# Improving prediction of eating-related behavioral outcomes with zero-sensitive regression models

Katherine Schaumberg[a,*], Erin E. Reilly[b,c], Lisa M. Anderson[b,d], Sasha Gorrell[b], Shirley B. Wang[e], Margarita Sala[f]

[a] *Department of Psychiatry, University of North Carolina School of Medicine, USA*
[b] *Department of Psychology, University at Albany – State University of New York, USA*
[c] *Department of Psychiatry, University of California – San Diego, USA*
[d] *Department of Psychiatry, University of Minnesota, USA*
[e] *Department of Psychology, Harvard University, USA*
[f] *Department of Psychology, Southern Methodist University, USA*

ABSTRACT

*Objective:* Outcome variables gauging the frequency of specific disordered eating behaviors (e.g., binge eating, vomiting) are common in the study of eating and health behaviors. The nature of such data presents several analytical challenges, which may be best addressed through the application of underutilized statistical approaches. While zero-sensitive models are well-supported by methodologists, application of these models has yet to gain traction among a widespread audience of researchers who study eating-related behaviors. The current study examined several approaches to predicting count-based behaviors, including zero-sensitive (i.e., zero-inflated and hurdle) regression models.
*Method:* Exploration of alternative models to predict eating-related behaviors occurred in two parts. In Part 1, participants (N = 524; 54% female) completed the Eating Disorder Examination-Questionnaire and Daily Stress Inventory. We considered the theoretical basis and practical utility of several alternative approaches for predicting the frequency of binge eating and compensatory behaviors, including ordinary least squares (OLS), logistic, Poisson, negative binomial, and zero-sensitive models. In Part 2, we completed Monte Carlo simulations comparing negative binomial, zero-inflated negative binomial, and negative binomial hurdle models to further explore when these models are most useful.
*Results:* Traditional OLS regression models were generally a poor fit for the data structure. Zero-sensitive models, which are not limited to traditional distribution assumptions, were preferable for predicting count-based outcomes. In the data presented, zero-sensitive models were useful in modeling behaviors that were relatively rare (laxative use and vomiting, 9.7% endorsed) along with those that were somewhat common (binge eating, 33.4% endorsed; driven exercise, 40.7% endorsed). Simulations indicated missing data, sample size, and the number of zeros may impact model fit.
*Discussion:* Zero-sensitive approaches hold promise for answering key questions about the presence and frequency of common eating-related behaviors and improving the specificity of relevant statistical models. The current manuscript provides practical guidance to aid the use of these models when studying eating-related behaviors.

Eating disorders are pervasive psychiatric conditions that have detrimental psychological and physical health outcomes (Mitchell & Crow, 2006). Recurrent episodes of binge eating and compensatory behavior use, including self-induced vomiting, laxative use, and compensatory exercise, are considered hallmark characteristics of many eating disorders (American Psychiatric Association, 2013) and are associated with marked psychosocial impairment (Spoor, Stice, Burton, & Bohon, 2007). The presence and frequency of binge eating and compensatory behaviors occurs on a continuum across both clinical and non-clinical populations. Depending on the population studied and the

---

**Table 1**
Comparison of regression-based models.

| | Assumptions | Benefits | Drawbacks |
|---|---|---|---|
| Ordinary Least Squares (OLS) Regression | • Outcome variable normally distributed<br>• Independence<br>• Homoscedasticity<br>• Continuous outcomes | • Familiar to most researchers<br>• Relatively easy to use<br>• Can be used with continuous, non-count variables | • Normality and homoscedasticity assumptions are rarely met<br>• Violations of normality and homoscedasticity can distort Type I and Type II error rates and reduce power<br>• Affected by outliers |
| OLS-Transformed | • Outcome variable normally distributed<br>• Independence<br>• Homoscedasticity<br>• Continuous outcomes | • Familiar to most researchers<br>• Relatively easy to use<br>• Can be used with continuous, non-count variables | • Transformation does not restore normality and homoscedasticity in all cases<br>• Outliers can remain after transforming data<br>• Difficult to interpret results due to change in scale |
| Logistic Regression | • Dichotomous outcomes<br>• Independence | • Only predicts possible probabilities<br>• Not affected by outliers | • Only appropriate for dichotomous outcomes (or those recoded to be dichotomous)<br>• Recoding variables into dichotomous outcomes may inflate Type II error<br>• Sample size must be large when outcomes are infrequent |
| Poisson Regression | • Outcome assumed to be distributed as a Poisson random variable<br>• Assumes variance is equal to the mean<br>• Continuous count outcomes | • Can be used in highly skewed distributions<br>• Appropriate for count data<br>• Appropriate when the mean count is a small value | • Selecting a Poisson model when the data are over-dispersed can result in Type I errors<br>• May not be appropriate for a large number of zeros<br>• Affected by outliers |
| Negative Binomial Regression | • Allows for independent specification of the mean and variance<br>• Continuous count outcomes | • Can be used in highly skewed distributions<br>• May be advantageous when over dispersion of outcomes occurs | • May not be appropriate for a large number of zeros<br>• Affected by outliers |
| Zero-inflated Regression | • Assumes a logistic regression model for the zero vs. non-zero portion of the outcome<br>• Assumes a Poisson or negative binomial distribution for the count portion of the model | • May be most successful in evaluating outcomes when there is a preponderance of zeros<br>• Able to maintain adequate power and Type I error control even when normality and heteroscedasticity assumptions are not met<br>• Can be used with highly skewed data | • Requires more power<br>• Affected by outliers |
| Hurdle Regression | • All zeros are structural zeros (i.e., true zeros)<br>• Assumes separate processes for zero and non-zero counts | • Appropriate when the zero portion of the model and the count portion of the model are considered to arise from discrete processes<br>• Able to maintain adequate power and Type I error control even when normality and heteroscedasticity assumptions are not met<br>• Can be used with highly skewed data<br>• Relatively easy to interpret | • Requires more power<br>• Affected by outliers |

context of a given investigation, the general distribution of these behaviors can vary greatly. Noting the degree to which the distributions of disordered eating behaviors may vary, research investigating risk and outcome related to these behaviors must carefully consider how such variables are quantified and analyzed.

## 1. Methodological considerations

When undertaking an investigation of pathological eating-related behaviors, researchers and clinicians have several choices related to the manner in which they analyze these outcomes (e.g., as continuous count data or categorically present vs. absent). As such, investigators should carefully consider the implications of these choices in informing scientific and clinical knowledge. Within eating behavior research, outcomes are commonly understood and measured as count variables. For instance, frequency of behaviors such as binge eating and purging may be used in diagnosis and as a proxy for the severity of an eating disorder. Additionally, change in the frequency of these behaviors is often used as a gauge of treatment progress (Agras et al., 2000; Jacobi, Völker, Trockel, & Taylor, 2012).

Although the evaluation of behavior frequency provides useful information, simple categorization of presence or absence of these behaviors also presents a clinically-informative indication of pathology. For instance, if one is assessing recovery from an eating disorder, full recovery may be conceptualized as a prolonged period of abstinence from compensatory behaviors. As such, researchers attempting to ascertain risk status may choose to dichotomize variables into "present" or "absent" within a specified time period (e.g., "non-binge eating" and "binge eating"; Herbozo, Schaefer, & Thompson, 2015). Depending on the purpose of the investigation, researchers may also elect to dichotomize count variables at thresholds other than zero, such as at an accepted diagnostic criterion.

Considering the categorical presence and absence of disordered eating symptoms in conjunction with continuously-defined symptom severity may be relevant for comprehensive models of risk and recovery. For example, depending on sample characteristics, certain variables of interest may not predict presence or absence of a behavior, but could indicate whether a behavior occurs more or less frequently for individuals who do endorse this behavior. For instance, evidence suggests that presence of weight suppression, or the difference between an individual's highest and current weight, predicts onset and maintenance of bulimic symptoms (Keel & Heatherton, 2010). However, recent findings indicate that weight suppression may not be useful in predicting the *frequency* of binge eating and vomiting among individuals who currently have an eating disorder diagnosis (Lavender et al., 2015). Considering questions of onset together with frequency of maladaptive behaviors will assist clinicians and researchers in better understanding complex patterns of risk.

