Avoiding Inferential Errors in Public Health Research: The Statistical Modelling of Physical Activity Behavior

Ann O. Amuta[1,*] and Dudley Poston Jr.[2]

[1]*Department of Health & Kinesiology, Texas A&M University, College Station, TX, 77843, USA*

[2]*Department of Sociology, Texas A&M University, College Station, TX, 77843, USA*

Abstract: *Background*: A review of the health behavior literature on the statistical modeling of days of physical activity (PA) indicates that in many instances linear regression models have been used. It is inappropriate statistically to model a count dependent variable such as days of physical activity with Ordinary Least Squares (OLS). Many count variables have skewed distributions, and, also, have a preponderance of zeroes. Count variables should not be treated as continuous and unbounded. If OLS is used, estimations of the regression will frequently turn out to be inefficient, inconsistent and biased, and such outcomes could well have incorrect impacts on health programs and policies.

Methods: We considered three statistical methods for modelling the distribution of days of PA data for respondents in the 2013 Health Information Trends Survey (HINTS). The three regression models analyzed were: Ordinary Least Squares (OLS), Negative Binomial (NBRM), and Zero-inflated Negative Binomial (ZINB). We used the exact same predictor variables in the three models. Our results illustrate the differences in the results.

Results: Our analyses of the PA data demonstrated that the ZINB model fits the observed PA data better than either the OLS or the NBRM models. The coefficients and standard errors differed in the zero-inflated count models from the other models. For instance, the ZINB coefficient for the association between income and PA behavior was not statistically significant (p>0.05), whereas in the NBRM and in the OLS models, it was statistically significant (p<0.05).

Conclusions: The inappropriate use of regression models could well lead to wrong statistical inferences. Our analyses of the number of days of moderate PA demonstrated that the ZINB count model fits the observed PA data much better than the OLS model and the NBRM.

Keywords: Count Regression, Inference error, Measurement, physical activity, Health behavior.

BACKGROUND

Much of the research in health behavior seeking to understand the variation among individuals in their amount of physical activity (PA) operationalizes the PA variable as the number of days in a typical week in which the person engages in PA. For instance, the Health Information National Trends Survey (2013) (HINTS), a nationally representative survey that has been administered every several years since 2003 by the National Cancer Institute, asked in the cycle conducted between October 2012 and January 2013 the following question: “In a typical week, how many days do you do any physical activity or exercise of at least moderate intensity, such as brisk walking, bicycling at a regular pace, and swimming at a regular pace?” Participants answered this question with values ranging from 0 to 7 days. Such data are referred to by statisticians as “count” data; the respondent’s answer to a question such as the above is a count of the number of days he/she engages in physical activity in a typical week.

There are several examples in the health behavior literature of research analyzing a count PA dependent variable measured in the same or in a similar way as the PA variable in the Health Information National Trends Survey [1-5]. Often studies such as these analyze the count dependent variable by estimating ordinary least squares (OLS) regression equations. However, it is inappropriate statistically to model a count dependent variable with OLS. For many count variables, the distribution is heavily skewed with a long right tail and is thus far from being normally distributed. This is certainly the case with the PA variable from the HINTS dataset because, as we will see later, many people in the HINTS data, almost 30 percent, engage in no activity, with just a few having 6 or 7 days of physical activity of moderate intensity. Moreover, even if the count data were normally distributed, OLS is still an incorrect statistical method to use because OLS can produce predicted counts that are negative, which is an impossible outcome for a count variable.

Count variables should not be treated as though they are continuous and unbounded. Count data are usually not independently and identically distributed. Thus, the use of OLS to predict count outcomes will often result in incorrect results if one or more of the OLS assumptions are not met. Hence statisticians have noted that while “the linear regression model has often been applied to count outcomes, this can result in inefficient, inconsistent and biased estimates ... It is

*Address correspondence to this author at the Texas A&M University, Texas A&M University, Department of Health & Kinesiology, College Station, TX 77843, USA; Tel: 979.845.9280; Fax: 979.847.8987;
E-mail: aamuta@hlkn.tamu.edu

(thus) much safer to use models specifically designed for count outcomes."[6]

In this paper we analyze physical activity (PA) data from Cycle 2 of the 4th Iteration of HINTS. We show that the use of an OLS model to estimate PA is statistically inappropriate, and that count regression models are preferred. We show specifically that a zero-inflated negative binomial count model is the statistically best and preferred model for modeling the PA data. We turn now to a more detailed statement of the methodology for estimating count models.

