



**University of Pittsburgh**

*University Center for Social and Urban Research*

## **Policy Research Brief - December 2015**

# **Connecting People and Place – Improving Communities through Integrated Data Systems**

## **Chronic School Absenteeism in Public Schools in Pittsburgh, PA**

**Sabina Deitrick**

**Feifei Ye**

**Joshua Childs**

**Caiyan Zhang**

University Center for Social and Urban Research

University of Pittsburgh

Pittsburgh, PA 15260



**NATIONAL NEIGHBORHOOD  
INDICATORS PARTNERSHIP**



THE ANNIE E. CASEY FOUNDATION

## INTRODUCTION

Kids who miss too many days of school – whether excused or unexcused -- begin a path to what is known as chronic absenteeism. A student's absence from the classroom becomes *chronic* when missing school is both persistent and long term, with chronic absence formally defined as missing 10 percent or more days of a school year.<sup>1</sup> Chronically absent students are on a pathway to lower grades, lower achievement, and lower prospects for high school graduation.

The impacts of chronic absenteeism affect kids across the nation. An estimated 5 million to 7.5 million students regularly missing school and the negative impacts of chronic absence begin in Kindergarten.<sup>2</sup> In the Pittsburgh Public Schools in 2013, 25 percent of students were chronically absent.<sup>3</sup>

Children miss school for many reasons, including illness; physical and mental health; family background and social structures; conditions in schools themselves, including the physical environment and student-peer interactions; and neighborhood context and residential conditions.<sup>4</sup>

Deliberative strategies and targeted outreach have been successful in getting students to school and improving school attendance.<sup>5</sup> These strategies shift districts

and stakeholders from punitive measures against students to encouragement and directed outreach approaches for students. In Pittsburgh and Allegheny County, the *Be There* campaign, begun in the region in 2013, has developed a wide network of school districts and nonprofit stakeholders working together on messaging and programming to reduce absenteeism.<sup>6</sup> Individual schools have also developed cost-effective strategies to get absent kids back in the classroom. These efforts demonstrate that improved student attendance is possible with targeted outreach and deliberative interventions, but much more is needed.

While the academic and achievement outcomes of students missing school have been well documented, additional understanding of why kids miss school has been enhanced by Integrated Data Systems (IDS). IDS bring together administrative data from different government agencies and other external indicators to provide data support and analysis of information unavailable before IDS were possible.

### ***Purpose of the Study***

The goals of this project are to expand integrated data coordinated by the Allegheny County Department of Human Services (DHS) that brings together DHS program data and other external data

<sup>1</sup> Chang & Romero (2008). While school districts use different definitions of chronic absence, the work of Hedy Chang and Attendance Works has achieved near universal acceptance with a definition of 10% or more days missed, excused + unexcused.

<sup>2</sup> Balfanz & Byrnes (2012); Balfanz & Byrnes (2013).

<sup>3</sup> Figure from analysis in this study.

<sup>4</sup> Balfanz & Byrnes (2012); Chang & Romero (2008); Childs (2015); Gottfried (2009); Gottfried (2011).

<sup>5</sup> Sheldon & Epstein (2004)

<sup>6</sup> Childs (2015). The school districts in this study are part of Allegheny County's *Be There* campaign.

sources with public school student records.<sup>7</sup> This project builds on DHS's IDS by adding indicators on neighborhood effects and residential conditions available from the University of Pittsburgh Center for Urban and Regional Analysis. What we call people, place, and parcel incorporates data on students from three public school districts, a representative group of schools in the region, with programmatic and neighborhood conditions. The project is working with school and community stakeholders on outreach and strategies to reduce school absences. Important findings from this project have already become part of a larger community conversation on reducing chronic absence in the schools and finding solutions to reducing barriers identified through this research

---

## **CHRONIC ABSENTEEISM AND NEIGHBORHOOD CONDITIONS**

For students missing school, this research examines how neighborhood effects interact with other conditions in the student's life that contribute to chronic absence. The key question is: Does the built environment affect school attendance and, if so, what features stand out and how are they related to other factors in a student's life?

Neighborhood conditions and the quality of the residence in which a student lives have many effects on social and personal

---

<sup>7</sup> Dalton (2012); Allegheny County (2013); Fraser (2015).

<sup>8</sup> Brooks-Gunn et al. (1993); Gottfried (2014).

decisions and have long been analyzed to understand their impact on educational outcomes.<sup>8</sup> Previous work has already shown that students living in impoverished urban neighborhoods have lower levels of academic achievement. Student absences have also been shown to be related to neighborhood conditions, such as poverty, family structure, homeownership status and race.

Neighborhood effects can be structured in different ways.<sup>9</sup> More nuanced and detailed measures of neighborhood effects are available through neighborhood information systems (NIS). NIS contain dozens and sometimes hundreds of indicators on neighborhood social, economic and physical conditions. In this study, we use neighborhood data collected at the University of Pittsburgh Center for Social and Urban Research. By integrating this NIS information with DHS administrative data and public school records, the IDS developed for this analysis allows for more robust investigation into neighborhood effects and school attendance.

This project does not extend to long-term impacts of neighborhood effects; analysis of time series of neighborhood effects extends beyond the scope of this study. Here we investigate the interactions of a number of variables of the person, place and parcel

<sup>9</sup> Cutsinger, Galster and Santiago (2010, p. 18) found "clear heterogeneity in neighborhood effects," where differential effects resulted across various student groups.

and how they affect student absenteeism in cross section in 2013.

