Application of Discrete Regression Models for Analyzing K-8 Students Nonchronic Absenteeism in the United States

Khairul Islam[1] & Tanweer J. Shapla[2]

[1,2]Department of Mathematics and Statistics, Eastern Michigan University, USA

Correspondence: Khairul Islam, Department of Mathematics and Statistics, Eastern Michigan University, Ypsilanti, MI 48197, USA.

Abstract

Absenteeism is a national crisis in the United States, and must be addressed adequately at the early stages or at its onset, to prevent consequential disaster and burden due to absenteeism. A pervasive and persuasive nonchronic absenteeism results in chronic absenteeism, and causes severe damage to students' life, schools and societies. While a good number of articles address various issues relating to chronic absenteeism, no evidence of research exists investigating nonchronic absenteeism. The aim of this article is to investigate factors affecting nonchronic absenteeism in K-8 students in the United States by applying discrete regression models. Initially, we investigate K-8 students nonchronic absenteeism discrepancies due to socio-demographic and parental involvement factors via descriptive analysis and then employ Poisson and negative binomial regression models for exploring significant factors of K-8 nonchronic absenteeism. The findings of this study will be of great use to stakeholders in developing appropriate incentive measures for reducing nonchronic absenteeism early and thereby reducing chronic absenteeism.

Keywords: goodness-of-fit, K-8 student, negative binomial regression, nonchronic absenteeism, poisson regression

1. Introduction

The state of being absent from schools for 18 days or more in a given academic year for any reason, including excused and unexcused absences or suspensions is referred to as a chronic absenteeism (U.S. Department of Education, 2016a). On the other hand, the state of being absent from schools for less than 18 days in a given academic year may be termed as the nonchronic absenteeism. As a matter of facts, one can, therefore, argue that the pervasive and persuasive form of nonchronic absenteeism is the chronic absenteeism. It appears that the chronic absenteeism is widespread in the United States, and is often termed as a national hidden crisis or unidentified problem (Bruner et al., 2011; Education Commission of the States, 2010; U.S. Department of Education, 2016b). About 10% of kindergarteners and first-graders in the United States are chronically absent (Chang & Romero, 2008). Chronic absenteeism has been highlighted in a great number of articles in the literature (Allen et al., 2018; Bauer et al., 2018; Bruner et al., 2011; Gottfried, 2014; Jordan, 2019). It is evident that the severity and rates of chronic absenteeism vary between states, school districts, and even schools within a district (Education Commission of the States, 2010).

Many articles provide evidence of chronic absenteeism at different levels of education, and how it may cause adverse impact on students' lives and academic performance (Chang et al., 2014; Gershenson et al., 2017; Gottfried, 2010, 2011, 2014; Kiani et al., 2018; National Association of Elementary School Principals, 2016; Washington, 2017; Reid, 2005, 2008), including interventions for reducing absenteeism (Jordan, 2019; Robinson et al., 2018; Rogers & Feller, 2018). According to Gottfried (2014), the chronic absenteeism reduces math and reading achievement outcomes, reduces educational engagement, and decreases social engagement. Washington (2017) notes that the chronic absenteeism can cause third graders unable to master reading, sixth graders failing courses and ninth graders dropping out of high school. Chronic absenteeism can devastate learning and is an early warning sign of academic trouble and dropping out of schools (Chang et al., 2014). Chronic absenteeism has been associated with increased likelihood to engage in criminal behavior, sexual risk behaviors, abuse of illicit substances, and dropout of school entirely (Henry et. al, 2009). A brief review of causes, course and treatment in relation to chronic absenteeism appears in Kiani et al. (2018). Jordan (2019) introduces a sample of practical strategies and intervention recommendations to local policymakers and educators for reducing chronic absenteeism, including a substantial sample of the leading work and latest thinking on improving attendance.

Chronic absenteeism is a complex national problem, and despite available literatures addressing causes, impacts and

interventions, dealing with the chronic absenteeism is a big challenge unless appropriate measures are taken at early stages. This proposition implies the need of studying the nonchronic absenteeism so as to provide a safeguard against the pervasive and persuasive form of the chronic absenteeism. However, as of now, no evidence of research seems to exist targeting nonchronic absenteeism in the United States at K-8 school levels. The proper understanding of the nonchronic absenteeism at K-8 schools at early stages and its onset would be of a great use in developing and implementing intervention strategies for reducing nonchronic absenteeism, which eventually would contribute in the reduction of chronic absenteeism. Therefore, the aim of this article is to investigate K-8 nonchronic absenteeism in the United States in detail by descriptive analyses and model-based analyses of a nationally representative survey database.

2. Methods

Using the National Household Education Surveys Program (2016), we evaluate the effect of various potential factors on nonchronic absenteeism of K-8 students in the United States. In this section, we provide a brief description of the data resource, potential factors and statistical data analysis methods in relation to K-8 students' nonchronic absenteeism in the United States.