METHODS

The most basic approach for predicting a count variable, such as PA, is the Poisson regression model (PRM). The Poisson model is fundamental for understanding count regression models. In the PRM, the dependent variable, namely, the number of events, i.e., in the case of this paper, the number of days in a typical week in which the respondent engages in PA, is a nonnegative integer and has a Poisson distribution with a mean that depends on the characteristics (the independent variables) of the respondents [6,7]. The PRM incorporates observed heterogeneity according to the following structural equation:

$$\mu_i = \exp\,(a + X_{1i}\, b_1 + X_{2i}\, b_2 + ... + X_{ki}\, b_k) \quad (1)$$

where:

μ_i is the expected number of days in a typical week in which the i^{th} respondent engaged in PA; $X_{1i}, X_{2i} \ldots X_{ki}$ are the characteristics of the i^{th} respondent; and a, b_1, $b_2 \ldots b_k$ are the Poisson regression coefficients.

The PRM is appropriate when the mean and the variance of the count distribution are similar, and is inappropriate when the variance of the distribution greatly exceeds the mean, that is, when there is a significant amount of over-dispersion in the count data. If this occurs, the estimates from the PRM will tend to be consistent, but inefficient. "The standard errors in the Poisson regression model will be biased downward, resulting in spuriously large z-values and spuriously small p-values" [6,8], which could lead the investigator to make incorrect statistical inferences about the significance of the independent variables.

This situation is addressed by adding to the PRM "a parameter that allows the conditional variance of (the count outcome) to exceed the conditional mean"[7]. This extension of the Poisson regression model is known as the negative binomial regression model (NBRM). The NBRM adds to the Poisson regression model the error term **ε** according to the following structural equation:

$$\mu_i = \exp\,(a + X_{1i}\, b_1 + X_{2i}\, b_2 + ... + X_{ki}\, b_k + \varepsilon_i) \quad (2)$$

However, sometimes there are many more zeroes in the count dependent variable than are predicted by the PRM or by the NBRM, resulting in an overall poor fit of the model to the data. Zero-inflated count regression models respond to this problem of excess zeroes "by changing the mean structure to allow zeroes to be generated by two distinct processes" [6].

Consider two examples of excess zeroes. Suppose one wishes to survey visitors to a national park to predict the number of fish they caught. Suppose that data were not available about the visitors to the park who did and who did not fish. The data gathered hence would likely have a preponderance of zeroes, some of which would apply to persons who fished and caught no fish, and others to persons who did not fish [9].

Or consider the issue of predicting the number of days in a typical week a person engages in physical activity. Some persons will never engage in PA either because they have chosen not to do so, or, perhaps, because they are not permitted for medical and other reasons to do so; in other words they would have a zero probability of ever engaging in PA. Conversely, some other persons would report zero days of PA because they were unable to do so; they had planned and wished to engage in one or more days of PA but for various reasons were not able to do so; their probability of engaging in PA is not always-zero.

In such a situation, there will likely be many zeroes in the dataset, and they will be two kinds of zeroes. Some of the zeroes will apply to respondents who tried to engage in one or more days of PA, but for one reason or another were not successful; and the other zeroes would apply to respondents who neither intended nor tried to engage in PA. When modelling PA, the researcher should be including in the sample persons with zero days of PA who tried to engage in PA.

Long and Freese (2006: 394) have noted that in zero-inflated models it is assumed that "there are two latent (i.e., unobserved) groups. An individual in the Always-0 Group (Group A) has an outcome of 0 with a probability of 1, while an individual in the Not Always-0 Group (Group ~A) might have a zero count, but there is a nonzero probability that he or she has a positive count."