## BUILDING INTEGRATED DATA

How parcel and place effects interact with students' attendance is possible to analyze through the IDS developed here.

Our first steps were to build on research that DHS and the University of Pittsburgh had previously conducted on student absenteeism. In that work, student data from the Pittsburgh Public Schools was integrated with DHS data.<sup>10</sup> The analysis found important positive relationships between students being in DHS programs and poor school attendance.

This project expands that work by incorporating neighborhood effects into the IDS and adding two suburban school districts identified by DHS as having attendance problems among its students, Woodland Hills and Clairton City (see Figure 1). These offer two distinct school districts. Woodland Hills encompasses a number of suburban and former steel mill communities. Clairton is a small municipality with a long industrial legacy.

Student data for each school district was integrated with DHS and external data sources to create datasets for each district. The process of data integration requires linking data sets together with unique information for each entry. Student records were matched with individual program data from DHS for 2013. Neighborhood conditions were aggregated or measured at

<sup>10</sup> Kim (2013a); Kim (2013b)

the census tract level, then integrated with the student data files. Students were classified as being chronically absent -- missing 10% or more days of school -- or not, the dependent variable in the study. The integrated data set was then cleaned to produce the final data set.<sup>11</sup>

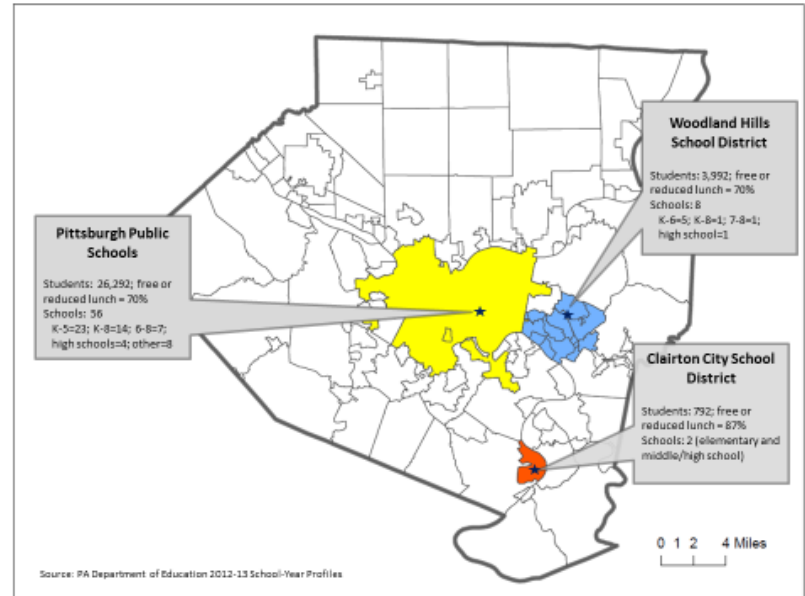


Figure 1. Pittsburgh, Woodland Hills and Clairton City school districts.

## METHODS

We conducted three types of statistical analyses for the study to bring reduce the number of potential predictors from the dozens of variables incorporated in the IDS. The tests were performed for grades in the Pittsburgh Public Schools and for grade groups for Woodland Hills and Clairton City Schools. The dependent variable, chronic absence, was defined for each student.

<sup>11</sup> Additional details on integrating data is included in the Appendix 1 and sources of data in Appendix 2.

1. We first examined bivariate relationships between the dependent variable *chronic absence* and potential predictors in the IDS, including an extensive set of neighborhood variables. For each grade (PPS) and grade group (Woodland Hills, and Clairton City), the bivariate relationship between chronic absence and other variables was examined using t-test for continuous variables and two-way chi-square for dichotomous variables. The effect sizes were examined to determine the meaningfulness of the results.<sup>12</sup> This produced the initial screening through dozens of neighborhood and parcel variables. Significant variables remained from this initial screening for further testing.
2. The second test conducted binary logistic regression using the listed predictors from the bivariate analysis. This gave the bivariate association between the attendance group and predictors, and was conducted for each grade. The simultaneous logistic regression showed the unique contribution of each variable.
3. We followed with a third test, a data mining technique -- classification and regression tree analysis (CART).<sup>13</sup> Data mining approaches, such as CART, can screen large numbers of predictor variables to

identify those that create the largest impact on the dependent variable, but unlike logistic regression, CART examines interactions among predictors through a simple tree system and is more powerful for interdisciplinary work, as being conducted here. As we proposed, neighborhood effects can be important and significant contributors to students missing school. The CART model allows us to examine those interactions of neighborhood effects with student data and student involvement in DHS programs.

Dependent variable – Chronic Absence = No. of excused + unexcused absences divided by the number of enrollment days for the student.

Yes = 1 = missed  $\geq$  10% days of school;  
 No = 0 = missed < 10% days of school.

## FINDINGS

In the Pittsburgh Public Schools, 24.5% of students chronically absent in 2013, and this rate was exceeded in both suburban districts (see Table 1). In all three school districts, school attendance problems began in Kindergarten and were evident across all grade levels (see Figures 1 and 2 for PPS and Woodland Hills).

<sup>12</sup> Point-Biserial correlation for continuous variables and Cramer's V for dichotomous variables.