2.1 Data Resource

This study consists of a sample of 8,188 students with nonchronic absenteeism in K-8 schools in the United States. The sample has been derived from the National Household Education Surveys (NHES) Program public use data, which is a significant data resource in relation to K-12 students in the United States. We utilize this data for exploring the significant factors of nonchronic absenteeism of K-8 students because we believe that the reduction of K-8 students nonchronic absenteeism contributes to the reduction of chronic absenteeism automatically. Below we describe response and related factors of interest in this study:

Response: The response is a discrete random variable denoting the nonchronic absenteeism of K-8 students in the United States, taking values 0, 1, …, 17 for qualifying to be nonchronic absenteeism as per U.S. Department of Education (2016a).

Socio-demographic factors: The following socio-demographic factors are being assessed in relation to nonchronic absenteeism of K-8 students:

Table 1. Socio-demographic factors

Socio-demographic factors in NHES data	Renamed Factors	Meaning	Values
CSEX	Gender	Gender of the child	1=Male, 2=Female
RACEETHN	Ethnicity	Ethnicity of the child	1=White, 2=Black, 3=Hispanic, 4=Others
PARGRADEX	Peduc	Parents education	1=Less than HS, 2=HS or equivalent, 3=Vocational /Technical, 4=College graduates, 5=Graduate or professional
-	Pov	Poverty	1=Poor, 2=nonpoor

The variable poverty (Pov) in Table 1 is defined and complied using total household members and total household income variables using the algorithm in McQuiggan and Megra (2017) and Hanson and Pugliese (2020).

Parental involvement types: A set of parent and family involvement (PFI) factors are being assessed for possible relationship with K-8 students' nonchronic absenteeism. The PFI factors have the following meanings and values:

Table 2. Parent and family involvement (PFI) factors

PFI Factors in NHES data	Renamed PFI factors	Meaning	Valid value
FSSPORTX	Ase	Attend a school event	1=Yes, 2=No
FSVOL	Vol	Serve as a volunteer	1=Yes, 2=No
FSMTNG	Asm	Attend a school meeting	1=Yes, 2=No
FSPTMTNG	Aptm	Attend a parent - teacher organization meeting	1=Yes, 2=No
FSATCNFN	Aptc	Attend parent - teacher conference	1=Yes, 2=No
FSFUNDRS	Fundr	Participate in fundraising	1=Yes, 2=No
FSCOMMTE	Sosc	Serve on school committee	1=Yes, 2=No
FSCOUNSLR	Mwgc	Meet with guidance counselor	1=Yes, 2=No

2.2 Test of Percent Discrepancies of Subjects Due to Factors

Initially, we test for the percent or proportion discrepancies of subjects due to underlying factors across groups by setting following null and alternative hypotheses:

H_0: *There is no percent discrepancy among subjects due to different groups of an underlying factor.*

H₁: *Percent discrepancy exists among subjects due to different groups of an underlying factor.*

The Chi-Square test of homogeneity of proportion or percent has been employed for each of the potential factors in regard to the specified hypotheses, by implementing SAS software.

2.3 Test of Median Absenteeism Due to Factors

We also test for the K-8 students' median nonchronic absenteeism discrepancies due to potential factors such as gender, ethnicity, parental education, poverty and various parental involvement factors. For discrepancies of an underlying factor across groups, we perform nonparametric Wilcoxon rank sum test (for comparing medians in two groups) and Kruskal-Wallis test (for comparing medians in more than two groups) as appropriate by implementing SAS software, for testing the following null and alternative hypotheses:

H_0: There is no difference in medians of K-8 students' nonchronic absenteeism across factor levels.

H₁: Differences do exist in medians of K-8 students' nonchronic absenteeism across factor levels.

2.4 Analysis of Factor Effects Using Regression Models

In order to evaluate if an underlying factor has any effect on K-8 students' nonchronic absenteeism, adjusted for other factors, we perform analyses using Poisson and Negative binomial regression models, along with the assessment of the goodness-of-fit of such models. Let Y_i (taking values 0,1,…,17) refer to the number of days of K-8 students' nonchronic absenteeism for individual i.

2.4.1 Poisson Regression

The Poisson regression is one the most popular models for count data, where mean and variance are equal. It assumes that an individual response Y_i, given the vector of covariates $\boldsymbol{x_i}$, is independently distributed as $Poisson(\mu_i)$ with probability mass function

$$P(Y_i = y_i | \boldsymbol{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}; y_i = 0, 1, \dots; \mu_i > 0.$$

where $\mu_i = E(Y_i|\boldsymbol{x}_i) = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_{k-1}X_{ik-1})$ is the mean incidence rate per unit of exposure time (i.e., a given academic year, in this article) and $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_{k-1})$ is a $k \times 1$ parameter vector with intercept β_0 and $k-1$ regressor coefficients $\beta_1, \cdots, \beta_{k-1}$. The parameters β_j, $j = 0, 1, 2, \cdots, k-1$, of the Poisson regression model are estimated by the maximum likelihood estimation (MLE) method by maximizing log-likelihood function

$$\mathcal{L}(\mu_i) = \sum_{i=1}^{n} log\left(\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}\right) = \sum_{i=1}^{n}\{y_i\boldsymbol{x}_i'\boldsymbol{\beta} - \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) - logy_i!\}$$

The goodness of fit of the model is evaluated using the chi-squared statistic given by $X^2 = \sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$, where