The estimation of zero-inflated count regression models involves three steps: 1) predicting membership in the two latent groups, Group A and Group ~A; 2) estimating the number of counts for persons in Group ~A; and 3) computing the observed probabilities as a "mixture of the probabilities for the two groups weighted by the proportion in each group" [6]

To analyze the count of days of PA in a typical week for a group of survey respondents, one would follow these steps: [6, 9]

In Step 1, use a logistic regression model to predict the respondent's group membership in Group A (never engage in PA) or Group ~A (may or may not engage in PA). The independent variables used in the logistic equation are "referred to as inflation variables since they serve to inflate the number of 0s" [6].

In Step 2, for respondents in Group ~A (may or may not engage in PA), depending on whether or not there is over-dispersion in the CEB dependent variable, use either a Poisson regression model or a negative binomial regression model to predict the probabilities of counts 0 to *y* (where *y* in our analysis will equal 7, the maximum number of days in a typical week in which the subject engages in PA).

In Step 3, the results from the preceding steps are used to determine the overall probability of 0's, which is "a combination of the probabilities of 0's from each group, weighted by the probability of an individual (survey respondent) being in the group" [7]. The probabilities of counts other than 0 are adjusted in a similar way.

RESULTS

As already noted we use data in this paper from the 2nd Cycle of the 4th Iteration of Health Information National Trends Survey (2013) (HINTS), a nationally representative survey administered every several years since 2003 by the National Cancer Institute; the data for this cycle were gathered *via* a single mail-mode survey in the period between October 2012 and January 2013. The survey is a sample of the U.S. adult civilian non-institutionalized population aged 18 or older. The final HINTS 4 Cycle 2 sample contains data for over 3,600 respondents. Owing to issues of missing data, we have responses for analysis in this paper for 3,422 subjects [10].

Table **1** reports descriptive data for the dependent variable of the count of days of physical activity and for the independent variables used in the analysis as predictors of the count (all four of the independent variables are dummy variables). The independent variables reflect the socioeconomic and demographic characteristics of the respondents that have been shown in the literature to be related to one's days per week of physical activity [11-15].

The respondents had a mean count of days of PA in a typical week of 2.61, with a standard deviation of 2.38, and a variance of 5.20; thus there is extensive overdispersion in the PA data. Figure **1** is a graph showing the percentage distribution of the respondents by their respective counts of PA, from 0 to 7. Almost 30% of the respondents reported they had zero days of PA in a typical week, and 8% had one day, 13% had two days, and so forth. The PA distribution is far from normally distributed.

Regarding the independent variables we use in our regressions to predict each person's count of PA days, over 61% of the participants were female, and 17% had annual incomes of $100,000 or more. Almost 62% of the participants had at least some college attained, and 27% were of age 65 or older (Table **1**).

We first estimate an equation predicting the count of PA days using ordinary least squares regression

Table 1: Descriptive Data: Count of Days of Physical Activity, and Four Independent Variables, 3,422 Respondents: Cycle 2 of the 4th Iteration of the Health Information National Trends Survey, 2013

Variable	Mean	Standard Deviation
Count of Days of Physical Activity	2.62	2.38
Gender (Female = 1)	0.61	0.46
Income ($100,000 or more = 1)	0.17	0.38
College (some college or more = 1)	0.62	0.49
Old Age (age 65+ = 1)	0.27	0.45