## 2. Statistical considerations

When investigating binge eating and compensatory behaviors, the

frequency distributions of these variables within a population of interest may impact choice of statistical approach. When studying these constructs— either at a single time point in low-risk populations, in populations with variation in eating disorder diagnosis, or as indicators of symptom remission across time in intervention studies— the distributions often take on a characteristic shape. They are bounded by zero and highly positively skewed. Although skewness in the distribution of potentially infrequent behaviors may be expected, extreme skewness may lead to difficulties in conducting and drawing conclusions from traditional statistical analyses (Erceg-Hurn & Mirosevich, 2008).

Distributions may not be problematically skewed in all circumstances; for example, in clinical samples of individuals with BN, measures of certain symptoms might yield a distribution that approaches normality and may not necessarily require transformation before conducting analyses. However, in the event that symptoms remit over the course of treatment, distributions of behaviors (e.g., objective binge episodes) may shift towards a more extreme skew. Furthermore, even in clinical samples of individuals with eating disorders, a minority of individuals report certain types of behaviors (e.g., chewing and spitting, diet pill use, laxative use; Song, Lee, & Jung, 2015). Accordingly, it is likely that researchers investigating a range of eating disordered behaviors encounter skew within routine statistical analyses and should consider distributional factors when making decisions regarding an analytic plan.

We outline the assumptions, mathematical basis, benefits, and drawbacks of several regression-based analytic approaches for eating-related behavioral count data in Table 1. Ordinary least squares (OLS) regression techniques are commonly-used statistical models for evaluating whether specific variables account for a significant proportion of variance within a given outcome of interest. OLS approaches rest on certain assumptions (see Table 1). Should violations of these assumptions occur, results from traditional techniques will prove invalid and unreliable. As such, when investigating variables that demonstrate notable skew (suggesting non-normal distribution of errors), researchers should consider the use of alternative statistical approaches to decrease the impact of skewness. Scale transformations are commonly-used approaches for normalizing skewed distributions; however, transformations may be unsuccessful for variables that are highly skewed and are not the most appropriate choice for behavioral count data. Instead, other types of distributions that were devised for use with count variables (e.g., Poisson or negative binomial) may be more suitable. However, these distributions have other assumptions that must be met (i.e., within the Poisson model, variance should be equivalent to the mean of the variable of interest).

## 3. The utility of zero-sensitive count models

Zero-inflated and hurdle count models (referred to together here as "zero-sensitive models") present an alternative approach to addressing concerns related to high skew; these models may also circumvent limitations of other alternative distributions. Hurdle models are two-part models in which one formula is specified for predicting zero values, and another formula is specified for predicting positive values. Zero-inflated models are highly similar to zero-hurdle models, with one key difference. Whereas zero-inflated models combine a model for excess zeros with a regression model for count data, which may include some zero values (e.g., Poisson, negative binomial), hurdle models first fit all zero values in a logistic regression, and then separately model a truncated count regression for all positive values (Neighbors et al., 2011).

Zero-sensitive models may be more successful than traditional count models in evaluating outcomes when there is a preponderance of zeros in the distribution of the measured behavior. Other fields that assess infrequent behaviors and highly skewed data have begun to adopt zero-sensitive approaches to improve modeling (e.g., substance use research; Atkins, Baldwin, Zheng, Gallop, & Neighbors, 2013). Demonstrating the

utility of these models for eating behavior investigation, Grotzinger, Hildebrandt, and Yu (2014) provided a simulated example of how zero-inflated approaches may improve prediction of binge eating frequency over the course of eating disorder treatment, given adequate sample size and modest amounts of missing data. To our knowledge, only a handful of researchers have directly applied a zero-inflated count regression model to assess disordered eating behaviors (Becker et al., 2017; Eichen, Conner, Daly, & Fauber, 2012; Emery, King, Fischer, & Davis, 2013; Farstad et al., 2015; Fischer, Peterson, & McCarthy, 2013; Linna et al., 2013; Pearson, Combs, Zapolski, & Smith, 2012). In each instance, the authors found zero-inflated models to be useful, as they allow for the examination of unique research questions related to how certain risk factors may relate to (1) the likelihood of behavioral engagement and (2) the frequency of behavioral engagement within the sample of non-zeros. Therefore, while current use is limited within the field of eating pathology, zero-inflated count models may be well-suited to the characteristics of eating disorder-related outcomes.

Although alternative models for analyzing count data have been proposed and discussed within the statistical/methodological field, we aim to translate this prior work to outcomes relevant to the field of eating disorders using both real-world and simulated data. The current investigation examines the use of zero-sensitive models to analyze binge eating and compensatory behaviors in a real-world sample of college students. Using this data, we provide a practical demonstration assessing theoretical and statistical support for various analytic approaches. Data analysis was conducted in two parts. First, we examined theoretically-relevant variables associated with risk for binge eating and compensatory behaviors (i.e., Cognitive Behavioral Theory of Eating Disorders; Fairburn, Cooper, & Shafran, 2003). Analyses were designed to demonstrate how various approaches may be useful in the context of eating disorder research and provide guidance on their application, rather than to rigorously test the relation between predictors and eating-related outcomes or provide an exhaustive examination of the statistical performance of zero-sensitive approaches. We then followed up with a limited set of simulations to explore the performance of alternative analytic approaches in simulated data with varying characteristics. To conclude, results from this two-pronged approach are considered in the broader context of relevant research questions in the area of eating behaviors.

## 4. Method

### 4.1. Participants

Participants ($N$ = 524; 54.0% female) were undergraduates at a university in the Northeastern United States. Members of the sample had a mean age of 18.9 ($SD$ = 2.0) years and endorsed varied ethnic and racial backgrounds, including White (n = 309; 59.0%), Black (n = 65; 12.4%), Asian (n = 61; 11.6%), Hispanic (n = 74; 14.1%), Other (n = 45; 8.6%) Mixed Race (n = 33; 6.3%) and Native American (n = 4; 0.8%), with 1.5% of individuals choosing not to respond.

### 4.2. Measures

#### 4.2.1. Eating Disorders Examination-Questionnaire (EDE-Q; Fairburn & Beglin, 1994)

The EDE-Q, a 28-item self-report questionnaire, was developed based on a well-validated semi-structured interview assessment of eating pathology (Cooper, Taylor, Cooper, & Fairburn, 1987). Participants are asked to rate each item according to their thoughts and behaviors "over the previous 28 days." The scale has shown good validity and reliability in previous research (Mond, Hay, Rodgers, Owen, & Beumont, 2004). In the current investigation, we summed the "Shape Concern" and "Weight Concern" subscales of the measure to create a variable representing body dissatisfaction, as past work has indicated that items on both subscales seem to load onto one factor (Peterson,

Crosby, & Wonderlich, 2007). Our measure of body dissatisfaction demonstrated good internal consistency (Cronbach's α = .93), as did the Restraint (Cronbach's α = .86) subscale.

We used several items of the EDE-Q as count measurements for binge eating (Item #14: "… On how many of these times did you have a sense of having lost control over your eating (at the time that you were eating)?"), purging (Item #16: "Over the past 28 days, how many times have you made yourself sick (vomit) as a means of controlling your shape or weight?"), laxative use (Item #17: "Over the past 28 days, how many times have you taken laxatives as a means of controlling your shape or weight?"), and driven exercise (Item #18: "Over the past 28 days, how many times have you exercises in a 'driven' or 'compulsive' way as a means of controlling your weight, shape or amount of fat, or to burn off calories?"). Although reliability estimates are not available for these items, prior work has suggested that single-item measurements yield comparably good reliability and validity to multiple-item scales, particularly in the case of behavior-related reports (De Boer, Van Lanschot, & Stalmeier, 2004; Dollinger & Malmquist, 2009; Elo, Leppänen, & Jahkola, 2003).

### 4.2.2. Daily Stress Inventory (DSI; Brantley & Jones, 1989)

The DSI is a 58-item measure that presents subjects with a list of 58 common stressful events (e.g., "was embarrassed"; "feared illness/pregnancy"). The questionnaire asks subjects to rate whether each event has occurred over the previous 24-h period, and if so, how subjectively stressful they found the event, using a 7-point Likert-type scale ranging from 1 (*Occurred, but was not very stressful)* to 7 (*Caused me to panic)*. We used participants' self-reported stress in response to experienced events as a measure of current stress. Cronbach's alpha for the measurement in the current study was excellent (α = .98).

### 4.3. Procedures

Subjects were recruited through the university's research pool. Participants attended one in-lab appointment during which they provided informed consent and filled out a series of self-report questionnaires. Participants earned course credit for participation. The university's Institutional Review Board (IRB) approved all study procedures.

