

| Honorable Mentioned | |
|----------------------------|--|
| Authors | Tarah Cole (coletl2@miamioh.edu) Department of Mathematics & Statistics, Northern Kentucky University 615 South Union St., OH, 45106 |
| Title | Assessment of Risk Factors for Truancy of Children in Grades K-12 Using Survival Analysis |
| Type of Project | Capstone Research Project |
| Instructor Sponsor | Joseph Nolan (nolanj1@nku.edu) Department of Mathematics & Statistics, Northern Kentucky University |
| Abstract | Survival Analysis is a time-to-event statistical analysis commonly used to assess risk of particular events based on available predictors. In this study, survival analysis is implemented to evaluate risk factors pertaining to truancy using data obtained from a large Kentucky school district. Truancy has been identified as an early warning sign that youth are headed for potential delinquent activity. By understanding what factors increase or decrease a students' risk of becoming truant, schools can allocate resources to the high risk groups and take necessary measures to lessen truancy rates which may in turn improve later quality of life. |

Assessment of Risk Factors for Truancy of Children in Grades K-12 Using Survival Analysis

Tarah Cole

Abstract

Survival Analysis is a time-to-event statistical analysis commonly used to assess risk of particular events based on available predictors. In this study, survival analysis is implemented to evaluate risk factors pertaining to truancy using data obtained from a large Kentucky school district. Truancy has been identified as an early warning sign that youth are headed for potential delinquent activity. By understanding what factors increase or decrease a students' risk of becoming truant, schools can allocate resources to the high risk groups and take necessary measures to lessen truancy rates which may in turn improve later quality of life.

Introduction

Although truancy can be defined differently from state to state, it is a national issue in educational systems. By law, all students are required to attend school every day; districts must record any absences as excused or unexcused. Laws vary from state to state. This study defines truancy according to the Kentucky School Law; KRS 159.50 states, "Any child who has attained the age of six years but has not reached his or her eighteenth birthday, who has been absent without a valid excuse for three or more days, or [who has been] tardy without a valid excuse on three or more days is truant. Any student enrolled in a public school who has attained the age of eighteen years but has not reached his or her twenty-first birthday, who has been absent from school without a valid excuse for three or more days, or [who has been] tardy on three or more days is truant." ¹

Preliminary findings from the Office of Juvenile Justice and Delinquency Prevention's (OJJDP) evaluation of Truancy Reduction Demonstration Program (TRDP) confirm that truancy is correlated to family and school factors, economic influences and student variables (such as mental health). Also according to a U.S. Department of Justice Bulletin, "It is also known that truancy has been clearly identified as one of the early warnings signs that youth are headed for potential delinquent activity, social isolation, and/or educational failure. Decades of research have also identified a link between truancy and later violence, marital and job problems, adult criminality and incarceration" ²

The purpose of this study is to investigate factors that influence truancy. By understanding what factors increase or decrease a student's risk of becoming truant, schools can allocate resources to the high risk groups and take necessary measures. In doing so, schools can lessen truancy rates thus improving the student's education and may in turn improve later quality of life.

Methods

The study of truancy in grades K-12 applies survival analysis, a time-to-event statistical analysis commonly used to assess risk of particular events based on available predictors. This method is an innovative approach that will help schools to not only identify high risk groups, but also allow for the comparison of the associated risks with other groups. Cox proportional hazard models were used to assess the risk factors affecting truancy. These models examine the difference in the survival curves while also accounting for covariates. The use of survival analysis requires a start date, end date, censoring status and survival time. For this study, start date is defined as the student's first day of school within the district and the end date is defined as the date on which a student either becomes truant is censored. Students are censored who finish the school year without becoming truant, move to another school district, begin a home school/online school program, drop out of school, or encounter any other circumstance resulting in their removal from school within the district. The survival time is measured as the total days between the start and end date.

This is a retrospective analysis using data obtained from a large midwestern school district (IRB 12-064). The sample consisted of 16,418 eligible students from a total of twenty-one

schools (twelve elementary, five middle and four high schools) for the 2009-2010 school year. Forty-eight students were excluded due to inconclusive start dates and/or truancy dates. All analysis was conducted using SAS 9.2.