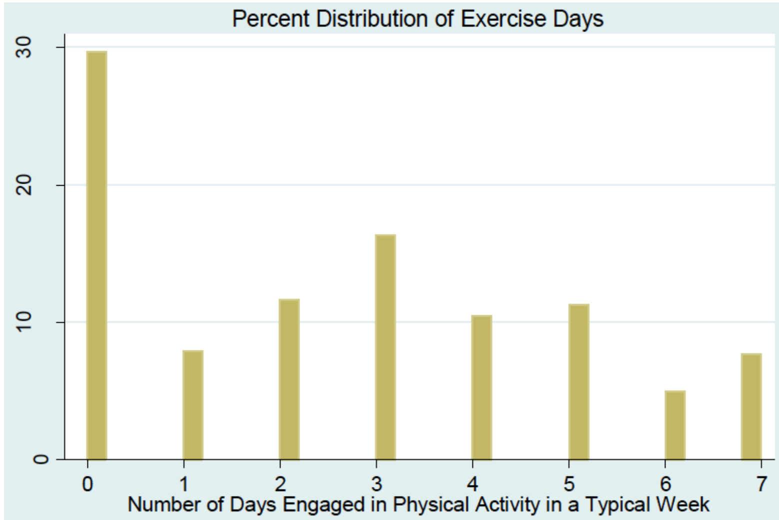

Figure 1: Showing Percent Distribution of Exercise Days.

(OLS). We do so even though it is neither the preferred nor the statistically correct method for predicting count variables. We later compare the OLS regression effects and fit with the results from models especially designed for predicting count variables.

In the first panel of Table **2**, we report the OLS results of regressing the PA count variable on the four independent variables of gender, high-income, college, and old-age. All four of the independent variables are dummy variables (see Table **1** for their 0-1 coding specifics). Of the four independent variables, three of them have statistically significant effects on the count of PA days. On average, females have about one-third of a PA day less than males, controlling for the effects of the other variables. Persons reporting high incomes ($100,000+) have about .37 more days of PA per week than persons without high incomes; and persons with at least some college have about .45 more days of PA than persons with no college. The old-age variable has no statistically significant effect on the count of PA days. Even though the adjusted value of R^2 is low (a value of 0.021), the F-test of the OLS model is statistically significant, ($p<0.001$).

How well does the OLS model fit the distribution of the PA count data? Figure **2** is a histogram of the predicted probabilities of the respondents at each PA

Table 2: Ordinary Least Squares (OLS) Regression Model, Negative Binomial Regression Model (NBRM), and Zero-inflated Negative Binomial Regression Model (ZINB), 3,422 Respondents, Health Information Trends Survey, 2013

	OLS		NBRM		ZINB	
Independent Variable	**b**	**t**	**b**	**z**	**b**	**z**
	Panel 1			**Panel 2**		**Panel 3**
X_1 Gender (female =1)	-.342	-4.31	-.132	-3.77	-.088	-3.88
X_2 High Income	.371	3.51	.130	2.83	.029	1.03
X_3 College	.446	5.40	.176	4.73	-.007	-0.28
X_4 Old Age	.077	0.88	.039	0.62	.119	4.70
Constant	2.811	18.99	1.031	15.82	1.396	33.30
F-test	19.02, P = 0.00					
Likelihood Ratio χ^2				55.22, P = 0.00		40.05, P = 0.00
Alpha (over-dispersion parameter)			0.623			
Likelihood Ratio χ^2 test of Alpha			1296.39, P = 0.00			
Vuong Test of ZINB vs. NBRM						95.00, P = 0.00

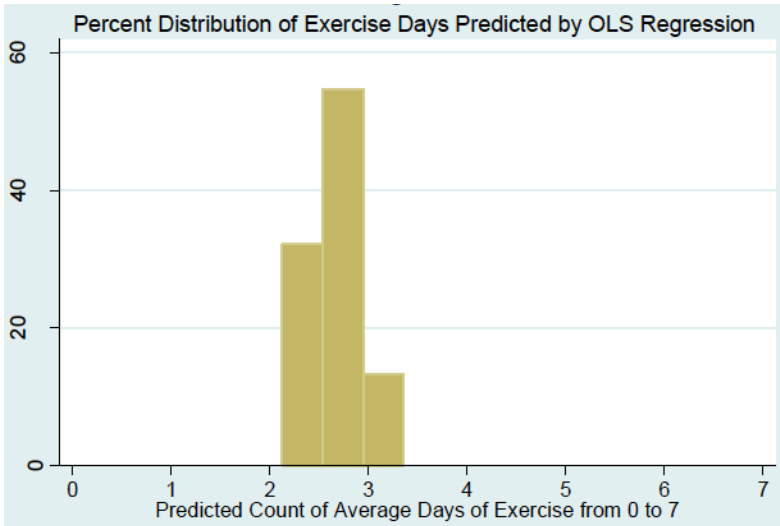

Figure 2: Showing Percent Distribution of Exercise Days Predicted by OLS Regression.