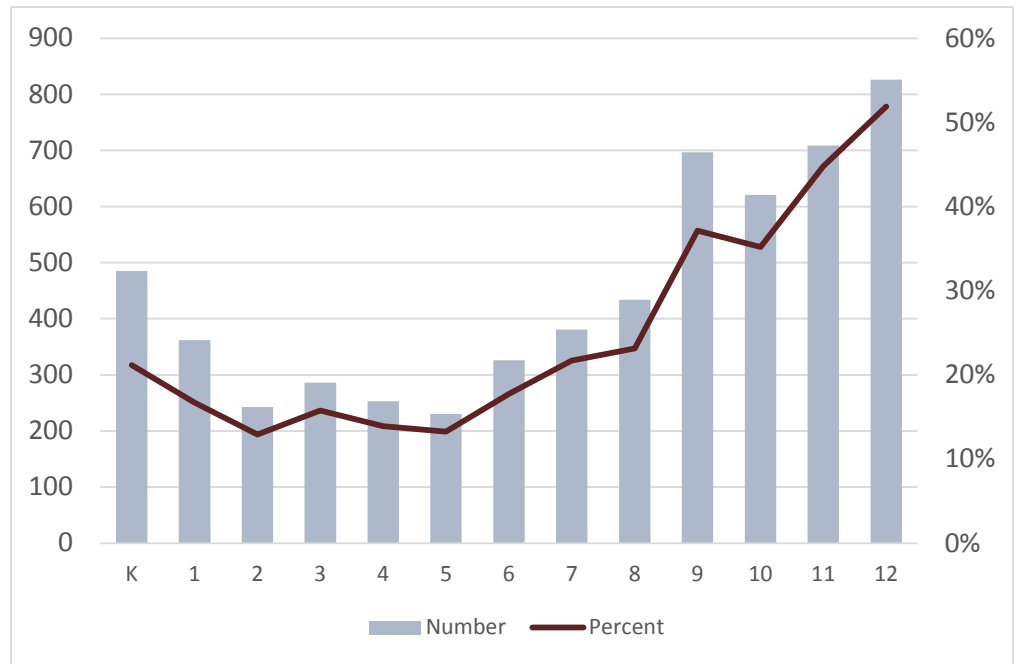
<sup>13</sup> Brieman et al. (1984); Hochachka et al. (2007). Regression models typically use a single model to

represent a data set. CART partitions data into smaller groupings with nodes classifying data interactions.

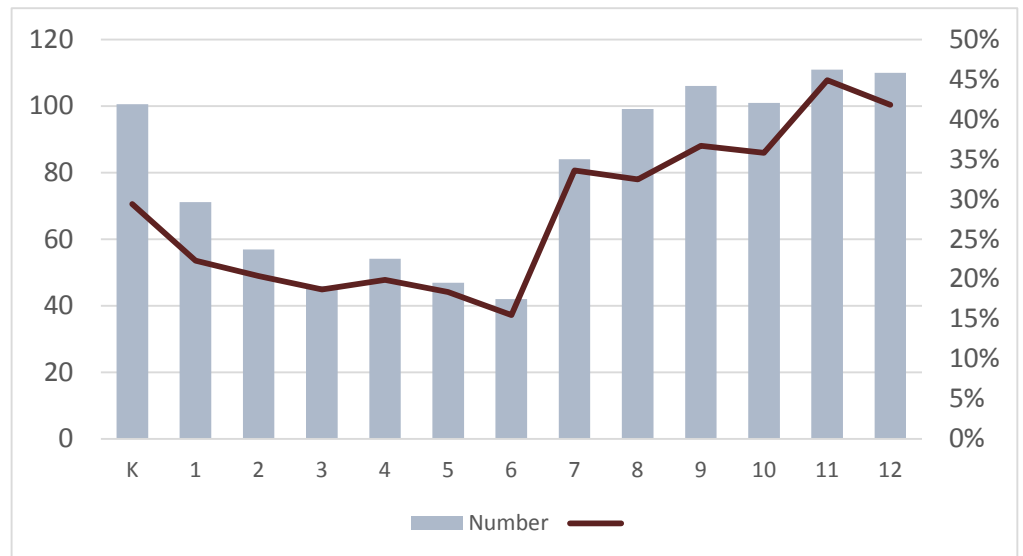
**Table 1. Number of chronically absent students, per school district, 2013**

School district	No. of students in study	No. chronically absent (≥10%)	Percent
Pittsburgh Public Schools	23,992	6,033	24.5%
Woodland Hills School District	3,620	1,029	28.4%
Clairton School District	732	223	30.5%

Generally, chronic absence typically drops from Kindergarten through elementary school, but increases in middle school and accelerates in high school. Woodland Hills offers a distinct jump from the elementary grades to junior high school and high school grades, where 40 percent or more students were chronically absent in 2013.



**Figure 2. Chronic Absence, Pittsburgh Public Schools Students, 2013**



**Figure 3. Chronic Absence, Woodland Hills School District, by Grade, 2013**

For the rest of this brief, we will focus on the results of the CART analyses for each school district.