### 4.4. Analytic plan

R version 3.2.2 software program was used to conduct statistical analyses. Zero-sensitive analyses were conducted using the R 3.2.2 pscl package 1.4.9 (Jackman, 2015; Zeileis, Kleiber, & Jackman, 2008). Simulations were completed with Mplus 7. Zero-order correlations were computed between variables of interest in the overall sample, along with the subsamples of individuals who endorsed binge eating and compensatory behaviors. These analyses were conducted to provide preliminary information regarding the relationships among variables of interest and facilitate comparison of variable relationships across subsamples.

### 4.4.1. Part 1: Real-Life Application

In subsequent analyses, variables of disordered eating over the past 28 days (i.e., binge eating, purging behaviors, and driven exercise) were entered as dependent variables. Due to similarity in the means and distributions of vomiting and laxative use, purging behaviors were conceptualized as a single dependent variable comprised of both behaviors, to reduce redundancy in analytic exemplars. Parameters included in the model as independent variables were derived from the Cognitive Behavioral Theory of Eating Disorders (Fairburn et al., 2003) and included: dietary restraint (as measured by the EDE-Q dietary restraint subscale), body dissatisfaction (comprised of the shape and weight concern subscales of the EDE-Q), stress related to life events (DSI score), and weight suppression as an indicator of low relative

weight (difference between highest weight at adult height and current weight; Lowe, 1993).

We conducted three sets of analyses for each dependent variable to test whether different statistical approaches yielded discrepant findings. Prior to analyses, we evaluated whether the assumptions of OLS regression were met. First, we examined normality of the distribution of dependent variables, by evaluating skewness and kurtosis values (identifying absolute values > 3 as demonstrating significant skew/kurtosis) and significance of the Shapiro-Wilk's test of normality for dependent variables. Second, we evaluated the normality of the OLS regression residuals (i.e., normal distribution of error via visual examination of scatterplots and normal P-P plots of standardized predicted and residual values). In the event that the dependent variables or residuals were not normally distributed, we transformed dependent variables using an inverse transformation. Normality assumptions were then re-evaluated. For the first set of analyses, OLS regression-based models were conducted using both the untransformed and the transformed dependent variable to determine whether transformation resulted in any changes within results. Second, we evaluated logistic regression models, which examine the contribution of different variables in accounting for variance in whether behaviors are either "present" or "absent" over the past 28 days. Logistic regression has been commonly used in prior work on the variables of interest (e.g. Fairburn, Peveler, Jones, Hope, & Doll, 1993; Keel & Heatherton, 2010; Reba et al., 2005). The final set of analyses evaluated relations between predictor variables and each count variable using Poisson, negative binomial (NB), and respective zero-inflated models. We also included a zero-truncated hurdle (Hurdle NB) model for the negative binomial approach. Table 1 presents a description and comparison of each regression-based approach.

In the current analyses, count models were statistically compared to their zero-sensitive counterparts through Vuong's test— a statistical test which can be used to assess and compare model fit between non-nested models (Vuong, 1989). Across models, we also computed the root mean square error (RMSE) and the mean absolute error (MAE) to provide indicators of model fit. The RSME represents the square-root of the variance of the residuals for a given model. The MAE is the mean of the absolute value of residuals, and, in comparison to RMSE, this index punishes larger errors less harshly. These indices provide a measurement of discrepancy between predicted values and observed values, allowing for a comparison of fit under conditions where outcome variables are on the same scale (i.e., data that are untransformed and not recoded). Ultimately, model selection may be influenced by a number of factors, including the degree to which model assumptions are met, theoretical considerations, comparative model fit, and parsimony. Thus, all factors are discussed when exploring the utility of various models in the current investigation.

### 4.4.2. Part 2: Data simulations

In order to compare performance of an appropriate count-based regression method with zero-inflated counterparts across scenarios that may mimic real-world research that seeks to predict binge eating and compensatory behavior count data, we completed a Monte Carlo simulation experiment (Mplus code provided in Supplementary File 1). We examined the performance of three models (Hurdle NB, zero-inflated negative binomial [ZINB], and NB) using the negative binomial distribution (see Supplementary Table 1). Based on evaluations of models in Part 1, we chose to focus on negative binomial approaches in simulations, as the characteristics of eating-related behavioral count data (i.e. count data with a large number of zero values and overdispersion) most closely fit the theoretical assumptions of a negative binomial-based approaches.

Specifically, we examined several scenarios with a range of sample sizes (100, 500, 1000) and probabilities of zero values on the outcome variable (20%, 50%, 80%). When considering application of zero-sensitive models within real-world settings, it is critical to consider the

effect of missing data on the performance of a given model. Accordingly, we tested the performance of NB, Hurdle NB, and ZINB under conditions of missing and complete data. Given past research documenting missing data rates ∼15–20% within large samples (Kelly et al., 2017), we evaluated models with 20% missing data in addition to violating models with complete data. Parameter values for the simulation were set by a Hurdle NB model with two covariates. When specifying population effect sizes for the two variables within the model, we drew from existing research on risk factors for eating disorders, which indicates that the majority of well-researched risk factor effects fall into the small to medium range (e.g., Culbert, Racine, & Klump, 2015; Stice, 2002). Given that use of zero-sensitive models within the field of eating disorders is limited, we were not able to draw upon existing work for hypothesizing effect sizes across the count and binomial portions of the model, and instead chose to implement a range of small effect sizes that varied in their relation to count and binomial portions of the model. The first predictor was set to a modest relationship with the outcome variable in the count portion of the model (.20) and also related to the likelihood of being a zero value (.30). The second covariate related only to frequency count portion of the model, at a moderate value (.40), and did not relate to the likelihood of being a zero value. These two variables were also set to covary with one another (.20). These values were chosen to represent variables with small to medium effects that varied in relationship to the binomial portion of the model. As such, this scenario is representative of a potential regression with zero-inflated data predicting eating-related behavioral outcomes in clinical and community samples.

Models were determined to be adequate when parameter bias did not exceed 10 percent for any parameter in the model (Muthén & Muthén, 2002). Second, coverage should remain above 90. Once these conditions were satisfied, power was also evaluated, and simulations in which power was close to .80 were identified (Muthén & Muthén, 2002).

## 5. Results

### 5.1. Part 1: real world exemplar

#### 5.1.1. Descriptive Statistics

Descriptive statistics for binge eating, laxative use and vomiting, and driven exercise are presented in Table 2. Individuals reported low average levels of binge eating and compensatory behaviors, with wide variation in number of occasions endorsed over the previous 28 days. The majority of participants reported no laxative use or vomiting over the past 28 days (90.3%), no driven exercise (59.3%), and no binge eating (67.6%), indicating that distributions for each variable are likely zero-inflated. Results are presented with and without skew transformations for OLS models.

For comparison of associations among variables of interest across subsamples, correlations are also reported for relevant variables within several subsamples in Supplemental Correlation Tables 1 and 2 A comparison of correlations across these groups indicates differences in

relations among variables of interest depending on the sample, indicating that relations among variables may be dependent on presence or absence of symptoms, consistent with use of a zero-inflated model.

Table 3 presents model parameters and model fit statistics for all models, allowing for standardized comparison of models. Table 4 presents information for comparison of count models with zero-inflated approaches. In comparison to less complex models, zero-inflated models include more parameters. Accordingly, we calculated a model comparison statistic (Vuong Z) that corrects for parsimony (Table 4). Table 5 offers data relevant to clinical interpretation, presenting how an increase in 1SD in each predictor would affect the predicted value of outcomes.

#### 5.1.2. Ordinary Least Squares (OLS) Regression

In the OLS model of binge eating, examination of the residuals in the model resulted in a leptokurtic distribution; thus, the assumption of normality of residuals was not met (see Supplementary Fig. 1). When binge eating was inversely transformed, the distribution of residuals appeared bimodal, continuing to violate normality of error assumptions. In the model predicting episodes of driven exercise, residuals violated assumptions of normality. In the OLS Inverse model, body dissatisfaction emerged as a significant predictor, differing from the untransformed model. Similar to what was observed within the binge eating model, residuals within the Driven Exercise OLS Inverse model took a bimodal shape. In the OLS model predicting purging, the inverse transformation model led to a change in the significance of stress as a predictor. Again, residuals were problematic in both models, indicating that alternative approaches would be necessary to provide a more accurate fit to the data.