Factors that may affect truancy include: age of student, number of schools changes within the year, special education status, socio economic status (SES), school, and ethnicity, language and limited English proficiency (LEP) status. Since some of these variables (SES, special education status, and LEP status) may change during the school year, all students were classified by their most extreme status within the school year. For example, students having free lunch status at any point during the year would never be classified as paying for their lunch.

Socio economic status was determined by two variables, lunch status and homelessness status. Lunch status consisted of paid, reduced and free lunch status. If a student was homeless that student was categorized as homeless regardless of lunch status. When combined the SES variable was taken to be an ordinal variable having four levels: 0=paid, 1=reduced, 2=free and 3=homeless. Within the sample, 67.91% had paid status, 5.69% had reduced status, 25.34% had free status and 1.05% of students were known to be homeless. Among other factors, language had two levels, English and non-English; Special Education was considered as having three levels, active, inactive, and none. Ethnicity was classified into seven categories: Caucasian, African American, Asian, Native Hawaiian or Other Pacific Islander, Hispanic, American Indian or Alaskan Native, or a person having two or more ethnicities.

The proportional hazards assumption necessary for use of the Cox Proportional Hazard model was examined by a test using Schoenfeld residuals.³ The residuals were plotted against the natural logarithm and square of total days. Pearson correlation coefficients were examined and there was no evidence to suggest than any of the potential variables violated the assumption.

The Cox model provides hazard ratios (HR) which is the ratio of two hazard functions as shown below. The baseline or reference hazard function is represented by $h_0(t)$.

$$HR = \frac{\text{Comparative Hazard}}{\text{Reference Hazard}} = \frac{h_0(t)^{(X_1\hat{\beta})}}{h_0(t)^{(X_0\hat{\beta})}}$$

Confidence intervals of the hazard ratios are used to determine a specific group's risk of truancy when compared to a reference group. Those intervals for HR which include one indicate no evidence of a difference in the hazards. If the comparative hazard is greater the confidence interval for the HR will be entirely greater than one; conversely a confidence interval for the HR that falls entirely below one indicates that the reference hazard is greater.

Results

During the 2009-2010 school year, there were 1,728 students (10.53%) who became truant within the district. Single predictor models indicated that age of student, number of schools changes, special education status, economic status, school and ethnicity were individually significant at the $\alpha=0.05$ level. Language and Limited English Proficiency (LEP) did not show evidence of importance. Table 1 below provides a 95% Confidence Intervals on the Hazard Ratios for these models. For ethnicity, Caucasian was used as reference group as the vast majority of students classify themselves as Caucasian.

Table1: Single predictor analyses

| Predictor | Hazard Ratio(95% CI)* | n |
|---|----------------------------|--------|
| Economic Status | Reference Group: Paid | 11,150 |
| Reduced | (1.51 , 2.23) | 935 |
| Free | (2.76 , 3.36) | 4,160 |
| Homeless | (3.71 , 6.72) | 173 |
| School Change | Reference Group: No | 16,136 |
| Yes | (1.49 , 2.57) | 282 |
| Special Ed Status | Reference Group: None | 12,452 |
| Active | (1.06 , 1.35) | 2720 |
| Inactive | <i>NS</i> | 1246 |
| Age | | N\A |
| Per one year increase | (1.14 , 1.18) | N\A |
| Ethnicity | Reference Group: Caucasian | 14,489 |
| Hispanic | (1.38 , 2.00) | 740 |
| American Indian or Alaska Native | (1.59 , 6.39) | 26 |
| Asian | (0.20 , 0.59) | 379 |
| Black or African American | (1.07 , 1.73) | 515 |
| Native Hawaiian or Other Pacific Islander | <i>NS</i> | 27 |
| A person with more than one ethnicity | <i>NS</i> | 241 |