### ***Pittsburgh Public Schools***

For the Pittsburgh Public Schools, chronic absenteeism was a problem across all grades in 2013, but one predictor of chronic absence stood out across the district across all grades. When all students were grouped together, we found that the variable indicating students who changed schools in the district during the school year – “within year school move” -- created the largest difference between students who are chronically absent and those who are not.<sup>14</sup> In 2013, 2,793 students changed schools within the district during the year, and of those students, 49.2% were chronically absent.

Changing schools in the school district also created the largest difference between chronically absent students and those not chronically absent when the modeling was done for larger grade groups (Grades K-5 and Grades 6-12). It was also the largest predictor of chronic absence for a number of individual grades, including 1<sup>st</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, 6<sup>th</sup>, 8<sup>th</sup> and 10<sup>th</sup> grade.

The importance of this finding has brought attention to the district’s current practices regarding bus transportation and has engaged stakeholders in a series of meetings to establish a means to change current practice. PPS contracts out to over twenty different bus companies who then

arrange transit for students. It often takes a week or more for students to receive their new bus transportation assignment, so for students without other means to travel to the new school, school absence is the result. The decentralization and lack of coordination of students’ bus transportation has created a practice that is a very strong contributor to students being chronically absent. This is also an area that is actionable by school officials who are working with stakeholders to revise and change their policies when students change schools and reduce the number of days the student will be absent will waiting for new bus assignments.

The strength of the CART modeling shows in the example above, but goes further. The CART modeling allows for detailed understanding of predictors of chronic absence and the interactions of predictors through what are called “nodes.” In the next section, we will focus CART results for first graders in the PPS to demonstrate the details through this model. It is beyond the scope of this brief to cover each grade individually, but this example will show how different student, DHS program and neighborhood factors interact and affect student attendance.

#### *The CART model: Chronic absence in 1st grade*

CART produces a “tree,” as shown in Figure 5 that is ordered by a variable’s predictive force on the dependent variable. The binary dependent variable is binary: the student

---

<sup>14</sup> With the IDS here, we could not account for the effects of students who change school districts, also a change in school effect.

missed 10% or more days of school in 2013 (=1) or not (=0). At the top of the tree, we see that 362 students, or 16.7% of first graders, were chronically absent in 2013.

The variable that created the biggest split in attendance rates for 1<sup>st</sup> graders in 2013 was, again, the variable for students who changed schools in the district during that year. If a first grader moved during the school year, 35.9% of them were also chronically absent. Moving down that branch of the tree, we find that of first graders who changed schools in the district during the year AND lived in a neighborhood where the median house sale price in 2013 was < \$33,500, 48.4% were also chronically absent from school in 2013. This “node” had the highest level of chronic absence in the model.

Moving to the left hand side of the tree, we start with first graders who did not change schools during the year. Among those students, the variable that created the largest difference in chronic absence was violent crime in the neighborhood, defined by the violent crime rates in the student’s home census tract.<sup>15</sup> Thus, we find that 175 students, or 20.6% of 1<sup>st</sup> graders, who did not move during the year but lived in a

neighborhood with high violent crime rates were chronically absent in 2013.

The CART adds predictors as it moves through the tree. For example, one more node lower on the left side, we find the variable “old for grade,” creates the next largest differences between students. Here, 36.3% of first graders whose age was more than a year older than the average first graders and lived in a neighborhood with high violent crime rates were chronically absent.

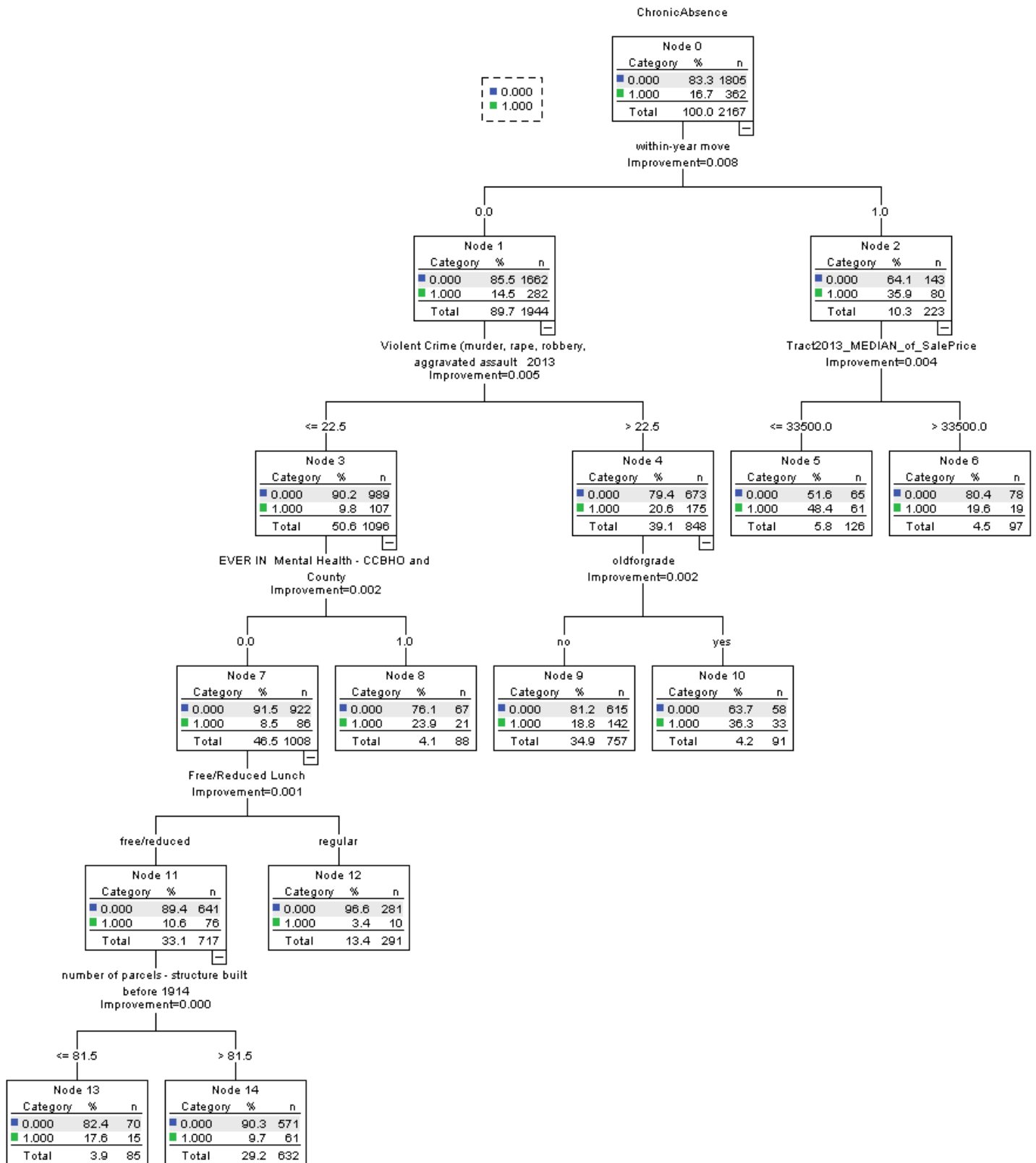
The CART model continues finding the variable of the largest predictive force on each branch of the tree. For instance, for first graders who did not change schools during the school year and did not live in a neighborhood with high rates of violent crimes, the largest predictor of chronic absence among that group were students who were enrolled in a DHS Mental Health program.<sup>16</sup> Additional CART predictors for chronic absence among first graders also included other neighborhood (older housing variable) and student effects (free/reduced lunch status).