#### 5.1.3. Logistic Regression

A logistic regression model examined outcomes as either absent ("0") or present ("1"), circumventing issues related to the distribution of the outcome data, but artificially truncating variability in those who endorse a behavior. The Binge Eating logistic regression model produced similar results to the findings within OLS and OLS inverse models. In predicting driven exercise and purging, patterns of significance among predictors were similar to those in the respective OLS Inverse models, but differing from the original OLS models.

#### 5.1.4. Poisson Regression (PR) and Zero-inflated Poisson Regression (ZIP)

In the PR model predicting binge eating, gender, which was not significant in OLS or logistic regression models, demonstrated a statistically significant effect size. An examination of deviance statistics indicated that the data remained problematically overdispersed in this model (Pearson Dispersion = 7.04). In the Driven Exercise PR model, patterns of significance were consistent with the OLS model. Body dissatisfaction was not significant, consistent with the original OLS model but discrepant from the OLS Inverse and logistic regression models. Again, dispersion was problematic in this model (Pearson Dispersion = 9.7). In the Purging PR model, gender was a significant predictor of purging counts, which differed from results found in other

**Table 2**
Descriptive statistics of binge eating and compensatory behaviors.

| Variable | Mean | Variance | % Zeros | % Subthreshold | % Threshold | Mean of Non-zero Distribution (N) | Variance of Non-zero distribution |
|---|---|---|---|---|---|---|---|
| Binge Eating | 1.50 | 14.07 | 67.6% | 18.1% | 14.3% | 4.26 (170) | 28.85 |
| Vomiting | .51 | 7.76 | 92.2% | 4.0% | 3.8% | 6.48 (41) | 62.05 |
| Laxative Use | .39 | 2.07 | 94.1% | 2.3% | 3.6% | 6.58 (31) | 32.78 |
| All Purging | .89 | 17.81 | 90.3% | 3.1% | 6.7% | 9.21 (51) | 108.77 |
| Exercise | 3.77 | 46.76 | 59.3% | 14.1% | 26.7% | 9.26 (214) | 64.09 |
| All Compensatory | 4.66 | 75.76 | 56.8% | 13.5% | 29.8% | 10.80 (227) | 109.31 |

Note: Reported count variables presented based on responses to the Eating Disorder Examination Questionnaire (EDE-Q). Subthreshold defined as frequency of 1–3 times over the past 28 days, threshold defined as 4 or more times over the past 28 days. 'All Purging' = composite mean of Vomiting and/or Laxative Use frequency over the past 28 days. 'All Compensatory' = composite mean of Vomiting, Laxative Use, and/or Exercise frequency over the past 28 days.

**Table 3**
Regression Coefficients and Model parameters.

| | Model Fit | | | | Stress | Restraint | Body Dissat | Gender | Wt Suppression |
|---|---|---|---|---|---|---|---|---|---|
| | RMSE | ZRMSE | MAE | ZMAE | b (SE) | b (SE) | b (SE) | b (SE) | b (SE) |
| Binge Eating – OLS | 3.48 | 0.93 | 1.76 | 0.47 | 0.29 (0.13)* | 0.45 (0.17)** | 0.54 (0.22)* | 0.62 (0.33) | 0.01 (0.01) |
| Binge Eating – OLS Inv | 0.30 | 0.87 | 0.24 | 0.71 | 0.05 (0.01)*** | 0.06 (0.01)*** | 0.06 (0.02)** | −0.02 (0.03) | 0.0003 (0.001) |
| Binge Eating – LR | 1.03 | – | 0.94 | – | 0.30 (0.08)*** | 0.32 (0.11)*** | 0.36 (0.14)** | −0.04 (0.22) | 0.003 (0.009) |
| Binge Eating – PR | 1.94 | – | 1.51 | – | 0.17 (0.03)*** | 0.19 (0.03)*** | 0.32 (0.05)*** | 0.41 (0.08)*** | 0.004 (0.002) |
| Binge Eating – ZIP | 1.39 | – | 0.77 | – | | | | | |
|   Binomial | | | | | 0.30 (0.09)*** | 0.32 (0.10)** | 0.34 (0.14)* | −0.11 (0.23) | 0.003 (0.008) |
|   Count | | | | | 0.04 (0.03) | 0.06 (0.03) | 0.16 (0.05)*** | 0.51 (0.09)*** | 0.004 (0.002) |
| Binge Eating – NB | 0.84 | – | 0.75 | – | 0.32 (0.08)*** | 0.32 (0.10)** | 0.32 (0.14)* | 0.60 (0.22)** | 0.008 (0.008) |
| Binge Eating - ZINB | 1.12 | – | 0.56 | – | | | | | |
|   Binomial | | | | | 0.70 (0.27)** | 1.27 (0.55)* | 0.72 (0.55) | −0.47 (0.58) | 0.08 (0.04) |
|   Count | | | | | 0.09 (0.08) | 0.13 (0.10) | 0.26 (0.14) | 0.76 (0.26)** | −0.001 (0.007) |
| Binge Eating – NB hurdle | 1.01 | – | 0.56 | – | | | | | |
|   Binomial | | | | | 0.31 (0.08)*** | 0.32 (0.10)** | 0.36 (0.14)** | −0.04 (0.22) | 0.003 (0.001) |
|   Count | | | | | 0.07 (0.09) | 0.12 (0.10) | 0.22 (0.15) | 0.87 (0.27)** | 0.006 (0.008) |
| Exercise – OLS | 6.02 | 0.87 | 4.07 | 0.59 | −0.15 (0.48) | 2.25 (0.28)*** | 0.05 (0.38) | 1.67 (0.58)** | 0.005 (0.24) |
| Exercise – OLS Inv | 0.36 | 0.87 | 0.38 | 0.75 | 0.02 (0.03) | 0.26 (0.04)*** | 0.12 (0.06)* | 0.24 (0.08)** | 0.001 (0.003) |
| Exercise – LR | 1.06 | – | 0.98 | – | 0.09 (0.08) | 0.58 (0.11)*** | 0.31 (0.14)* | 0.55 (0.22)* | 0.003 (0.008) |
| Exercise – PR | 2.70 | – | 2.31 | – | −0.03 (0.02) | 0.41 (0.02)*** | 0.05 (0.03) | 0.47 (0.05)*** | 0.001 (0.002) |
| Exercise – ZIP | 1.38 | – | 0.96 | – | | | | | |
|   Binomial | | | | | 0.09 (0.08) | 0.57 (0.11)*** | 0.32 (0.14) | 0.54 (0.22)* | 0.003 (0.009) |
|   Count | | | | | −0.06 (0.02)*** | 0.22 (0.02)*** | −0.08 (0.03)* | 0.25 (0.05)*** | 0.003 (0.001) |
| Exercise – NB | 0.90 | – | 0.81 | – | 0.006 (0.07) | 0.45 (0.10)*** | 0.20 (0.13) | 0.71 (0.21)*** | 0.006 (0.008) |
| Exercise – ZINB | 0.89 | – | 0.61 | – | | | | | |
|   Binomial | | | | | 0.12 (0.11) | 0.85 (0.25)*** | 0.62 (0.33) | 0.70 (0.31)* | 0.003 (0.01) |
|   Count | | | | | −0.07 (0.06) | 0.25 (0.08)** | −0.10 (0.11) | 0.27 (0.18) | 0.002 (0.006) |
| Exercise – NB hurdle | 0.93 | – | 0.64 | – | | | | | |
|   Binomial | | | | | 0.08 (0.08) | 0.57 (0.11)*** | 0.31 (0.14)* | 0.54 (0.22)* | −0.003 (0.008) |
|   Count | | | | | −0.08 (0.05) | 0.24 (0.07)*** | −0.06 (0.09) | 0.29 (0.16) | 0.003 (0.005) |
| Purging - OLS | 4.04 | 0.95 | 1.58 | 0.37 | 0.27 (0.14) | 0.56 (0.19)** | 0.06 (0.25) | 0.56 (0.39) | 0.04 (0.02)* |
| Purging – OLS Inv | 0.23 | 0.95 | 0.12 | 0.51 | 0.03 (0.01)*** | 0.03 (0.01)** | 0.002 (0.01) | 0.01 (0.02) | 0.002 (0.001)* |
| Purging - LR | 0.74 | – | 0.54 | – | 0.40 (0.12)*** | 0.37 (0.15)* | 0.03 (0.21) | 0.02 (0.36) | 0.02 (0.01)* |
| Purging - PR | 1.97 | – | 1.35 | – | 0.23 (0.03)*** | 0.42 (0.04)*** | 0.09 (0.06) | 0.58 (0.11)*** | 0.02 (0.002)*** |
| Purging - ZIP | 1.60 | – | 0.51 | – | | | | | |
|   Binomial | | | | | 0.39 (0.12)*** | 0.37 (0.14)* | 0.03 (0.21) | −0.03 (0.36) | 0.02 (0.01)* |
|   Count | | | | | 0.002 (0.03) | 0.13 (0.05)* | 0.04 (0.08) | 0.59 (0.10)*** | 0.009 (0.003)** |
| Purging – NB | 0.49 | – | 0.45 | – | 0.43 (0.19)* | 0.29 (0.24) | 0.22 (0.32) | 0.47 (0.51) | 0.03 (0.02) |
| Purging - ZINB | 1.08 | – | 0.35 | – | | | | | |
|   Binomial | | | | | 0.44 (0.14)** | 0.37 (0.17)* | 0.02 (0.24) | −0.12 (0.39) | 0.02 (0.02) |
|   Count | | | | | −0.04 (0.12) | 0.11 (0.24) | 0.08 (0.38) | 0.63 (0.44) | 0.003 (0.02) |
| Purging – NB hurdle | 1.09 | – | 0.35 | – | | | | | |
|   Binomial | | | | | 0.39 (0.12)*** | 0.37 (0.15)* | 0.03 (0.21) | −0.02 (0.36) | 0.02 (0.01)* |
|   Count | | | | | −0.04 (0.12) | 0.10 (0.21) | 0.10 (0.33) | 0.63 (0.42) | 0.01 (0.01) |