$\hat{\mu}_i = \exp(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}) = \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_{k-1}X_{ik-1})$ and the statistic X^2 has an approximately chi-square distributions with $n - k$ degrees of freedom (DF). While the mean and variance of Poisson distribution are equal, theoretically, often with real-life data variance is greater than the mean, and the situation is said to have an

overdispersion estimated by $\hat{\phi} = \frac{1}{n-k}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$. Details about the Poisson regression method, parameter estimation

procedure and goodness of fit of the model are available in Cameron and Trivedi (2013), NCSS Statistical Software (2018a) and SAS/ETS® 15.1 User's Guide (2018a). In this article, we estimate parameters and goodness of fit statistics by implementing SAS proc genmod procedure (SAS/ETS® 15.1 User's Guide, 2018a). Given a continuous predictor X_j, if $\hat{\beta}_j > 0, j = 1, 2, \cdots, k-1$, then the mean incidence rate for a unit increase of X_j value will be e^{β_j} times the mean incidence rate when $X_j = 0$. If X_j is a dichotomous variable, and $\beta_j > 0$, then the mean response of a given

level is e^{β_j} times the mean response at base level. However, if $\hat{\beta}_j < 0$, it will have reverse interpretation.

2.4.2 Negative Binomial Regression

The negative binomial regression model is a generalized Poisson regression model with a random heterogeneity terms τ_i, independent of the vector of regressors $\boldsymbol{x}_i$, for individual i. This leads to the fact that $Y_i | \boldsymbol{x}_i, \tau_i$ follows a Poisson distribution

$$f(Y_i|\boldsymbol{x}_i, \tau_i) = \frac{e^{-\mu_i \tau_i}(\mu_i \tau_i)^{y_i}}{y_i!} , y_i = 0, 1, ..$$

with mean rate $\mu_i \tau_i$, where τ_i has the density $g(\tau_i)$ such that $f(Y_i|\boldsymbol{x}_i) = \int_0^\infty f(Y_i|\boldsymbol{x}_i, \tau_i) \, g(\tau_i) d\tau_i$.

In practice, the distribution of τ_i is assumed to be a gamma $G\left(\frac{1}{\theta}, \theta\right)$ distribution with the density $g(\tau_i) =$

$\dfrac{(\tau_i)^{\frac{1}{\theta}-1}\left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} e^{-\tau_i/\theta}}{\Gamma\left(\frac{1}{\theta}\right)}$ and mean $E(\tau_i) = 1$ to allow constant (mean) term in the model. Then, it follows that

$$f(Y_i|x_i) = \int_0^\infty \frac{e^{-\mu_i \tau_i}(\mu_i \tau_i)^{y_i}}{y_i!} \frac{(\tau_i)^{\frac{1}{\theta}-1}\left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} e^{\frac{-\tau_i}{\theta}}}{\Gamma\left(\frac{1}{\theta}\right)} d\tau_i$$

By re-arranging the terms to match a gamma distribution, it follows that

$$f(Y_i|x_i) = \frac{\mu_i^{y_i}\left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} \Gamma\left(y_i + \frac{1}{\theta}\right)}{y_i! \; \Gamma\left(\frac{1}{\theta}\right) \left(\mu_i + \frac{1}{\theta}\right)^{y_i+\frac{1}{\theta}}} \int_0^\infty \frac{(\tau_i)^{y_i+\frac{1}{\theta}-1} \, e^{-\left(\mu_i+\frac{1}{\theta}\right)\tau_i}\left(\mu_i + \frac{1}{\theta}\right)^{y_i+\frac{1}{\theta}}}{\Gamma\left(y_i + \frac{1}{\theta}\right)} d\tau_i$$

By the property of gamma distribution, the integral of the equation is 1, and it follows that

$$f(Y_i|x_i) = \frac{\Gamma(y_i + \theta^{-1})}{y_i! \; \Gamma(\theta^{-1})}\left(\frac{\mu_i}{\mu_i + \theta^{-1}}\right)^{y_i} \left(\frac{\theta^{-1}}{\mu_i + \theta^{-1}}\right)^{\theta^{-1}} ; y_i = 0, 1, 2, ...$$

Or, equivalently,

$$f(Y_i|x_i) = \frac{\Gamma(y_i + \theta^{-1})}{y_i! \; \Gamma(\theta^{-1})}\left(\frac{\theta\mu_i}{\theta\mu_i + 1}\right)^{y_i} \left(\frac{1}{\theta\mu_i + 1}\right)^{\theta^{-1}} ; y_i = 0, 1, 2, ...$$

which is negative binomial distribution with

$$E(Y_i|\boldsymbol{x}_i) = \theta^{-1}\left(\frac{\theta\mu_i}{\theta\mu_i+1}\right) / \left(\frac{1}{\theta\mu_i+1}\right) = \mu_i$$

And

$$V(Y_i|\boldsymbol{x}_i) = \theta^{-1}\left(\frac{\theta\mu_i}{\theta\mu_i+1}\right) / \left(\frac{1}{\theta\mu_i+1}\right)^2 = \mu_i(1 + \mu_i\theta) = \mu_i + \mu_i^2\theta > E(Y_i|\boldsymbol{x}_i).$$

