conv. | Non-Par | GEE |
|---|---|---|---|
| Main effect of group on discounting | - | - | X |
| Effect of delay on probability discounting | * | X | X |
| Effect of probability on delay discounting | * | X | X |
| Interaction of group and delay on probability discounting | - | X | X |
| Interaction of group and probability on delay discounting | - | X | X |
| Post-hoc: Differences in probability discounting between groups at each level of delay | X | X | X |
| Post-hoc: Differences in delay discounting between groups at each level of probability | X | X | X |

**Table 2.**

Counts of significant post-hoc for between group differences in probability discounting at each delay to the destination

| | Within group resampling | | | Random assignment | | |
|---|---|---|---|---|---|---|
| **Delay to the destination** | **Conv.** | **Non-par.** | **GEE** | **Conv.** | **Non-par.** | **GEE** |
| 30 seconds | 636 * | 627 * | 468 * | 45 | 44 | 67 |
| 3 minutes | 920 * | 984 * | 877 * | 44 | 51 | 54 |
| 15 minutes | 999 * | 1000 * | 966 * | 44 | 49 | 52 |
| 1 hour | 998 * | 998 * | 967 * | 42 | 43 | 55 |
| 3 hours | 997 * | 998 * | 996 * | 41 | 34 | 43 |

Note:

* indicate that a statistically significant difference between the groups was found in the experimentally obtained data ($p$ <.5).

**Table 3.**

Counts of significant post-hoc for between group differences in delay discounting across each odds of a car crash

| | Within group resampling | | | Random assignment | | |
|---|---|---|---|---|---|---|
| **Odds against a car crash** | **Conv.** | **Non-par.** | **GEE** | **Conv.** | **Non-par.** | **GEE** |
| 9 | 910* | 932* | 754* | 36 | 55 | 58 |
| 99 | 982* | 973* | 963* | 41 | 43 | 51 |
| 299 | 997* | 993* | 988* | 41 | 39 | 53 |
| 999 | 998* | 998* | 991* | 39 | 35 | 47 |
| 2999 | 998* | 998* | 998* | 51 | 39 | 48 |

Note:

* indicate that a statistically significant difference between the groups was found in the experimentally obtained data ($p$ <.05).