*Note.* Purging includes laxative use and vomiting. Exercise refers to endorsement of driven exercise. Outcome variables measured by the Eating Disorder Examination – Questionnaire (EDE-Q). Stress = Stress related to life events measured by the Daily Stress Inventory. Body Dissatisfaction measured by the EDE-Q. Weight Suppression calculated as the difference between an individual's highest adult weight and self-reported current weight. OLS = ordinary least squares regression. OLS-Inverse = Inverse transformation applied to dependent variable. Inverse coefficients were reversed in sign to aide interpretation. LR = logistic regression. PR = Poisson regression *Significant at the 0.05 level **Significant at the 0.01 level ***Significant at the 0.001 level. Regression coefficients are unstandardized and reflect the influence of a 1-unit change in the predictor on the predicted level of the outcome variable when all other factors are set to their mean values. For example, a 1-unit increase in stress would predict a .29 increase in number of binge eating episodes in the OLS model, and an increase of .05 inverse units in the inverse transformed model. Coefficients are interpreted in the units relevant to outcomes. Thus, while they are interpretable within models as a measure of effect size, they are not comparable across models.

**Table 4**
Model fit comparisons.

| | Poisson v. ZIP | | | Negative Binomial v. ZINB | | |
|---|---|---|---|---|---|---|
| | Vuong Z | AIC Corrected | BIC Corrected | Vuong Z | AIC Corrected | BIC Corrected |
| Binge Eating | 6.03*** | 5.93*** | 5.74*** | 4.22*** | 3.33*** | 1.43 (p = .07) |
| Exercise | 10.84*** | 10.77*** | 10.64*** | 5.96*** | 5.34*** | 4.05*** |
| Purging | 4.84*** | 4.79*** | 4.71*** | 3.47*** | 2.40** | 0.15 |

*Note.* Purging includes laxative use and vomiting. Exercise refers to endorsement of driven exercise. Outcome variables measured by the Eating Disorder Examination – Questionnaire (EDE-Q). Stress = Stress related to life events measured by the Daily Stress Inventory. Body Dissatisfaction measured by the EDE-Q. Weight Suppression calculated as the difference between an individual's highest adult weight and self-reported current weight. NB = negative binomial regression. ZINB Count = count portion of the zero-inflated negative binomial model. ZINB Binomial = Binomial portion of the zero-inflated negative binomial model. Vuong Z = model comparison statistic. *Significant at the 0.05 level **Significant at the 0.01 level ***Significant at the 0.001 level.

**Table 5**
Clinical interpretation - predicted levels of behavior at varying levels of risk across models.

| | Mean Levels of Predictors | Stress +1SD | Restraint +1SD | Body Dissat +1SD | Wt Suppress + 1SD | Female | All +1SD + Female |
|---|---|---|---|---|---|---|---|
| **Binge Eating** | | | | | | | |
| Raw Data | 1.50 | | | | | 1.59 | |
| *Zeros* | 67.6% | | | | | 61.8% | |
| *Nonzero M* | 4.62 | | | | | 5.54 | |
| OLS | 1.72 | 2.10 | 2.06 | 2.48 | 1.88 | 1.43 | 3.06 |
| OLS.Inv | 2.70 | 2.93 | 2.87 | 3.07 | 2.68 | 2.66 | 3.51 |
| LR | 66.57% | 57.50% | 60.99% | 54.31% | 65.67% | 66.13% | 37.40% |
| PR | 1.23 | 1.53 | 1.43 | 1.93 | 1.30 | 1.02 | 2.40 |
| NB | 1.16 | 1.74 | 1.47 | 1.83 | 1.26 | 0.88 | 2.86 |
| ZIP | 1.34 | 1.78 | 1.63 | 2.23 | 1.44 | 1.09 | 2.80 |
| *Excess Zeros* | 65.51% | 56.36% | 59.90% | 53.84% | 64.66% | 64.3% | 36.3% |
| *Count M* | 3.88 | 4.07 | 4.06 | 4.83 | 4.08 | 3.06 | 4.40 |
| ZINB | 1.77 | 2.12 | 2.08 | 2.76 | 1.86 | 1.38 | 2.45 |
| *Excess Zeros* | 10.35% | 4.49% | 4.20% | 4.02% | 4.41% | 8.51% | 0.21% |
| *Count M* | 1.97 | 2.12 | 2.17 | 2.87 | 1.94 | 1.27 | 2.46 |
| Hurdle NB | 1.28 | 1.73 | 1.60 | 2.19 | 1.38 | 1.00 | 2.68 |
| *Zeros* | 43.99% | 30.56% | 36.41% | 29.79% | 43.68% | 35.01% | 0.00% |
| *Count M* | 2.29 | 2.49 | 2.51 | 3.13 | 2.46 | 1.53 | 2.68 |
| **Exercise** | | | | | | | |
| Raw Data | 3.78 | | | | | 3.65 | |
| *Zeros* | 59.3% | | | | | 57.9% | |
| *Nonzero M* | 9.26 | | | | | 7.83 | |
| OLS | 4.88 | 4.69 | 6.57 | 4.95 | 4.95 | 4.11 | 5.73 |
| OLS.Inv | 3.24 | 3.31 | 3.72 | 3.67 | 3.21 | 3.02 | 3.96 |
| LR | 52.22% | 49.42% | 41.46% | 41.34% | 53.01% | 58.39% | 35.11% |
| PR | 3.62 | 3.50 | 4.92 | 3.86 | 3.67 | 2.91 | 4.12 |
| NB | 3.41 | 3.44 | 4.81 | 4.53 | 3.64 | 2.46 | 4.95 |
| ZIP | 4.16 | 4.08 | 6.04 | 4.61 | 4.23 | 3.24 | 5.14 |
| *Excess Zeros* | 52.2% | 49.37% | 42.49% | 41.22% | 53.00% | 58.36% | 35.02% |
| *Count M* | 8.72 | 8.05 | 10.32 | 7.85 | 9.02 | 7.78 | 7.91 |
| ZINB | 4.78 | 4.61 | 6.88 | 5.20 | 4.86 | 3.72 | 5.25 |
| *Excess Zeros* | 34.27% | 30.78% | 21.56% | 17.99% | 34.98% | 41.86% | 12.33% |
| *Count M* | 7.28 | 6.66 | 8.76 | 6.34 | 7.47 | 6.41 | 6.30 |
| Hurdle NB | 4.11 | 3.99 | 5.93 | 4.67 | 4.20 | 3.18 | 5.12 |
| *Zeros* | 47.02% | 43.3% | 36.22% | 34.32% | 48.13% | 53.17% | 27.26% |
| *Count M* | 7.77 | 7.03 | 9.30 | 7.11 | 8.10 | 6.79 | 7.04 |
| **Purging** | | | | | | | |
| Raw Data | 0.89 | | | | | 0.84 | |
| *Zeros* | 90.3% | | | | | 88.8% | |
| *Nonzero M* | 9.22 | | | | | 8.60 | |
| OLS | 1.72 | 1.52 | 1.59 | 1.26 | 1.65 | 0.91 | 2.26 |
| | Mean Levels of Predictors | Stress +1SD | Restraint +1SD | Body Dissat +1SD | Wt Suppress + 1SD | Female | All +1SD + Female |
| OLS.Inv | 2.21 | 2.29 | 2.27 | 2.21 | 2.27 | 2.19 | 2.43 |
| LR | 91.72% | 86.94% | 89.34% | 91.37% | 89.74% | 91.64% | 78.97% |
| PR | 0.66 | 0.88 | 0.91 | 0.75 | 0.79 | 0.50 | 1.29 |
| NB | 0.57 | 0.99 | 0.71 | 0.78 | 0.78 | 0.46 | 1.83 |
| ZIP | 0.62 | 0.98 | 0.88 | 0.67 | 0.86 | 0.47 | 1.55 |
| *Excess Zeros* | 91.70% | 86.90% | 89.32% | 91.36% | 89.71% | 91.59% | 78.91% |
| *Count M* | 7.47 | 7.48 | 8.20 | 7.87 | 8.31 | 5.69 | 7.36 |
| ZINB | 0.65 | 1.02 | 0.91 | 0.76 | 0.87 | 0.51 | 1.61 |
| *Excess Zeros* | 89.37% | 82.71% | 86.44% | 89.06% | 86.40% | 88.81% | 71.74% |
| *Count M* | 6.15 | 5.87 | 6.71 | 6.96 | 6.40 | 4.60 | 5.64 |
| Hurdle NB | 0.64 | 0.97 | 0.88 | 0.76 | 0.84 | 0.51 | 1.56 |
| *Zeros* | 89.74% | 83.68% | 86.94% | 89.54% | 87.41% | 89.10% | 73.71% |
| *Count M* | 6.28 | 5.96 | 6.77 | 7.25 | 6.73 | 4.69 | 1.56 |