count from 0 to 7. The OLS model performs very poorly predicting the actual counts of days of PA (compare the predicted counts in Figure **2** with the actual counts in Figure **1**). The OLS model has predicted counts for all 3,422 respondents within the narrow range of 2.1 to 3.4. To illustrate, Person #1 in the sample reports 5 PA days; the OLS model predicts 2.6 PA days for this person. Person #6 reports 0 PA days; the OLS model predicts 2.2 PA days for this person. The OLS model does a very poor job predicting the range of PA counts for the respondents.

We noted earlier that irrespective of the fit of the OLS model with the count data, it is inappropriate statistically to model a count dependent variable with OLS. For one thing, many count variables have skewed distributions, and, also, many of them have a preponderance of zeroes. Count variables should not be treated as though they are continuous and unbounded. Count data are usually not independently and identically distributed. Also, even if the count data were normally distributed, OLS would still be an incorrect statistical method to use because OLS can produce predicted counts that are negative. Hence it is preferred to use a regression model specifically designed for count data.

We noted above that the most basic model for count data is the Poisson regression model (PRM). But it is only appropriate when the mean and the variance of the count distribution are similar, and is less appropriate when the variance of the distribution exceeds the mean, that is, when there is over-dispersion in the count data. We noted above that the mean number of PA days among the respondents in our sample is 2.61 and the variance is 5.20. This indicates over-dispersion in the PA count data. There are several formal ways for determining if there is over-dispersion in the count data (see Poston, 2002)[16], and one test will be discussed and shown below. For now we conclude that there is a significant amount of over-dispersion in the count data, so that a negative binomial regression model is preferred over a Poisson model.

The second panel of Table **2** reports the results of a negative binomial regression (NBRM). The same independent variables are used in this regression equation, as were used in the OLS regression shown in the first panel of Table **2**.

The same three independent variables have statistically significant negative binomial regression coefficients (panel 2 of Table **2**) as had statistically significant OLS regression coefficients (panel 1 of Table **2**), namely, gender, high-income, and college. Thus, even if the researcher were to use the less appropriate OLS model instead of the more appropriate NBRM, there would be no differences in statistical inference. Moreover, the likelihood ratio chi-square test (analogous to the global T-test in the OLS model) indicates that the NBRM is statistically significant, ($p<0.001$).

The dependent variable in the negative binomial regression model (NBRM) shown above in equation #2 is the predicted count of PA days; it is related nonlinearly to the right-hand side variables. Another way, therefore, to gauge the fit of the NBRM is in terms of predicted probabilities. Figure **3** shows the NBRM

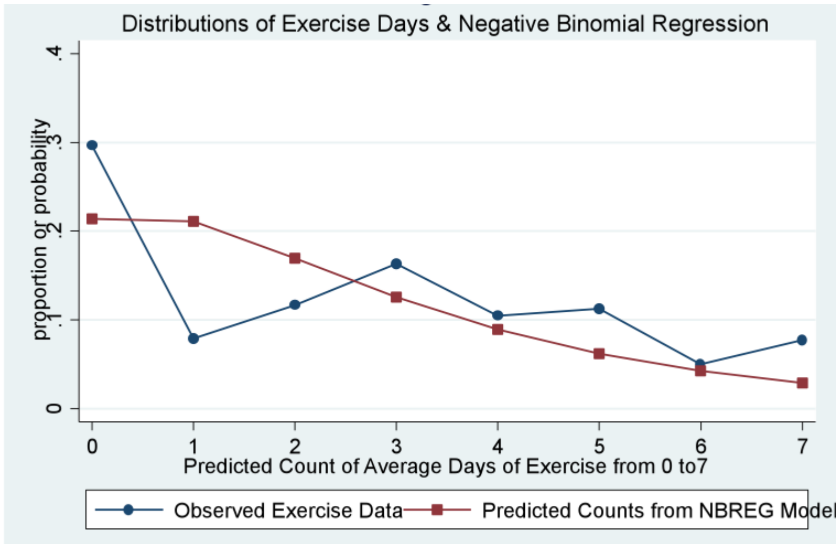

Figure 3: Showing Distributions of Exercise Days & Negative Binomial.