The distribution of $Y_i | \boldsymbol{x}_i$ obtained by the mixture of Poisson and gamma distribution is popular due to the fact that it allows Poisson heterogeneity in modeling count or discrete data. Relating the mean $E(Y_i|\boldsymbol{x}_i) = \mu_i$ to a set of $k - 1$ predictor variables by the expression $\mu_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_{k-1}X_{ik-1})$, the parameters θ and β_j, $j = 0, 1, 2, \cdots, k - 1$, of the negative binomial regression model are estimated by the MLE method by maximizing log-likelihood function

$$\mathcal{L}(\mu_i) = \sum_{i=1}^{n}\left\{log\left(\frac{\Gamma(y_i + \theta^{-1})}{\Gamma(\theta^{-1})}\frac{1}{y_i!}\left(\frac{\theta\mu_i}{\theta\mu_i + 1}\right)^{y_i}\left(\frac{1}{\theta\mu_i + 1}\right)^{\theta^{-1}}\right)\right\}$$

By noting the fact that $\frac{\Gamma(y_i+\theta^{-1})}{\Gamma(\theta^{-1})} = \prod_{j=0}^{y_i-1}(j+\theta^{-1})$ and re-arranging other terms, it appears that

$$\mathcal{L}(\mu_i) = \sum_{i=1}^{n}\left\{\sum_{j=0}^{y_i-1}\log(j+\theta^{-1}) - \log(y_i!) - (y_i+\theta^{-1})\log(\theta\mu_i+1) + y_i\log\theta + y_i\log\mu_i\right\}$$

For evaluating the goodness of fit of the negative binomial regression model, one can use the deviance statistic

$D = 2\{\mathcal{L}(y_i) - \mathcal{L}(\mu_i)\} = 2\sum_{i=1}^{n}\left\{y_i\log\left(\frac{y_i}{\mu_i}\right) - (y_i+\theta^{-1})\log\left(\frac{\theta y_i+1}{\theta\mu_i+1}\right)\right\}$ or the chi-squared statistic $X^2 = \sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i}$.

Details about the negative binomial regression method, parameter estimation procedure and goodness of fit of the models are available in Cameron and Trivedi (2013), NCSS Statistical Software (2018b) and SAS/ETS® 15.1 User's Guide (2018b). In this article, we estimate parameters and goodness of fit statistics of negative binomial regression by implementing SAS proc genmod procedure (SAS/ETS® 15.1 User's Guide, 2018b)

3. Results and Discussions

In Table 3, we present frequency (f) and percent (%) of subjects due to different groups of underlying factors, and the value of Chi-Squared test statistic (Chisq) along with degrees of freedom (DF) and p-value for the significance of tests specified in 2.2.

Table 3. Percent discrepancies of subjects due to labels of underlying factors

Factors	Labels	f	%	Chisq	DF	p-value
Gender	1	4230	51.66	9.0	1	**0.0026**
	2	3958	48.34			
Ethnicity	1	4554	55.62	4340.2	3	**<.0001**
	2	800	9.77			
	3	1771	21.63			
	4	1063	12.98			
Peduc	1	472	5.76	1959.2	4	**<.0001**
	2	879	10.74			
	3	2388	29.16			
	4	2330	28.46			
	5	2119	25.88			
Poverty	1	1208	14.75	4068.9	1	**<.0001**
	2	6980	85.25			
Ase	1	6891	84.16	3821.8	1	**<.0001**
	2	1297	15.84			
Volunteering	1	4016	49.05	3.0	1	0.0847
	2	4172	50.95			
Asm	1	7347	89.73	5169.5	1	**<.0001**
	2	841	10.27			
Aptm	1	4072	49.73	0.2	1	0.6268
	2	4116	50.27			
Aptc	1	6936	84.71	3945.8	1	**<.0001**
	2	1252	15.29			
Fundr	1	5441	66.45	886.4	1	**<.0001**
	2	2747	33.55			
Sosc	1	1122	13.7	4315.0	1	**<.0001**
	2	7066	86.3			
Mwgc	1	2309	28.2	1556.5	1	**<.0001**
	2	5879	71.8			

On the basis of results presented in Table 3, one may conclude that discrepancies of the percent of subjects due to factor labels are statistically significant for all underlying factors except for volunteering (Vol) and attending parent-teacher organization meeting (Aptm).

The results of test specified in 2.3 have been reported for each of the underlying factors in Table 4. The results include means, variances (Vars), standard error of means (SEMs) and medians (Meds) for K-8 students' nonchronic absenteeism

for different groups of underlying factors. It also reports value of chi-squared test statistic for comparing median absenteeism in two groups via nonparametric Wilcoxon rank sum test (equivalent to Mann-Whitney U test) or more than two groups via Kruskal-Wallis (chi-squared) test, along with degrees of freedom (DF) and p-value for the significance of the test.