---

<sup>15</sup> While all crime variables were significant in the bivariate test, we included only the violent crime variable in the final models because it produced the largest relationship to chronic absence.

<sup>16</sup> The splits in the tree continue to a cut off, here set at 50 students.





**Figure 4. CART Results for 1<sup>st</sup> Grade, Pittsburgh Public Schools, 2013**

### ***Chronic Absence in Other Grades***

CART modeling allows for classification of students through the IDS, and the model was successful across grades in classifying most students through data interactions. We found different neighborhood and DHS effects becoming predictors for older students that were not evident in elementary school grades.

Before turning to grade groups in the Woodland Hills School district, we'll examine neighborhood, student and DHS program interactions for 7<sup>th</sup> grade, as an example from middle school, and 10<sup>th</sup> grade, for high schoolers. What we find from CART modeling is that often different variables are the most important predictors when grades are examined individually.

The predictors of chronic absence in 7<sup>th</sup> grade were, again, a combination of individual, neighborhood, and DHS program variables. The variable creating the largest difference among 7<sup>th</sup> graders attendance were students in any Department of Public Welfare benefit. On that side of the CART, if 7<sup>th</sup> graders were in a DPW program in 2013 AND changed schools in the district, 52.4% of them were chronically absent, 6% of all 7<sup>th</sup> grade students in 2013 and represented the highest level of chronic absence of any outcome from the CART.

Seventh grade also included several neighborhood predictors of chronic absence, including indicators measuring delinquent properties, violent crimes, and median sales prices. Households of renters were also a predictor, interacting with high levels of violent crimes in a neighborhood to increase students missing school.

Finally, turning to 10<sup>th</sup> graders, 35.2 percent were chronically absent in 2013. The indicator measuring a school change during the school year created the largest difference between students who were chronically absent and those who weren't. Additional individual indicators were also among the largest predictors of chronic absence among 10<sup>th</sup> graders, including free/reduced lunch status, living in a renter household, and being in a DHS Mental Health program. The largest predictors also included the neighborhood conditions of older housing stock and large number of delinquencies.

In summary, the CART model shows that person, place and parcel variables were important predictors of chronic absence in the PPS in 2013 and interactions of those variables created more likely scenarios for conditions causing chronic absence among students. Stakeholders can use these predictors to examine students in the groupings and begin direct outreach to their families and them. For instance, in one neighborhood, there is an Early Warning System for high school students and school attendance and achievement. They are now examining the CART results to incorporate additional indicators for their Early Warning System on improving student attendance.

### ***Woodland Hills School District***

For the Woodland Hills School District, we conducted the logistic and CART modeling by grade groupings, which generated more predictors for each group (grades K-5; grades 6-12). As a smaller district, this was not unexpected.

For students in grades K-5, the predictor contributing to the largest difference in attendance was median home sales price in 2013. For those in tract with median price < \$16,975, 28.8% of K – 5 students were chronically absent, reflecting the impoverished communities that are part of the district. The next most important predictor of chronic absence was for those in the DHS Children, Youth and Families program. The neighborhood indicator measuring older housing stock by number of parcels built before 1914 also contributed to the classification of chronic absence in Woodland Hills.