predicted probabilities compared with the empirical distribution of the PA count data. The predicted probabilities from the NBRM fit the empirical PA data in a fair way for counts 2 through 7, but not well for counts 0 and 1.

We concluded above that the NBRM was preferred over a Poisson model because there was over-dispersion in the PA count data. One of the formal statistical tests for appraising the presence of dispersion in the count data is the alpha parameter shown at the base of the NBRM results (panel 2 of Table **2**). Alpha has a value of .623, indicating the presence of over-dispersion in the count data; the likelihood-ratio χ^2 test of alpha (bottom of panel 2, Table **2**) has a high value of 1296.4, with a probability of .000, indicating that the probability that one would observe these data if the process was Poisson, i.e., if alpha = 0, is virtually zero. The PA count data are clearly not Poisson [6, 17].

We showed in Figure **1** that there is a very large number of zeroes in the PA count data. Indeed almost 30% of the respondents in the 2013 HINTS data reported they had zero PA days in a typical week. Given this preponderance of zeroes in the PA data, we estimated a zero-inflated negative binomial regression model (ZINB) to see if the model fit would be improved over that of the NBRM. The ZINB results in the third panel of Table **2** may be compared to those of the NBRM shown in the second panel of Table **2**, and to the OLS coefficients in the first panel.

But let us first determine if the zero-inflated negative binomial model in panel 3 is preferred over the straight-forward negative binomial model in panel 2. There is a formal test statistic, the Vuong test [18], which determines statistically whether the ZINB model is a significant improvement over the NBRM. The Vuong statistic is asymptotically normal; if its value is > 1.96, the ZINB model is preferred over the NBRM, and if not, the NBRM is preferred. The Vuong test statistic is shown at the base of the third panel of Table **2**; Vuong = 95.0. This is clear evidence that the zero-inflated negative binomial regression results are preferred over the straight-forward negative binomial regression results.

The ZINB coefficients predicting PA counts of the respondents (panel 3 of Table **2**) are very different from those from the NBRM (panel 2 of Table **2**). Note first that although the magnitude of the gender coefficient in the ZINB model is less than that in the NBRM, in both instances the z values are above 2.0, indicating the statistical significance of gender on PA counts. This is also a conclusion indicated by the gender coefficient in the OLS model. In all three models females are reported on average to have less PA days than males.

However, and very importantly, the ZINB coefficients in panel 3 for two of the remaining three independent variables, namely high-income and college are not statistically significant, whereas in the NBRM and in the OLS model, these coefficients are significant. Thus were a researcher to predict the count of PA days with an OLS model or with a negative binomial model, he/she would conclude that high income people and people with at least some college have on average more PA days in a typical week than persons without high income or without some college