*Note.* Purging includes laxative use and vomiting. Exercise refers to endorsement of driven exercise. Outcome variables measured by the Eating Disorder Examination – Questionnaire (EDE-Q). Stress = Stress related to life events measured by the Daily Stress Inventory. Body dissat = Body Dissatisfaction measured by the EDE-Q. Wt Suppress = Weight Suppression calculated as the difference between an individual's highest adult weight and self-reported current weight. OLS = ordinary least squares regression. OLS-Inverse = ordinary least squares regression with inverse transformation applied to dependent variable. LR = logistic regression. PR = Poisson regression. NB = negative binomial. Numbers reflect predicted scores, percentage (or percentage likelihood) of zero scores, and predicted means of the non-zero distribution. Zero scores in zero-inflated models represent structural zeros, with the assumption that some zero scores may be accounted for in the count portion of the model as sampling zeroes. Predicted models were used to estimate means, percentage zero, and mean of non-distributions for appropriate values. In models indicating +1SD, a value of a predictor was chosen at one standard deviation above the mean value for that predictor, with all other predictors set to their mean level.

models. Dispersion of the distribution was higher than acceptable (Pearson Dispersion = 16.05).

One reason for overdispersion in PR models may be due to excess zeros. As such, ZIP models, which model zeros and values greater than zero separately, were evaluated as a way to account for excess zeros in the current data. In the Binge Eating ZIP model, gender was significant in the count portion, but not the binomial portion, of the model. Moreover, stress and restraint reached significance only in the binomial portion of the model, suggesting that effects of different predictors may be specific to presence vs. severity of binge eating. RMSE and MAE in

the ZIP model were lower than OLS and PR models, indicating improvement in fit. Dispersion of the non-zero distribution was attenuated, but remained high (Pearson Dispersion = 5.37), and the variance of the non-zero distribution (28.85) remained higher than the mean (4.62), indicating that overdispersion may not be adequately addressed in the ZIP model. Similar patterns of results emerged for the exercise and purging models, in that the relative importance of specific predictors differed across the binomial and count portions of these models, while overdispersion of non-zero values remained high (Pearson Dispersion = 5.92 and 10.53, respectively). In both cases, RMSE and MAE were improved in the ZIP as compared to PR and OLS models (see Table 3). Empirical comparison of ZIP with PR models also favored ZIP models (see Table 4).

### 5.1.5. Negative binomial (NB), zero-inflated negative binomial regression (ZINB) and negative binomial hurdle (hurdle NB) models

When count models are overdispersed, another option is to use NB regression, which draws from a different distribution, to more accurately approximate the data. In the Binge Eating NB model, results were generally consistent with the PR model. In the Binge Eating ZINB model, predictors varied in significance across the binomial and count portions of the model. Stress and restraint significantly accounted for variance in the likelihood of excess zeros. In the count portion of the model, gender predicted the number of binge eating episodes. Comparison of the NB and ZINB models favored the ZINB model in both raw and AIC-corrected Vuong Z statistics, and the more conservative BIC-corrected statistic bordered on significance (see Table 4).

In the NB model evaluating driven exercise, findings were again similar to the PR model. In the Exercise ZINB model, predictors again varied in significance across the two parts of the model. Restraint and gender both predicted likelihood of classification as an excess zero. In the count portion of the ZINB model, only restraint predicted the number of times an individual engaged in driven exercise in the previous month, and restraint scores 1SD above the mean were associated with a predicted exercise frequency of 6.88 overall, and 8.76 in the count portion of the model, a marked increase from the NB model (Table 5; NB predicted value = 4.81). Model comparison favored the zero-inflated model.

Finally, in the Purging NB model, stress was the only significant predictor, differing from other models where gender and weight suppression demonstrated significant effects. Stress and restraint both predicted likelihood of classification as an excess zero in the Purging ZINB model, and no significant predictor emerged in the count portion of the model. Raw and AIC-corrected Vuong Z statistics favored the ZINB model.

NB hurdle models generally revealed similar model fit to ZINB models. Notably, almost all zeros in the purging model were classified as excess zeros in the ZINB model (see Table 5); thus, patterns of effect sizes are strikingly similar for the NB hurdle and ZINB models for this outcome. On the other hand, relatively few zero values were classified as excess zeros in the Binge Eating ZINB model (see Table 5). Therefore, constraining all zeros to be reflective of a discrete process in this case led to a more noticeable change in effect sizes and patterns of significance in predictors when comparing the Binge Eating ZINB and Binge Eating Hurdle NB models (see Table 3, Table 5). Ultimately, theoretically-based decisions of whether abstinence over the sampling period reflects meaningful abstinence from this behavior would inform model preference. In the current sample, for instance, occasional episodes of binge eating may be expected in the general population, supporting a ZINB approach. On the other hand, abstinence from purging may be conceived as a more truly dichotomous process, supporting a hurdle model.

### 5.1.6. Model selection

Among the models investigated, RMSE and MAE consistently favored approaches with negative binomial and poisson distributions.

This choice is consistent with model assumptions and data structure that support count models in these circumstances. Among the count models, model selection was slightly more ambiguous. As a result of overdispersion, PR models were not ideal. ZIP models reduced overdispersion, improved model fit statistics, and showed statistically significant improvement in fit. However, overdispersion of the nonzero distributions within ZIP models remained higher than recommended. NB models generally showed a good fit to the data on RMSE and MAE statistics, similar or better than the ZIP models. ZINB models significantly improved fit over the negative binomial model values for most outcomes (See Table 4). In addition, NB Hurdle models may be appropriate when a researcher wishes to assume that all zero values represent a meaningful abstinence from behavior while accounting for overdispersion in the non-zero distribution. NB hurdle models also have interpretive benefit in cases where research questions of presence vs. absence and severity are assumed to be discrete processes, as predictors can be interpreted as reflecting these processes as discrete entities.

### 5.2. Part 2: simulations (Supplementary Table 1)

#### 5.2.1. Parameter Bias

Parameter bias is defined as the percentage by which the mean of parameter values within replications over- or under-estimates the population parameter value. All parameters for Hurdle NB models had less than 10% bias, with the exception of the simulation with a sample size of 100 and a 80% probability of a zero value on the outcome variable. Parameter bias was acceptable for X1 when there were only 20% zero values in the ZINB model, but increased as the probability of zero values in the sample increased. The relationship between X2 and U1 was underestimated by 19.7–32.5% across ZINB estimation models. Using the NB model, parameter bias for X1 was acceptable when the probability of a zero value was 20% in the sample. With larger proportions of zero values, the relationship between X1 and U1 was overestimated (46.4–85.6%). Similar to the ZINB model, the relationship between X2 and U1 was underestimated across model options.