The remaining five predictors with the highest average importance across K – 5 include a mix of individual student characteristics, neighborhood effects, and involvement in a DHS program: the indicator for a student being more than a year older than the average age for a grade; the number of tax delinquent properties in an area where a student lives; renter status of the student’s household; changing schools in the district during the school year, being enrolled in a Department of Public Welfare program, and involvement in a DHS Mental Health program.

The contributors to chronic absence for K – 5<sup>th</sup> grade students in Woodland Hills in 2013 marked a combination of neighborhood affects, student characteristics and participation in DHS programs in Mental Health and Public Welfare. The highest level of chronic absence was found in Node 16 (57.7%) with 41 students who lived in a

tract with a median sales price in 2013 < \$16,975, were in Children, Youth and Families, lived in a tract with a sum of total delinquencies < \$25,054, and were in a Department of Public Welfare program.

We grouped 6<sup>th</sup> – 12<sup>th</sup> graders together for the remaining statistical tests in Woodland Hills. The 7 predictors which had the highest average importance value across grades 6-12 are: 1) old for grade, 2) Department of Public Welfare program, 3) renter status of student’s household, 4) older age of housing stock measure, number of parcels built before 1914, 5) lower medium house sales price in 2013, 6) percent below poverty level, 7) percent black.

### ***Clairton School District***

Clairton School District presents a third level of district in the study, one of the very smallest school districts in the county, with a population about half what it was in 1980. In the 2010-14 period, 16.7% of families were in poverty.<sup>17</sup> Nonetheless, stability continues in the community with homeownership rates of 57.0 percent, though vacant housing units are on the rise and reached 21.5 percent of the total housing units over those same years.

Most prominent in Clairton is the Clairton Coke Plant of US Steel’s Mon Valley Works, the largest coke facility in the U.S., producing 4.7 million tons of coke per year. This certainly becomes a factor for kids attending school.<sup>18</sup> Students’ health is an important determinant of children missing school days, and the Clairton City Schools, students are more likely to suffer from asthma because of the environmental

<sup>17</sup> U.S. Census Bureau, American Community Survey, 5-year estimates, 2010-2014.

<sup>18</sup> Templeton (2015).

conditions in the community, with some of the worst fine particle pollution rates in the nation.<sup>19</sup>

Because of the small size of the school, students were categorized into three grade groups for the analysis: K-5; 6-8; and 9-12. Chronic absence by grade grouping:

---

**Table 2. Chronic Absence in Clairton City School District**

Grade group	Students	Chronic absence	Percent
K-5	391	108	27.6%
6-8	160	43	26.9%
9-12	181	72	39.8%

\*Missing 10% or more days of the school year.

---

For 391 students in grades K-5 in 2013, four predictors contributed to the classification of chronic absence: school move, living in one particular part of the community along the river, enrolled at any point in the year in the DHS Children, Youth and Families program, and being in a renter household.

Results for the other two grade groupings were more limited, owing to the small number of students (160 in Grades 6-8 and 181 in Grades 9-12). For students in grades 6-8 in 2013, being enrolled in a DHS Children, Youth and Families program was the largest predictor of chronic absence. In 2013, 26.9% of students in grades 6-8 were chronically absent, and of these students, 60.5% were in a DHS Children, Youth and Families program at some point during the

year. For students in grades 9 – 12 in high school, almost 40 % of students were chronically absent in 2013, and being “old for grade,” more than one year older than the average age for a grade, was the largest determinant of chronic absence and the only predictor from the CART model.

As a small district, the CART offers the opportunity for more direct outreach to families and students. Particularly among middle school students, DHS program involvement is the largest contributor the classification of chronic absence and marks a distinction from the other two districts, which showed stronger neighborhood effects as a contributor.

---

## **POLICY IMPLICATIONS AND RECOMMENDATIONS**

The results of this study point to the interwoven factors that affect students’ school attendance and, also, how they vary by different grades. The researchers are currently engaged with a number of stakeholders and stakeholder groups across the area on the results of the study, the implications of those results, and follow up work that can be done from the IDS.

Across all grades, indicators across both people, place and parcel status affected student absenteeism, and the CART analysis allows for specific factors to be understood as predictors of chronic absence for a large city school district, such as PPS. By generating greater understanding of interactions among predictors, these results

---

<sup>19</sup> Katz (2012).

can give more direction to intervention and outreach efforts of schools and stakeholders.

Certainly, the impacts of students changing schools during the school year consistently showed strong, positive impacts on student attendance. For example, in 10<sup>th</sup> grade for PPS students, changing schools during the school year produced the highest level of chronic absence (68.5%) via the CART model, in a grade where 13.7% of students changed schools in just that year. It was the first level predictor in six grades. The impacts of changing schools on attendance has become a critical concern of school officials and other stakeholders who are working to overcome barriers in the Pittsburgh Public Schools and other school districts.

Stakeholder engagement on the impacts of school changes during the school year and students missing school have focused on the district's policies regarding busing and the decentralization of student access to the bus companies. The district is currently in process of engaging stakeholders across schools and communities to change their policies and not make what is already disruptive to a child even more disruptive by not having the opportunity to start in the new school until a week after the new school assignment.<sup>20</sup>

School and neighborhood stakeholders also discussed how results here might impact their own risk factor analyses and early

warning systems that have been developed. While we do not have examples yet of frameworks that have changed various composite measures because of results from these models, we do know that the impacts are being examined.