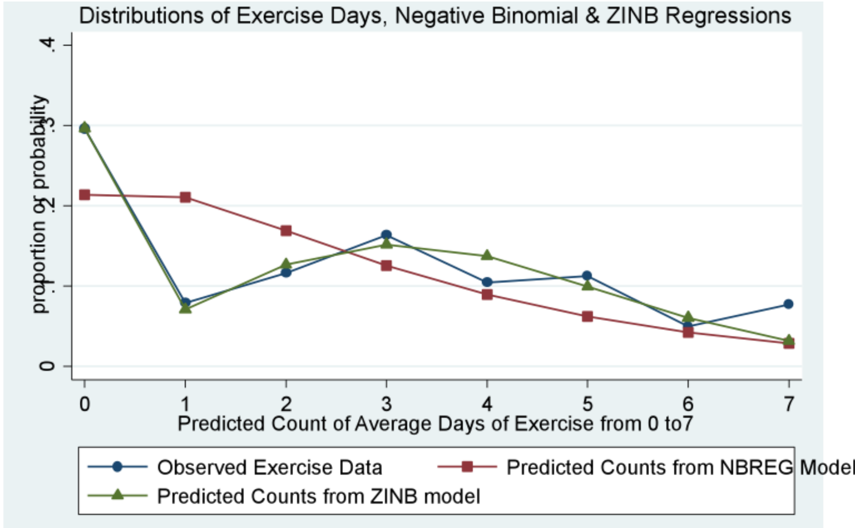

Figure 4: Showing Distributions of Exercise Days, Negative Binomial & ZINB Regressions.

years. But this is exactly the opposite of the ZINB results. In the zero-inflated model, there are no statistically significant effects, on average, of having high income or of having at least some college on the count of the number of PA days. Were the researcher to predict PA days with an OLS model or with a NBRM, he/she would end up making statistical inference errors. In the statistically preferred and correct model, the ZINB model, these two variables do not have a statistically significant effect on the count of PA days.

Also, in the NBRM and in the OLS model, there is no statistically significant effect of old age on the count of PA days. That is, both of these models indicate that on average the predicted count of PA days for old people is not different from the predicted count for people who are not old. But in the statistically preferred and correct ZINB model (panel 3 of Table **2**), persons of old age are shown on average to have more PA days than person who are not old. Thus a researcher predicting PA days with an OLS model or with a NBRM, instead of with the statistically correct ZINB model, would make another error of statistical inference.

Finally, we inquire about the fit of the predicted probabilities of the ZINB model with the empirical distribution of the PA count data. We also ask about how does the ZINB fit compare with the fit of the NBRM predicted probabilities? Figure **4** shows the ZINB and NBRM sets of predicted probabilities compared with the empirical distribution of the PA count data. We conclude that the predicted probabilities from the ZINB fit the empirical PA data considerably better –- indeed much better -- than NBRM sets of predicted probabilities. The amount of over- and under-prediction of the PA counts with the ZINB predicted probabilities is at a minimum. This provides even more support for the ZINB results compared to the NBRM results.

DISCUSSION

This paper discussed statistical methods for modelling the distribution of days of physical activity (PA) data for respondents in the Health Information Trends Survey of 2013. We first showed that the distribution of the PA data was not normal (Gaussian), but, rather, skewed with a right tail, and also, with a preponderance of zeroes. Given such a distribution, a linear regression (OLS) model is inappropriate for statistical modeling. Four socioeconomic and demographic variables were then used as independent variables to model the counts of PA days for the respondents. We estimated an ordinary least squares (OLS) regression model, a negative binomial regression model (NBRM), and a zero-inflated negative binomial (ZINB) regression model. We showed that the ZINB model was by far the statistically preferred and correct model to predict the count of PA days. We showed also that three of the four slope coefficients in the OLS model and in the NBRM would have resulted in major errors of statistical inference had their interpretations been based only on the results of these models rather than on the basis of the more correct ZINB model.

The health behavior literature on the statistical modeling of days of physical activity indicates that in many instances linear regression models have been used. The decision to use a linear model to predict a count variable such as PA days, however, is almost always wrong. Health researchers should not treat such a dependent variable as an interval variable, but as a count, i.e., as a non-negative integer. We echo the observation of Long and Freese (2006: 349) that while "the linear regression model has often been applied to count outcomes, this can result in inefficient, inconsistent and biased estimates ... It is (thus) much safer to use models specifically designed for count outcomes."

Our analysis of the HINTS 2013 survey of the number of days of moderate physical activity (PA) in a typical week clearly demonstrated that the zero-inflated negative binomial count model fits the observed PA data much better than the OLS model and the NBRM. The coefficients were so very different in the ZINB count model compared to the other two models that there would be several errors of statistical inference were the researcher to rely only on the results of the OLS model and the negative binomial regression model.