#### 5.2.2. Coverage

Coverage refers to the percentage of replications for which the 95% confidence interval for the estimated parameter value includes the defined population parameter value. The Hurdle NB model provided adequate coverage for both parameters when sample sizes were either 500 or 1000, across samples. Coverage slightly diminished when sample size was 100. In the ZINB model, coverage was adequate for that parameter of U1 on X1 with sample sizes of both 100 and 500, but was inadequate for X2 across sample sizes. In the NB model, coverage for X1 was adequate at varying sample sizes when the probability of being a zero value was 20%, but not when the probability of being a zero value was 50% or 80%. Coverage for U1 on X2 was inadequate.

#### 5.2.3. Power

In this simulated data, power refers to the proportion of replications for which the null hypothesis that a parameter is equal to zero is rejected at the .05 level. For the Hurdle NB model, power for the U1 on X1 parameter approached 80% only with a sample size of 500 with 20% zero values, or with a sample size of 1000 with either 20% or 50% zero values. When 20% of the data was missing, power was further reduced, with only the simulation involving a sample size of 1000 with 20% zeros reaching over 80% power. The U1 on X2 parameter was more robust, approaching or surpassing 80% power for all simulations with sample sizes of 500 or 1000 without missing data. In the simulation with 20% missing data, the parameter of U1 on X2 retained power at sample sizes of 1000 and at sample sizes of 500 with 20% and 50% zero values. Both the ZINB and the NB models showed increased power for the U1 on X1 parameter with 50% and 80% zero values, due to overestimation of the parameter. On the other hand, these models evidenced decreased power for the U1 on X2 parameter due to underestimation of this parameter.

## 6. Discussion

Binge eating and compensatory behaviors represent commonly-studied constructs within disordered eating that are often skewed across both clinical and non-clinical populations. Count variables of these behaviors are essential to our understanding and assessment of disordered eating, and appropriate modeling of such data is necessary in efforts to properly evaluate predictors of risk and treatment outcome related to these behaviors. Often, count measures of these behaviors have a preponderance of zeros and are positively skewed, which violates assumptions of traditional OLS regression techniques and thus limits the validity and reliability of results derived from these models. Existing, commonly-used methods for increasing the normality of a skewed distribution, such as scale transformations, are often inadequate to fully normalize a distribution of variables that are highly skewed. Moreover, count regressions (i.e., poisson, negative binomial) may not fully account for the large number of zeros that is often observed in data sets with eating disordered behaviors. In these cases, the identification and increased use of alternative procedures for addressing skewed count data with a preponderance of zeros is recommended.

The current study first examined the use of zero-sensitive regression models for ED behaviors including binge eating, purging, and driven exercise behaviors in a large, non-clinical sample. Due to high skew and the large number of zeros associated with each behavior in the current sample, assumptions for OLS regression approaches were not met. Traditional scale transformations also failed to adequately correct normality of residuals across the variables of interest. Altogether, consideration of model fit statistics and data assumptions supported the use of ZINB models on all three proposed outcomes in the current investigation, with Hurdle NB models also proving useful when theoretically justified. NB and ZIP models were preferential to OLS and PR approaches, though ZINB and Hurdle NB approaches remained the methods of choice based on a comprehensive evaluation of methodological and statistical considerations. Utilization of zero-sensitive models improved model fit across all variables, supporting hypothesized benefits conferred by use of these approaches. Further, application of zero-sensitive models shifted relations between risk and outcome variables, with different associations reaching statistical significance in the binomial and count portions of the models and noticeable variations in effect sizes emerging across the two parts of the models. This pattern of results suggests that utilization of this method may more appropriately characterize processes relevant to predicting disordered eating behaviors in similar samples.

Hurdle models, which fit separate models for predicting zeros (logistic regression) and positive values (count regression), may be of particular interest to researchers who might otherwise dichotomize count variables and those who wish to examine the presence or absence of behaviors as discrete outcomes. For example, in the Purging Hurdle NB model, while dietary restraint accounted for variability in whether an individual engaged in purging or did not, it did not relate to the frequency of purging behaviors among individuals that did endorse purging. Therefore, it is possible that restraint is most relevant in predicting initiation of purging, while other factors determine progression in severity. Analyses that allow for dual examination of binomial and count processes may be particularly useful for answering key questions related to behavioral patterns within disordered eating.

It is important to note that while we attempted to identify the most appropriate models to address questions related to the current data, in so doing we simultaneously considered factors related to assumptions, model fit, and theoretical considerations. As such, the significance of any given variable, while important, does not specifically indicate which model should be used. Instead, variations in significance across models highlight the key proposition of this study—that researchers should be evaluating model options judiciously, as they may profoundly affect what one might consider or interpret as an important predictor.

In a limited series of simulations, we also investigated bias,

coverage, and power in a range of scenarios when data was generated using a Hurdle NB model. Sample sizes of at least 500 were required for adequate power in the Hurdle NB model with two predictors of small-to-moderate effect. The Hurdle NB model outperformed the NB model, and did so more substantially as the percentage of zero values in the sample increased. The ZINB model, on the other hand, had less substantial parameter bias and did not evidence a decrement in performance as the proportion of zero values on the outcome variable increased. Previous studies that have further examined simulated data in relation to binge eating outcomes (Grotzinger et al., 2014) had similar findings that model misspecification produced biased parameter estimates.

Within the current literature, there is no "rule of thumb" for determining the level at which zeros become problematic, though results from the current study indicated that performance of the NB model deteriorated in simulated data as the proportion of zeroes increased. Specifically, our simulations suggested that zero-sensitive models were particularly beneficial when the proportion of zeros was 50% or 80%, as compared to 20% zeros, specifically when the parameter in question was related to the binomial portion of the model. Future use of zero-sensitive models and exploration of their utility through simulation may establish a standard that researchers can use to make appropriate choices for analysis of a given dataset, and comprehensive examination of such questions within the field of eating disorders are worthwhile endeavors for more detailed simulation studies. In practice, savvy researchers may rely on two simple approaches when examining whether the number of zeros in their data exceed what might be tenable in traditional approaches. First, individuals can examine the intercept value in the binomial portion of zero-inflated models. When this value is significant, it indicates more zeros than expected and suggests that consideration of zero-sensitive models may be prudent. Second, the Vuong Z statistic (Table 4), assists researchers in determining more specifically whether a zero-sensitive model improves model fit. When designing a study where a preponderance of zero values is expected, investigators can also employ Monte Carlo simulation techniques by modifying the code provided (Supplementary File 1) to estimate sample size needed for adequate power given expected effects.

Findings from the present investigation suggest that zero-sensitive models represent an acceptable and advisable method for analyzing pathological eating-related behaviors. Such models should be considered when evaluating these and other related constructs (e.g., use of laxatives; vaping or cigarette use; use of muscle enhancing steroids; chew and spit). Zero-sensitive models may also be appropriate for predicting these behaviors and recovery over time in clinical samples, although future research is needed to directly test the performance of these models within this population. Given that these approaches are widely accepted in the existing statistical literature, we hope that this tutorial on the application of zero-sensitive approaches will promote their widespread use within the field of eating disorders. Of note, count-regression and zero-sensitive count regression models can be extended to a range of complex data structures, including generalized linear mixed models (GLMM; see Atkins et al., 2013 for an overview) and generalized estimating equations (GEE; Monod, 2011). Altogether, we encourage researchers who investigate eating-related constructs to consider the use of zero-inflated and hurdle methods in efforts to increase the validity of their findings and more accurately conceptualize risk and resilience.

or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.appet.2018.06.030.

## References

Agras, W. S., Crow, S. J., Halmi, K. A., Mitchell, J. E., Wilson, G. T., & Kraemer, H. C. (2000). Outcome predictors for the cognitive behavior treatment of bulimia nervosa: Data from a multisite study. *American Journal of Psychiatry, 157*, 1302–1308. https://doi.org/10.1037/e323122004-004.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5\*)*. American Psychiatric Pubhttps://doi.org/10.1176/appi.books.9780890425596.