Table 4. Results of K-8 students' median absenteeism discrepancy due to factors

Factors	Labels	Means	Vars	SEMs	Meds	Chisq	DF	p-value
Gender	1	3.47	10.1	0.049	3	0.8442	1	0.3582
	2	3.49	9.8	0.050	3			
Ethnicity	1	3.66	9.6	0.046	3	96.9646	3	**<.0001**
	2	2.87	8.8	0.105	2			
	3	3.52	11.0	0.079	3			
	4	3.11	10.0	0.097	2			
Peduc	1	3.30	11.0	0.153	2	64.2828	4	**<.0001**
	2	3.90	12.2	0.118	3			
	3	3.81	11.1	0.068	3			
	4	3.34	8.9	0.062	3			
	5	3.13	8.3	0.063	2			
Poverty	1	3.89	12.4	0.101	3	13.0034	1	**0.0003**
	2	3.41	9.5	0.037	3			
Ase	1	3.47	9.6	0.037	3	0.2652	1	0.6065
	2	3.57	11.6	0.095	3			
Volunteering	1	3.47	9.3	0.048	3	1.2968	1	0.2548
	2	3.49	10.6	0.050	3			
Asm	1	3.43	9.7	0.036	3	11.357	1	**0.0008**
	2	3.92	12.1	0.120	3			
Aptm	1	3.31	9.5	0.048	3	29.7638	1	**<.0001**
	2	3.66	10.3	0.050	3			
Aptc	1	3.47	9.8	0.038	3	0.2538	1	0.6144
	2	3.57	10.7	0.092	3			
Fundr	1	3.41	9.5	0.042	3	4.6366	1	**0.0313**
	2	3.62	10.7	0.062	3			
Sosc	1	3.32	8.8	0.088	3	1.257	1	0.2622
	2	3.51	10.1	0.038	3			
Mwgc	1	3.75	11.9	0.072	3	11.6138	1	**0.0007**
	2	3.38	9.1	0.039	3			

Results presented in Table 4 provide evidence that K-8 students' median nonchronic absenteeism are statistically significant due to ethnicity, parental education (Peduc), poverty, attending school meeting (Asm), attending parent teacher organization meeting (Aptm), fundraising (Fundr) and meeting with guidance counselor (Mwgc).

Note that variances (Vars) of K-8 students' nonchronic absenteeism are higher than that of the means (Means) due to all factors, which suggest that negative binomial regression would be a better fit compared to Poisson regression for modeling K-8 students' nonchronic absenteeism given the predictors.

In Tables 5.1a and 5.1b, we report results of assessment of goodness of fit and label-specific significance of each factor due to Poisson regression model. In Tables 5.2a and 5.2b, we report results of assessment of goodness of fit and label-specific significance of each factor due to negative binomial regression model. In Table 6, we report results of comparative analyses for the assessment of goodness of fit in modeling K-8 students' nonchronic absenteeism due to Poisson and negative binomial regression models.

Table 5.1a. Assessment of Goodness of fit of Poisson regression model

Criterion	DF	Value	Value/DF
Deviance	8170	22552.7	**2.76**
Scaled Deviance	8170	8170	1.00
Pearson Chi-Square	8170	22700.7	**2.78**
Scaled Pearson X2	8170	8223.6	1.01
AIC (smaller is better)		43766.7	
BIC (smaller is better)		43892.9	

Table 5.1b. Label-specific factor effect on K-8 nonchronic absenteeism using Poisson regression model

Parameter	Label	DF	Estimate	SE	Wald 95% CI		Chisq	p-value
					LCL	UCL		
Intercept		1	1.365	0.046	1.274	1.456	870.7	<.0001
Gender	1	1	-0.016	0.020	-0.055	0.023	0.7	0.4135
	Ref=2							
Ethnicity	2	1	-0.304	0.038	-0.379	-0.229	63.5	**<.0001**
	3	1	-0.075	0.026	-0.126	-0.023	8.0	**0.0046**
	4	1	-0.146	0.032	-0.209	-0.083	20.9	**<.0001**
	Ref=1							
Peduc	1	1	0.013	0.053	-0.091	0.116	0.1	0.8116
	2	1	0.190	0.038	0.116	0.264	25.3	**<.0001**
	3	1	0.181	0.028	0.126	0.236	41.7	**<.0001**
	4	1	0.059	0.028	0.004	0.113	4.4	**0.0358**
	Ref=5							
Poverty	2	1	-0.128	0.030	-0.186	-0.069	18.3	**<.0001**
	Ref=1							
Ase	2	1	-0.013	0.030	-0.072	0.046	0.2	0.6627
	Ref=1							
Volunteering	2	1	-0.065	0.023	-0.110	-0.019	7.8	**0.0053**
	Ref=1							
Asm	2	1	0.095	0.034	0.029	0.161	7.9	**0.0049**
	Ref=1							
Aptm	2	1	0.091	0.022	0.048	0.133	17.6	**<.0001**
	Ref=1							
Aptc	2	1	-0.009	0.029	-0.065	0.048	0.1	0.765
	Ref=1							
Fundr	2	1	0.047	0.023	0.002	0.092	4.1	**0.0428**
	Ref=1							
Sosc	2	1	0.020	0.032	-0.043	0.083	0.4	0.5296
	Ref=1							
Mwgc	2	1	-0.125	0.022	-0.168	-0.082	32.4	**<.0001**
	Ref=1							
Scale			1.662	0.000	1.662	1.662		

Note. LCL=Lower confidence level and UCL= Upper confidence level.