The authors wish to acknowledge Texas A&M University for providing office space and statistical software to run all the analysis presented in this paper.

REFERENCES

[1] McCully SN, Don BP, Updegraff JA. Using the Internet to Help With Diet, Weight, and Physical Activity: Results From the Health Information National Trends Survey (HINTS). J Med Inter Res 2013; 15(8). http://dx.doi.org/10.2196/jmir.2612

[2] Boone-Heinonen J, Diez Roux AV, Kiefe CI, Lewis CE, Guilkey DK, Gordon-Larsen P. Neighborhood socioeconomic status predictors of physical activity through young to middle adulthood: The CARDIA study. Soc Sci Med 2011; 72(5): 641-9. http://dx.doi.org/10.1016/j.socscimed.2010.12.013

[3] Sallis JF, Richard Hofstetter C, Faucher P, Elder JP, Blanchard J, Caspersen CJ, *et al*. A multivariate study of determinants of vigorous exercise in a community sample. Prevent Med 1989; 18(1): 20-34. http://dx.doi.org/10.1016/0091-7435(89)90051-0

[4] Jago R, Fox KR, Page AS, Brockman R, Thompson JL. Parent and child physical activity and sedentary time: do active parents foster active children? BMC Public Health 2010; 10(1): 194. http://dx.doi.org/10.1186/1471-2458-10-194

[5] Vandewater EA, Shim M-s, Caplovitz AG. Linking obesity and activity level with children's television and video game use. J Adoles 2004; 27(1): 71-85. http://dx.doi.org/10.1016/j.adolescence.2003.10.003

[6] Long JS, Freese J. Regression Models for Categorical Dependent Variables Using Stata. Second ed. College Station, Texas: Stata Press 2006.

[7] Long JS. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, California: Sage Publications 1997.

[8] Cameron AC, Trivedi. PK. Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. J Appl Economet 1986; 1: 29-53. http://dx.doi.org/10.1002/jae.3950010104

[9] Colin CA, Trivedi PK. Regression analysis of count data. Cambridge, UK: Cambridge Univ. 1998.

[10] Health Information National Trends Survey [Internet]. National Cancer Institute 2013.

[11] Carroll-Scott A, Gilstad-Hayden K, Rosenthal L, Peters SM, McCaslin C, Joyce R, *et al*. Disentangling neighborhood contextual associations with child body mass index, diet, and physical activity: The role of built, socioeconomic, and social environments. Soc Sci Med 2013; 95: 106-14. http://dx.doi.org/10.1016/j.socscimed.2013.04.003

[12] Worsley A, Wang WC, Hunter W. Gender differences in the influence of food safety and health concerns on dietary and physical activity habits. Food Policy 2013; 41: 184-92. http://dx.doi.org/10.1016/j.foodpol.2013.05.011

[13] Scheers T, Philippaerts R, Lefevre J. Compliance with different physical activity recommendations and its association with socio-demographic characteristics using an objective measure. BMC Public Health 2013; 13(136): 10.1186.

[14] Strong LL, Anderson CB, Miranda PY, Bondy ML, Zhou R, Etzel C, *et al*. Gender differences in sociodemographic and behavioral influences of physical activity in Mexican-origin adolescents. J Phys Activ Health 2012; 9(6): 829-39.

[15] Lämmle L, Worth A, Bös K. Socio-demographic correlates of physical activity and physical fitness in German children and adolescents. Eur J Public Health 2012; 22(6): 880-4. http://dx.doi.org/10.1093/eurpub/ckr191

[16] Poston Jr DL. The statistical modeling of the fertility of Chinese women. J Modern Appl Statist Method 2002; 1(2): 47.

[17] StataCorp [Internet]. StataCorp LP. 2013.

[18] Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: J Economet Soc 1989: 307-33. http://dx.doi.org/10.2307/1912557

Received on 21-07-2014 Accepted on 11-10-2014 Published on 06-11-2014

http://dx.doi.org/10.6000/1929-6029.2014.03.04.7