In one neighborhood school, researchers in the Homewood Children's Village are looking into using the data set available via DHS for one high school and examine student attendance in the subsequent years since they have been involved in a set of activities and programs to improve student attendance, including their own early warning efforts and *the Be There* campaign.

The number of neighborhood factors related to housing conditions in a neighborhood – older age of housing stock and number of tax delinquent properties, for instance – that affected students' attendance raised cognizance of housing conditions for families. The presentation of this project's results, with neighborhood indicators as important predictors of chronic absence, led to discussions among stakeholders of the importance of working to stabilize housing opportunities and housing conditions for families with children in the districts. Stakeholders discussed the need for greater outreach to community development corporations (CDCs) and the awareness of conditions of neighborhood distress affecting students going to school. While only in discussion stages, our partners were suggesting a more formal collaboration with CDC and

---

<sup>20</sup> The United Way has begun a series of meetings focused on transportation impacts with school districts and stakeholders in Allegheny County.

other service organizations assisting families with housing needs. There is certainly excitement in the county to tie successful efforts in promoting students school attendance to community development and housing organizations more formally. As one stakeholder in the education and attendance community commented, the results of this work has rolled back some bigger problems in understanding ways to improve school attendance. “Attendance folks don’t think about housing and housing folks don’t think about attendance,” she commented, but clearly poor housing conditions and housing instability are clearly linked to the students’ school moves and, thus, school attendance and school absence.

One stakeholder group added that education opportunities and school attendance should be issues that newcomers to the neighborhood are made aware of upon arrival. While some neighborhoods have “welcome wagon” materials, many neighborhoods provide little education information to new residents, particularly information on assistance with kids going to school and other support for increasing attendance, including, in Allegheny County, the United Way’s successful *Be There Attendance Campaign* that has been a major effort in developing partnerships to promote awareness of the problem and create strategies to get kids in school.

Work on this project extends to area stakeholders, with continued outreach and dissemination of the results. There are many important and successful efforts well underway to promote awareness about

chronic absenteeism and increase students’ school attendance in Allegheny County, and the work from this IDS will continue through these efforts. The United Way of Allegheny County and University of Pittsburgh Office of Child Development will convene additional forums from this research in early 2016.

---

## **APPENDIX 1: DATA INTEGRATION**

The process of data integration requires linking data sets together with unique information for each student entry. We followed a protocol of data checking, variable identification and data cleaning to build the new, expanded IDS. Student records were matched with individual DHS program data for 2013. Neighborhood conditions were aggregated or measured at the census tract level, then integrated with student data files. Bivariate analysis of the neighborhood variables was then run against the dependent variable, chronic absence, to include only significant predictors for the final models.

These contained information on the neighborhood housing market, poverty levels, population by race, incidence of violent crimes, the number of tax delinquent properties, and age of housing, with additional integration of parcel information linked by the student's address. (Sources of data and files appear in Box 1 and Box 2). These indicators were selected after a much larger set of neighborhood effects were tested for a significant relationship to chronic absence. Student data was also geocoded.

Only students with total number of enrollment days greater than 50 were included in the sample. The attendance record of each student was analyzed. Chronic absences was defined by adding the number of excused + unexcused absences, then dividing by the number of days in a student's school year. A binary dependent

variables was created: Yes -- students with 10% or more days absent; No < 10% days absent. We then included other information from school records, including sex and race of the pupil, lunch code for reduced lunch status, if the student moved within the year to capture the effects of potential instability, and "old for grade," a variable computed by student's age if more than 6 years older than current grade.

The number of variables was reduced through each process. For instance, we began by analyzing several different crime variables -- total crime, violent crime, crimes involving a gun, and property crime -- through bivariate analysis. Though other crime measures were significant, the variable for "number of violent crimes" in a neighborhood consistently presented the strongest positive relationship to chronic absence compared to the other crime variables and thus was chosen for the logistic regression and CART analysis.

---

## APPENDIX 2: SOURCES OF DATA AND VARIABLES

*The major data sources integrated into this study include:*

People (individual records):

- Allegheny County Department of Human Services program data
- Housing Authority of City of Pittsburgh public housing
- Allegheny County court records and juvenile justice
- Pittsburgh Public Schools student data
- Clairton City School District data
- Woodland Hills School District data

Place (aggregated to census tract of student home residence):

- American Community Survey
- City of Pittsburgh Police crime data
- Allegheny County Department of Court Records (foreclosure filings not significant)
- Home Mortgage Disclosure Act (lending data not significant)
- Allegheny County Office of Property Assessment
- Allegheny County Treasurer

Parcel (individual property records):

- Allegheny County Office of Property Assessment

*The student level data included whether the individual was in one of the DHS or Allegheny County programs (1=yes if the student was in the program at **any** time*

*during the school year; 0=no, the student was not in that program at any time during the school year):*

- Children, Youth, and Families
- Drug & Alcohol
- Early Intervention
- Family Support Center
- Hunger and Homeless
- Mental Health
- Intellectual Disability
- Department of Public Welfare (includes Food Stamps, Social Security Insurance, and Temporary Assistance for Needy Families)
- Juvenile Justice (this variable is excluded for grades K-5)
- Housing Authority of City of Pittsburgh family housing communities (does not include voucher program)

The neighborhood related data were captured by the following indicators at the Census tract level:

- Median sales price of houses in 2013
- Percent of population below poverty level
- Percent of population non-Hispanic Black
- Number of violent crimes in 2013
- Number of tax delinquency properties 2013
- Number of parcels/structures built before 1914)

The final component of the neighborhood was parcel-level housing unit data, including:



- Homestead Exemption – this is a proxy for homeowner status or renter status.