Table 5.2a. Assessment of Goodness of fit of negative binomial regression model

Criterion	DF	Value	Value/DF
Deviance	8170	9361.7	**1.15**
Scaled Deviance	8170	9361.7	1.15
Pearson Chi-Square	8170	8066.8	**0.99**
Scaled Pearson X2	8170	8066.8	0.99
AIC (smaller is better)		**38205.5**	
BIC (smaller is better)		**38338.7**	

Table 5.2b. Label-specific factor effect on K-8 nonchronic absenteeism using negative binomial regression model

Parameter	Label	DF	Estimate	SE	Wald 95% CI		Chisq	p-value
					LCL	UCL		
Intercept		1	1.372	0.048	1.278	1.465	829.9	<.0001
Gender	1	1	-0.015	0.020	-0.054	0.024	0.6	0.4557
	Ref=2							
Ethnicity	2	1	-0.306	0.037	-0.378	-0.233	67.9	**<.0001**
	3	1	-0.071	0.027	-0.124	-0.019	7.1	**0.0075**
	4	1	-0.144	0.032	-0.206	-0.082	20.6	**<.0001**
	Ref=1							
Peduc	1	1	0.009	0.053	-0.095	0.112	0.0	0.8711
	2	1	0.192	0.039	0.116	0.267	24.5	**<.0001**
	3	1	0.182	0.029	0.126	0.238	40.8	**<.0001**
	4	1	0.058	0.028	0.003	0.112	4.3	**0.0372**
	Ref=5							
Poverty	2	1	-0.133	0.031	-0.194	-0.071	17.9	**<.0001**
	Ref=1							
Ase	2	1	-0.017	0.031	-0.076	0.043	0.3	0.5888
	Ref=1							
Volunteering	2	1	-0.068	0.024	-0.114	-0.021	8.2	**0.0042**
	Ref=1							
Asm	2	1	0.092	0.035	0.023	0.161	6.8	**0.0091**
	Ref=1							
Aptm	2	1	0.088	0.022	0.045	0.131	15.9	**<.0001**
	Ref=1							
Aptc	2	1	-0.012	0.029	-0.069	0.046	0.2	0.6844
	Ref=1							
Fundr	2	1	0.045	0.024	-0.002	0.091	3.6	0.0596
	Ref=1							
Sosc	2	1	0.017	0.032	-0.047	0.080	0.3	0.6041
	Ref=1							
Mwgc	2	1	-0.119	0.023	-0.163	-0.075	27.7	**<.0001**
	Ref=1							
Dispersion		1	0.529	0.014	0.502	0.557		

Note that the Deviance and Pearson Chi-Squared statistics reported in Table 5.1a, both follow chi-squared distribution with DF=8170, and as such the Value/DF is expected to be close to 1 for the Poisson regression model to fit nonchronic absenteeism well. Since Value/DF for Deviance is 2.76 (>1), and the Value/DF for Pearson Chi-Square is 2.78 (>1), both suggesting that the K-8 students' nonchronic absenteeism is subject to over-dispersion. The overdispersion of K-8 students' nonchronic absenteeism resulting from the assessment of Poisson regression model agrees with results reported in Table 4, where variance of nonchronic absenteeism exceeds the mean of nonchronic absenteeism for each factor labels, suggesting over dispersion. Given over dispersion, the estimates of standard errors (SEs) from Poisson regression model are incorrect, and may result in an invalid chi-square test or inference. The two common fixes due to over dispersion are either to use a negative binomial regression or correct the estimated standard errors by scaled criterion. In SAS, the scaling is achieved by specifying the descale or scale=d option. This option forces the Scaled Deviance to be 1 by forcing Value/DF to be 1 (dividing Value/DF by itself). It also adjusts standard errors by a factor (correction factor), which is the square root of Value/DF. In this study, the scaling parameter or correction factor is equal to sqrt(2.76)=1.661325 (SAS reported value is 1.662). After scaling, the corrected SEs is equal to pre-scaling SEs*sqrt (of pre-scaling Value/DF).

In Table 5.1b, the adjusted standard errors have been reported for each factor labels and thus the test and inference made via confidence interval estimates are approximately correct. Based on results of Poisson regression analysis reported in Table 5.1b, it appears that ethnicity, poverty, volunteering, attending school meeting (Asm), attending parent teacher organization meeting (Aptm), participation in fundraising (fundr) and meeting with guidance counselor (Mwgc) have statistically significant impact on K-8 students' nonchronic absenteeism. It also appears that all parental education (Peduc) labels are statistically significant for K-8 students' nonchronic absenteeism except for students with parental educational label below high school.

In order to compare performance of scaled Poisson regression model compared to the alternative approach using negative binomial regression model, the results of negative binomial regression analyses have been reported in Tables 5.2a and 5.2b. From the results reported in Table 5.2a, it is evident that the Value/DF for Deviance is 1.15 (close to 1) and the Value/DF for Pearson Chi-Square is 0.99 (close to 1) for negative binomial regression model. Therefore, it may be concluded that negative binomial regression fits to K-8 students' nonchronic absenteeism well. Indeed, it appears that p-value for the Pearson Chi-Square test is 0.78975, which is obtained by computing probability of the area to the right of the Chi-Square value of 8066.8 with DF=8170. This p-value provides statistically significant evidence that the negative binomial regression model fits absenteeism of K-8 students well.