Atkins, D. C., Baldwin, S. A., Zheng, C., Gallop, R. J., & Neighbors, C. (2013). A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of Addictive Behaviors, 27*, 166–177. https://doi.org/10.1037/a0029508.

Becker, K. R., Plessow, F., Coniglio, K. A., Tabri, N., Franko, D. L., Zayas, L. V., et al. (2017). Global/local processing style: Explaining the relationship between trait anxiety and binge eating. *International Journal of Eating Disorders, 50*, 1264–1272. https://doi.org/10.1002/eat.22772.

Brantley, P. J., & Jones, G. N. (1989). *Daily stress Inventory: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairburn, C. G. (1987). The development and validation of the body shape questionnaire. *International Journal of Eating Disorders, 6*, 485–494. https://doi.org/10.1002/1098-108X.

Culbert, K. M., Racine, S. E., & Klump, K. L. (2015). Research Review: What we have learned about the causes of eating disorders–a synthesis of sociocultural, about the causes of eating disorders–a synthesis of sociocultural, psychological, and biological research. *Journal of Child Psychology and Psychiatry, 56*(11), 1141–1164. https://doi.org/10.1111/jcpp.12441.

De Boer, A. G. E. M., Van Lanschot, J. J. B., Stalmeier, P. F. M., Van Sandick, J. W., Hulscher, J. B. F., De Haes, J. C. J. M., et al. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Quality of Life Research, 13*, 311–320. https://doi.org/10.1023/b:qure.0000018499.64574.1f.

Dollinger, S. J., & Malmquist, D. (2009). Reliability and validity of single-item self-reports: With special relevance to college students' alcohol use, religiosity, study, and social life. *The Journal of General Psychology, 136*, 231–242. https://doi.org/10.3200/GENP.136.3.231-242.

Eichen, D. M., Conner, B. T., Daly, B. P., & Fauber, R. L. (2012). Weight perception, substance use, and disordered eating behaviors: Comparing normal weight and overweight high-school students. *Journal of Youth and Adolescence, 41*, 1–13.

Elo, A. L., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health, 29*, 444–451. https://doi.org/10.5271/sjweh.752.

Emery, R. L., King, K. M., Fischer, S. F., & Davis, K. R. (2013). The moderating role of negative urgency on the prospective association between dietary restraint and binge eating. *Appetite, 71*, 113–119.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63*, 591. https://doi.org/10.1037/0003-066x.63.7.591.

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 363–370.

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41*, 509–528. https://doi.org/10.1016/s0005-7967(02)00088-8.

Fairburn, C. G., Peveler, R. C., Jones, R., Hope, R. A., & Doll, H. A. (1993). Predictors of 12- month outcome in bulimia nervosa and the influence of attitudes to shape and weight. *Journal of Consulting and Clinical Psychology, 61*, 696. https://doi.org/10.1037/0022-006x.61.4.696.

Farstad, S. M., von Ranson, K. M., Hodgins, D. C., El-Guebaly, N., Casey, D. M., & Schopflocher, D. P. (2015). The influence of impulsiveness on binge eating and problem gambling: A prospective study of gender differences in Canadian adults. *Psychology of Addictive Behaviors, 29*, 805. https://doi.org/10.1037/adb0000069.

Fischer, S., Peterson, C. M., & McCarthy, D. (2013). A prospective test of the influence of

negative urgency and expectancies on binge eating and purging. *Psychology of Addictive Behaviors, 27*(1), 294.

Grotzinger, A., Hildebrandt, T., & Yu, J. (2014). The benefits of using semi-continuous and continuous models to analyze binge eating data: A Monte Carlo investigation. *International Journal of Eating Disorders, 48*, 746–758. https://doi.org/10.1002/eat.22351.

Herbozo, S., Schaefer, L. M., & Thompson, J. K. (2015). A comparison of eating disorder psychopathology, appearance satisfaction, and self-esteem in overweight and obese women with and without binge eating. *Eating Behaviors, 17*, 86–89. https://doi.org/10.1016/j.eatbeh.2015.01.007.

Jackman, S. (2015). *pscl: Classes and Methods for R Developed in the political science computational laboratory. R package version 1.4.9*.

Jacobi, C., Völker, U., Trockel, M. T., & Taylor, C. B. (2012). Effects of an internet-based intervention for subthreshold eating disorders: A randomized controlled trial. *Behaviour Research and Therapy, 50*, 93–99. https://doi.org/10.1016/j.brat.2011.09.013.

Keel, P. K., & Heatherton, T. F. (2010). Weight suppression predicts maintenance and onset of bulimic syndromes at 10-year follow-up. *Journal of Abnormal Psychology, 119*, 268–275. https://doi.org/10.1037/a0019190.

Kelly, N. R., Cotter, E. W., Lydecker, J. A., & Mazzeo, S. E. (2017). Missing and discrepant data on the Eating Disorder Examination Questionnaire (EDE-Q): Quantity, quality, and implications. *Eating Behaviors, 24*, 1–6. https://doi.org/10.1016/j.eatbeh.2016.11.002.

Lavender, J. M., Shaw, J. A., Crosby, R. D., Feig, E. H., Mitchell, J. E., Crow, S. J., ... Lowe, M. R. (2015). Associations between weight suppression and dimensions of eating disorder psychopathology in a multisite sample. *Journal of Psychiatric Research, 69*, 87–93. https://doi.org/10.1016/j.jpsychires.2015.07.021.

Linna, M. S., Raevuori, A., Haukka, J., Suvisaari, J. M., Suokas, J. T., & Gissler, M. (2013). Reproductive health outcomes in eating disorders. *International Journal of Eating Disorders, 46*(8), 826–833.

Lowe, M. R. (1993). The effects of dieting on eating behavior: A three-factor model. *Psychological Bulletin, 114*, 100–121. https://doi.org/10.1037/0033-2909.114.1.100.

Mitchell, J. E., & Crow, S. (2006). Medical complications of anorexia nervosa and bulimia nervosa. *Current Opinion in Psychiatry, 19*, 438–443. https://doi.org/10.1097/01.yco.0000228768.79097.3e.

Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. (2004). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy, 42*, 551–567. https://doi.org/10.1016/s0005-7967(03)00161-x.

Monod, A. (2011). Generalized estimating equations for zero-inflated spatial count data. *Procedia Environmental Sciences, 7*, 281–286. https://doi.org/10.1016/j.proenv.2011.07.049.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8.

Neighbors, C., Atkins, D. C., Lewis, M. A., Lee, C. M., Kaysen, D., Mittmann, A., ... Rodriguez, L. M. (2011). Event-specific drinking among college students. *Psychology of Addictive Behaviors, 25*(4), 702. https://doi.org/10.1037/a0024051.

Pearson, C. M., Combs, J. L., Zapolski, T. C., & Smith, G. T. (2012). A longitudinal transactional risk model for early eating disorder onset. *Journal of Abnormal Psychology, 121*(3), 707.

Peterson, C. B., Crosby, R. D., Wonderlich, S. A., Joiner, T., Crow, S. J., Mitchell, J. E., ... Le Grange, D. (2007). Psychometric properties of the eating disorder examination-questionnaire: Factor structure and internal consistency. *International Journal of Eating Disorders, 40*, 386–389. https://doi.org/10.1002/eat.20373.

Reba, L., Thornton, L., Tozzi, F., Klump, K. L., Brandt, H., Crawford, S., ... Kaplan, A. S. (2005). Relationships between features associated with vomiting in purging-type eating disorders. *International Journal of Eating Disorders, 38*, 287–294. https://doi.org/10.1002/eat.20189.

Song, Y. J., Lee, J. H., & Jung, Y. C. (2015). Chewing and spitting out food as a compensatory behavior in patients with eating disorders. *Comprehensive Psychiatry, 62*, 147–151. https://doi.org/10.1016/j.comppsych.2015.07.010.

Spoor, S. T., Stice, E., Burton, E., & Bohon, C. (2007). Relations of bulimic symptom frequency and intensity to psychosocial impairment and health care utilization: Results from a community-recruited sample. *International Journal of Eating Disorders, 40*, 505–514. https://doi.org/10.1002/eat.20410.

Stice, E. (2002). Risk and maintenance factors for eating pathology: A meta-analytic review. *Psychological Bulletin, 128*(5), 825.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society, 307*–333. https://doi.org/10.2307/1912557.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software, 27*. https://doi.org/10.18637/jss.v027.i08.