---

## ACKNOWLEDGEMENTS

The authors would like to thank Allegheny County Department of Human Services for their involvement and extension of the IDS. We benefitted from engagement with the University of Pittsburgh -- Office of Child Development, United Way of Allegheny County, Homewood Children's Village, and the Pittsburgh Public Schools. The authors would also like to thank Bob Gradeck and Mike Blackhurst of UCSUR for their contributions to this study.

This research was funded by the Annie E. Casey Foundation as part of the Integrated Data Systems work of the National Neighborhood Indicators Partnership. We would also like to acknowledge the continued engagement and support of the Urban Institute for this work.

The views expressed in this work as those of the authors.

---

## AUTHORS

**Sabina Deitrick, PhD**, Director, Urban and Regional Analysis Program, University

Center for Social and Urban Research, University of Pittsburgh.

**Feifei Ye, PhD**, Assistant Professor of Research Methodology, School of Education, University of Pittsburgh.

**Joshua Childs, PhD**, Assistant Professor, Educational Policy and Planning, College of Education, University of Texas at Austin.

**Caiyan Zhang, PhD**, Assistant Psychometrician, College Board.

*The University Center for Social and Urban Research (UCSUR) was established in 1972 to serve as a resource for researchers and educators interested in the basic and applied social and behavioral sciences. As a hub for interdisciplinary research and collaboration, UCSUR promotes a research agenda focused on the social, economic, and health issues most relevant to our society.*

---

## REFERENCES

Allegheny County Department of Human Services. (2013). Data Brief: School Attendance and Academic Achievement in the Pittsburgh Public Schools. Pittsburgh: Allegheny County Department of Human Services, June.

Balfanz, R., & Byrnes, V. (2012). *The importance of being in school: A report on absenteeism in the nation's public schools*. Baltimore: John Hopkins University Center for Social Organization of Schools, May.

Balfanz, R., & Byrnes, V. (2013). Meeting the challenge of combating chronic absenteeism: Impact of the NYC mayor's interagency task force on chronic absenteeism and school attendance and

- its implications for other cities. Baltimore, MD: John Hopkins School of Education.
- Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). *Classification and Regression Trees*. Belmont, California: Wadsworth.
- Brooks-Gunn, J., Duncan, G.J., Klebanov, P.K., & Sealander, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology*, 99, 353-395.
- Chang, H.N., & Romero, M. (2008). Present, engaged, and accounted for: The critical importance of addressing chronic absence in the early grades. New York, NY: National Center for Children in Poverty, September.
- Childs, J. (2015). *Getting Students to "Be There" in School: the Be There Campaign and a Network Approach to Addressing Chronic Absenteeism in Pittsburgh*. Unpublished doctoral dissertation, University of Pittsburgh, July.
- Cutsinger, J.M., Galster, G.C. and Santiago, A.M. (2010). Neighborhood Effects on High School Outcomes for Latino and Black Youth: Evidence of a Natural Experiment in Denver. Presented at APPAM meeting, Boston, November.
- Dalton, E. (2012). Integrated Data Systems for an Integrated Practice Model. Presentation at the National Neighborhood Indicators Partnership, Providence, RI, September.
- Fraser, Jeffery. (2015). *Improving Educational and Well-Being Outcomes: School-DHS Data Sharing in Allegheny County*. Allegheny County Department of Human Services, Pittsburgh, PA.
- Gottfried, M.A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis*, 31(4), 392-415.
- Gottfried, M.A. (2011). The detrimental effects of missing school: Evidence from urban siblings. *American Journal of Education*, 117(2), 147-182.
- Gottfried, M. A. (2014). Can neighbor attributes predict school absences? *Urban Education*, 49(2), 216-250.
- Hochachka, W.M. et al. (2007). Data-Mining Discovery of Pattern and Process in Ecological Systems. *Journal of Wildlife Management* 71(7): 2427-2437.
- Katz, Cheryl. (2012). Unequal exposures: People in poor, non-white neighborhoods breathe more hazardous particles. Environmental Health News, November 1: <http://www.environmentalhealthnews.org/ehs/news/2012/unequal-exposures>
- Kim, K.H. (2013a). Pittsburgh Public Schools: Attendance summary report. Unpublished raw data. Pittsburgh: University of Pittsburgh, School of Education, June.
- Kim, K.H. (2013b). Pittsburgh Public Schools: Attendance summary report. Unpublished raw data. Pittsburgh: University of Pittsburgh, School of Education, September.
- Sheldon, S.B., & Epstein, J.L. (2004). Getting students to school: Using family and community involvement to reduce chronic absenteeism. *School and Community Journal*, 4(2), 39-56.
- Templeton, David. (2015). Allegheny County offers help to Northgate students with asthma. *Pittsburgh Post-Gazette*, July 9. <http://www.post-gazette.com/local/north/2015/07/09/Allegheny-County-offers-help-to-Northgate-students-with-asthma/stories/201507080226>