Regarding factor effects, based on results of negative binomial regression analysis reported in Table 5.2b, it appears that ethnicity, poverty, volunteering, attending school meeting (Asm), attending parent teacher organization meeting (Aptm) and meeting with guidance counselor (Mwgc) have statistically significant impact on K-8 students' nonchronic absenteeism. It also appears that all parental education (Peduc) labels are statistically significant for K-8 students' absenteeism except for parents with educational label below high school. Based on the results of Tables 5.1b and 5.2b, it appears that both Poisson regression model with scaling due to over dispersion and negative binomial regression models provide identical conclusion in regarding the significance of factors affecting nonchronic absenteeism of K-8 students in the United States, except for Poisson model identifies fundraising as a significant factor. For significant factors, the p-values differ insignificantly between two models.

In order to decide which of the two regression models fits better to the K-8 students' nonchronic absenteeism, let us take a look at the comparative analyses of goodness-of-fit statistics due to Poisson and negative binomial regression models as appears in 6.

Table 6. Comparison of Goodness of fit Models

	Goodness of Fit of PR model			Goodness of Fit of NBR model		
Criterion	DF	Value	Value/DF	DF	Value	Value/DF
Deviance	8170	22552.7	2.76	8170	9361.7	**1.15**
Scaled Deviance	8170	8170	1.00	8170	9361.7	1.15
Pearson Chi-Square	8170	22700.7	2.78	8170	8066.8	**0.99**
Scaled Pearson X2	8170	8223.6	1.01	8170	8066.8	0.99
AIC (smaller is better)		43766.7			**38205.5**	
BIC (smaller is better)		43892.9			**38338.7**	

Note. PR=Poisson regression and NBR=Negative binomial regression.

As reported in Table 6, the Deviance/DF or Pearson Chi-square/DF is closer to 1 for negative binomial regression than is for Poisson regression, suggesting that negative binomial regression is a better fit compared to Poisson regression. Identical conclusion can be reached on the basis of AIC (Akaike information criteria) and BIC (Bayesian information criteria) in regarding the goodness of fit of negative binomial regression to K-8 students' nonchronic absenteeism.

4. Concluding Remarks

Many studies address chronic absenteeism of K-12 students in the United States. The chronic absenteeism has been termed as an unidentified problem or hidden crisis (Bruner et al., 2011; Education Commission of the States, 2010; U.S. Department of Education, 2016b). Many articles address adverse impact in students' life and society chronic absenteeism could have (Chang et al., 2014; Gershenson et al., 2017; Gottfried, 2014; Henry et al., 2009; Jordan, 2019; Kiani et al., 2018; National Association of Elementary School Principals, 2016; Washington, 2017), including causes (Jordan, 2019; Kiani et al., 2018; National Association of Elementary School Principals, 2016), and interventions (Allen et al., 2018; Robinson et al., 2018; Rogers et al, 2018) in reducing chronic absenteeism. While all these studies relate to chronic absenteeism, none of them addresses nonchronic absenteeism. We believe that paying attention to nonchronic absenteeism and of its reduction would ultimately contribute to the reduction of the chronic absenteeism. In particular, an adequate understanding of the factors affecting nonchronic absenteeism at early stages and its onset is important for setting up interventions and strategies targeting the reduction of nonchronic absenteeism. Given this proposition, in this study, we utilize a nationally representative survey database due to NHES to study K-8 students' nonchronic absenteeism in the United States.

As appears in this study, negative binomial regression fits nonchronic absenteeism of K-8 students in the United States better as compared to the Poisson regression, given the sets of predictors in this study. It is noted that ethnicity, parental education, poverty, volunteering, attending school meeting, attending parent teacher organization meeting and meeting with guidance counselor are significant factors affecting K-8 students' nonchronic absenteeism in the United States due to negative binomial regression analysis. The Poisson regression model suggests that fundraising is also a significant

factor affecting K-8 students' nonchronic absenteeism, in addition to other factors found significant using negative binomial regression. While there has been no significant evidence of research addressing nonchronic absenteeism, the findings of this study are expected to contribute in providing a safeguard against the pervasive and persuasive form of the chronic absenteeism in the United States given the implementation of intervention strategies at early stages and its onset. Given the findings of this study, we believe that interventions addressing differences in factors such as ethnicity, poverty, parental education, encouraging and ensuring parental involvement in education effectively could have an important role in reducing nonchronic absenteeism.

Acknowledgements

References

Allen, C. W., Diamond-Myrsten, S., & Rollins, L. K. (2018). School absenteeism in children and adolescents. *American Family Physician*, *98*(12), 738-744.

Bauer, L., Liu, P., Schanzenbach, D. W., & Shambaugh, J. (2018). *Reducing chronic absenteeism under the every student succeeds act*. The Hamilton Project. Brookings.

Bruner, C., Discher, A., & Chang, H. (2011). *Chronic elementary absenteeism: A problem hidden in plain sight*. A Research Brief from Attendance Works and Child & Family Policy Center.