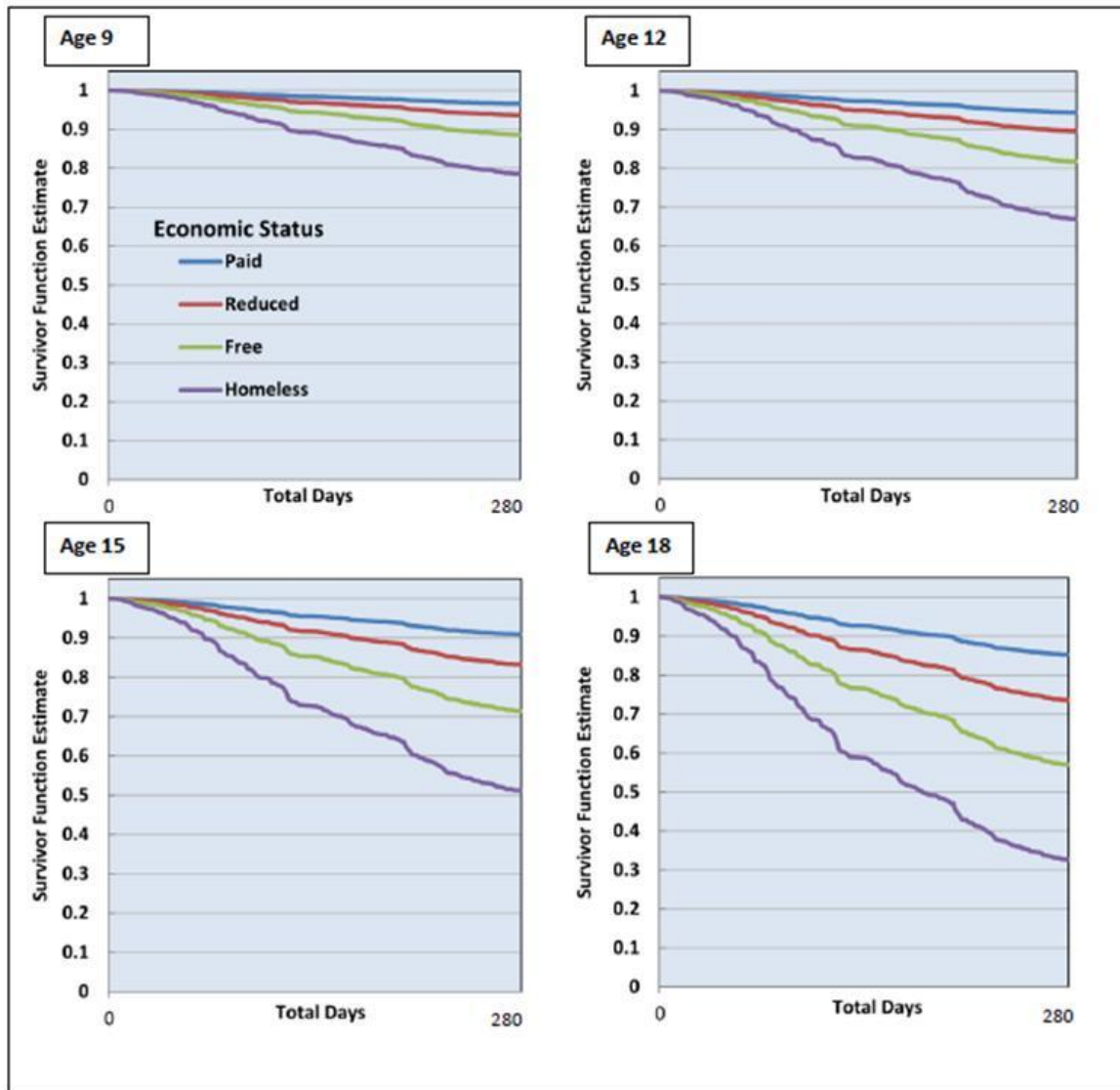
*NS = No evidence of a significant difference from the reference group

Results for economic status indicate that with 95% confidence, students who have a reduced lunch status are 1.51 to 2.23 times more likely to become truant, students who have a free lunch status are 2.76 to 3.36 times more likely to become truant and homeless students are 3.71 to 6.72 times more likely to become truant than students who have a paid lunch status. Also with 95% confidence, the risk of truancy increases by around 16% for every year in age of the student. In advancing from the kindergarten to twelfth grade, a student's risk of truancy is increased by a factor of almost 6.