Cameron & Trivedi (2013). *Regression analysis of count data*, Second Edition, Cambridge University Press.

Chang, H. N., Gomperts, J., & Boissiere, L. (2014, October 07). *Chronic absenteeism can devastate K-12 learning*. Education Week.

Chang H. N., & Romero, M. (2008, September). *Present, engaged, and accounted for: The critical importance of addressing chronic absence in the early grades*. National Center for Children in Poverty. Retrieved from https://www.nccp.org/wpcontent/uploads/2008/09/text_837.pdf

Education Commission of the States (2010, February). Chronic early absence. *The Progress of Education Reform*, *11*(1). Retrieved from https://files.eric.ed.gov/fulltext/ED544564.pdf

Gershenson, S., Jacknowitz, A., & Brannegan, A. (2017). Are student absences worth the worry in U.S. primary schools? *Education Finance and Policy, 12*(2), 137-165. https://doi.org/10.1162/EDFP_a_00207

Gottfried, M. A. (2010) Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal*, *47*(2), 434-465. https://doi.org/10.3102/0002831209350494

Gottfried, M. A. (2011). The detrimental effects of missing school: Evidence from urban siblings. *American Journal of Education*, *117*(2), 147-182. https://doi.org/10.1086/657886

Gottfried, M. A. (2014). Chronic absenteeism and its effects on students' academic and socioemotional outcomes. *Journal of Education for Students Placed at Risk, 19*(2), 53–75. https://doi.org/10.1080/10824669.2014.962696

Hanson, R., & Pugliese, C. (2020). *Parent and family involvement in education*: *2019* (NCES 2020-076). U.S. Department of Education. National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020076.

Henry, K. L., Thornberry, T. P., & Huizinga, D. H. (2009). A discrete-time survival analysis of the relationship between truancy and the onset of marijuana use. *Journal of Studies on Alcohol and Drugs*, *70*(1), 5-15. https://doi.org/10.15288/jsad.2009.70.5

Jordan, P. (2019, July). *Attendance playbook: smart solutions for reducing chronic absenteeism*. FutureEd and Attendance Works. Retrieved from https://cut.ly/It4cPk3

Kiani, C., Otero, K., Taufique, S., & Ivano, I. (2018). Chronic absenteeism: A brief review of causes, course and treatment. *Adolescent Psychiatry*, *8*(3), 214-230. https://doi.org/10.2174/2210676608666180709155116

McQuiggan, M., & Megra, M. (2017). *Parent and family involvement in education: Results from the National Household Education Surveys Program of 2016 (NCES 2017-102)*. National Center for Education Statistics. https://nces.ed.gov/pubs2017/2017102.pdf

National Association of Elementary School Principals (2016). Six Causes—and Solutions—for Chronic Absenteeism. *Communicator*, *40*(1). Retrieved from https://www.naesp.org/resource/six-causes-and-solutions-for-chronic-absenteeism/

National Household Education Surveys Program (2016). *National household education surveys program of 2016 data file user's manual*. U.S. Department of Education. https://nces.ed.gov/pubs2018/2018100.pdf

NCSS Statistical Software (2018a). *Poisson Regression*. Chapter 325. Retrieved from https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Poisson_Regression.pdf

NCSS Statistical Software (2018b). *Negative Binomial Regression*. Chapter 326. Retrieved from https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Negative_Binomial_Regression.pdf

Reid, K. (2005). The causes, views and traits of school absenteeism and truancy: An analytical review. *Research in Education*, *74*(1), 59-82. https://doi.org/10.7227/RIE.74.6

Reid, K. (2008). The causes of non-attendance: An empirical study. *Educational Review*, *60*(4), 345-357. https://doi.org/10.1080/00131910802393381

Robinson, C. D., Lee, M. G., Dearing, E., & Rogers, T. (2018). Reducing student absenteeism in the early grades by targeting parental beliefs. *American Educational Research Journal*, *26*(3), 353-383. https://doi.org/10.3102/0002831218772274

Rogers, T., & Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour, 2*(5), 335-342. https://doi.org/10.1038/s41562-018-0328-1

SAS/ETS® 15.1 User's Guide (2018a). *The countreg procedure: Poisson regression*. SAS Institute Inc. Retrieved from https://support.sas.com/documentation/onlinedoc/ets/141/countreg.pdf

SAS/ETS® 15.1 User's Guide (2018b). *The countreg procedure: Negative binomial regression*. SAS Institute Inc. Retrieved from https://support.sas.com/documentation/onlinedoc/ets/141/countreg.pdf

U.S. Department of Education (2016a, August 23). *New research shows nearly half of American parents underestimate the harm of school absences*. Retrieved from https://cutt.ly/TnIQBJT

U.S. Department of Education (2016b). *Chronic absenteeism in the nation's schools: A hidden educational crisis*. Retrieved from https://www2.ed.gov/datastory/chronicabsenteeism.html

Washington, V. (2017, September 6). *Student absence leads to academic failure*. Michigan State University Extension. Retrieved from https://www.canr.msu.edu/news/student_absence_leads_to_academic_failure