Age and SES were examined in a two-factor interaction. The interaction effect showed significance (p-value = 0.0015), but main effects for age and SES were also evident (p-values < 0.0001). The largest effect is that of SES, as students receiving free lunch, or who are homeless, are at substantially increased risk of truancy. Truancy risk also increased with age, and the interaction implied that lower SES exacerbates the increased risk. The interaction confirms that the reference (paid) and comparative (reduced, free, homeless) survival curves have a different proportionality constant for each age; the constant of proportionality generally seems to increase as age increases. However, the increase in proportions is relatively small compared to main effects; the main effects of age and SES appear to overshadow the interaction.

Kaplan-Meier curves were created to illustrate these key risk factors, age and SES. It can be seen as a student ages they have a higher risk of becoming truant but within each age group the risks for those of a particular economic status also change. The survival curves remain in the same order (paid, reduced, free, homeless) for every age, but as the age increases they become more widespread. Figure 1 includes Kaplan-Meier curves for ages nine, 12, 15 and 18.

Figure1: Time to truancy by economic status by age



Additional results from Table 1 indicate that students who change schools one or more times (within the district) during the school year are 1.49 to 2.57 times more likely to become truant than the students that do not change schools during the school year. We believe this might be attributed to lack of stability in the home. Among special education students, those with an active status, (1.06, 1.35) are also at a slightly higher risk of becoming truant than students with no status. There was no evidence in a difference in risks between a student with an inactive status and no status.

Finally, the model also suggests that ethnicity may play a role in the risk of becoming truant. Hispanic (1.38, 2.00), African American (1.07, 1.73) and American Indian or Alaska Native (1.59, 6.39) students are at a higher risk of becoming truant as compared to Caucasian students. Asian students (0.20, 0.59) are at a lower risk of becoming truant compared to Caucasians. Although American Indians or Alaska Natives are at a higher risk,

it should be noted that the sample sizes for those groups were quite small ($n < 50$). Small sample size leads to a wider and much less informative confidence interval.

Limitations and Future Analysis

It is clear that economic status plays an integral role in the risk of truancy. Faculty members from several departments at the University (Social Work, Education, and Mathematics/Statistics) will collaborate to provide a better understanding of meaning of these results.

There are some known limitations of the economic status variable. Students are classified according to enrollment; however classification according to qualification would be preferred. In order to attain reduced or free lunch status, a parent or guardian must submit application for their child. It is quite likely that some students belong in a different classification. It is also quite likely that the homeless population is under reported.

An important question concerns the influence of Title 1 status in increasing/decreasing a student's risk of becoming truant. This potential predictor has proven to be more complex as Title 1 eligibility is based on individual students but the status is applied at the level of schools. Risks might be assessed for students that would qualify for Title 1; however the impact of programming would be examined at the level of the school. Differences between schools did exist and were briefly examined; however there was no clear evidence that such differences had any relationship to the Title 1 status of these schools.

Another future project would assess data from different years. Such comparative analyses may lead to more conclusive and informative results. Following students for several years in a longitudinal study could also provide interesting results in regards the effect of age on truancy. Expanding the analysis to include multiple districts would be beneficial as well. In particular, using multiple districts in the same area might allow students who move into neighboring districts could still be accounted for rather than censored at the date in which the student moves.

Conclusion

Truancy is a serious issue that goes beyond academic failure. Truancy has been identified as an early warning signs of negative and criminal behavior. If truancy rates are not lessened, how can society expect crime to lessen? It is imperative that children attend school daily and get the education that every child deserves. This project showed that Socio Economic Status and Age are two key predictors for risk of truancy. Further study may eventually show Title 1 to be an important predictor as well. School administrations nation-wide may use this knowledge to help identify students at-risk of truancy and then provide programs for these students designed to mitigate the problem.

References

1. Kentucky Legislature. Kentucky Revised Statutes. Title XIII –Education. 2011. Feb 15, 2012.
<http://www.lrc.ky.gov/krs/titles.htm>
2. Baker M, Sigmon J, Nugent E: **Truancy Reduction: Keeping Students in School. U.S.** Department of Justice Juvenile Justice Bulletin. September, 2001.
3. Manchin, D., Cheung, Y., and Parmar, M. Survival Analysis: A Practical Approach. Second Edition. 2006.
2006
4. SAS Publishing. SAS/Stat 9.3 User's Guide: Survival Analysis. SAS Institute Inc. NC, USA. 2011.