University Center for Social and Urban Research

Integrated Data to Predict Chronic Absence in the Clairton City Public School District

August 2015

Prepared By

Feifei Ye, PhD
Sabina Deitrick, PhD
Josh Childs, PhD
Caiyan Zhang

This research was conducted through the *Connecting People and Place: Improving Communities through Multi-agency Integrated Data Systems (IDS)* project, funded by the Annie E. Casey Foundation and National Neighborhood Indicators Partnership, Urban Institute, in partnership with Allegheny County Department of Human Services.

University of Pittsburgh

Integrated Data to Predict Chronic Absence in Clairton City School District

08/18/2015

Table of Contents

This report addresses chronic absence in Clairton City Public School District and its association with student demographic variables, within-year move, Human Service variables, census tract variables, and house assessment data. Student files from the year of 2013 were combined with student records from DHS. The census tract data and housing information were integrated with student level data. Only students with total number of enrollment dates greater than 50 were included, resulting with 738 students. Census tract variables were missing for six students, who were excluded in the analysis. The final analysis sample has 732 students.

Attendance rate was computed by summing excused and unexcused absences divided by enrollment days (i.e., (iExc + iUnexc)/iMemb).

- Dependent variables:
    - chronic absence – was created (No, < 10%; yes, ≥ 10%).
- Predictors of CART:
    - School data:
        - sex (0=male, 1=female),
        - Within school year move (defined as whether the student had a start date after 09/30)
        - Old for grade
    - Department of Human Service Data (all the following variables are binary; 1=yes, 0=no)
        - CYF (EVER IN Children, Youth, and Families Children)
        - DA (EVER IN Drug & Alcohol - CCBHO and County)
        - EI (EVER IN Early Intervention)
        - FSC (EVER IN Family Support Center)
        - HH (EVER IN Hunger and Homeless)
        - MH (EVER IN Mental Health - CCBHO and County)
        - MR (EVER IN Office of Mental Retardation and Developmental Disabilities)
        - HACP (EVER IN Housing Authority of City of Pittsburgh)
        - DPW (EVER IN Department of Public Welfare-FS, SSI, and TNF)
        - JPO (Ever in JPO children): this is excluded for grades k-5.
    - Census Tract data
        - Median House Sale price of 2013
        - Percentage below poverty level
        - Percentage Black
        - sumoftotal (sum of total delinquency)
        - HsgAgeBefore1914_tract_CountOfma (number of parcels-structure built before 1914)
        - Tract ranking suburbs
            - RankSuburbs_Poverty_12
            - RankSuburbs_Poverty200_12
            - RankSuburbs_FHF_12
            - RankSuburbs_MaleLF_12
            - RankSuburbs_Vacant_12
            - RankSuburbs_NoVehicle_12
            - RankSuburbs_Dropout_12

        - RankSuburbs_Sum
    - House assessment data
        - Homestead status
- Three different types of analyses were performed for each grade:
    - ***Section 1: Bivariate relationship*** between chronic absence and other variables were examined using t-test (for continuous variables) and two-way chi-square (for dichotomous variables).
    - ***Section 2: Binary logistic regression*** was performed predicting chronic absence by the aforementioned predictors.
    - ***Section 3: Classification and tree analysis*** (CART) is a data mining method to determine important predictors of categorical dependent variable using a non-parametric iterative process. Unlike logistic regression, CART examines interactions among predictors. CART was performed for the outcome variable Chronic Absence for subjects by each grade.
- There is no missing data on these variables.

1. Bivariate relationship

1.1 Description of house parcel and census tract variables

There are 422 house parcels and 28.91% is homestead. There are 3 census tracts from which Clairton students were from.

Table 1 presents the specific values of census tract variables for the three tracts respectively.

Table 1. Census tract variables measured at tract level

TRACT	492700	492800	492900
SumOfTotal	37511.35	69942.66	22356.79
HsgAgeBefore1914_tract_CountOfma	183	149	10
Tract2013_MEDIAN_of_SalePrice	16500	16500	13500
percent_poverty	25.4	36.4	23.1
percent_black	0.09	0.11	0.06
RankSuburbs_Poverty_12	241	255	247
RankSuburbs_Poverty200_12	237	250	228
RankSuburbs_FHF_12	258	229	231
RankSuburbs_MaleLF_12	211	259	223
RankSuburbs_Vacant_12	241	251	229
RankSuburbs_NoVehicle_12	196	257	223
RankSuburbs_Dropout_12	1	263	1
RankSuburbs_Sum	1385	1764	1382

1.2 Description of student-level variables by chronic absence

Due to small sample size, students were categorized into three grade groups, K-5, 6-8, and 9-12. For each grade group, bivariate relationship between chronic absence and other variables were examined using t-test (for continuous variables) and two-way chi-square (for dichotomous variables). Effect sizes (Point-Biserial correlation for continuous variables and Cramer's V for dichotomous variables) was examined in order to determine meaningfulness of results. Cramer's V and point-biserial correlation are on the same metric as correlation coefficient (small, 10-.30; medium, .30-.50; large, >.50).

Table 2. Bivariate association between attendance group and predictors—Kindergarten to Grade 5 (N=391)

	Non-Chronic absent 72.40%		Chronic absent 27.60%		Correlation
	Proportion		Proportion		
Female	0.51		0.49		.013
old for grade	0.07		0.11		.059
School Move	0.03		0.16		.225***
CYF	0.25		0.40		.145**
DA	0.00		0.00		.03
MH	0.18		0.19		.021
EI	0.10		0.05		.085
FSC	0.19		0.16		.043
HH	0.13		0.11		.027
MR	0.00		0.00		.031
JPO	0.00		0.00		na
DPW	0.36		0.40		.035
HACP	0.02		0.00		.077
homestead	0.25		0.19		.059
	M	SD	M	SD	
Sum of Total Delinquency in 2013	48964.83	21708.83	45052.36	20459.72	0.082
Number of parcels-structure built before 1914	110.92	70.15	115.63	73.29	0.03
Medium house sales price in 2013	15545.94	1399.60	15555.56	1399.82	0.003
% below poverty level	30.19	6.29	28.85	6.00	0.096
% black	0.09	0.02	0.09	0.02	0.033
Tract ranking Poverty	249.93	5.49	248.20	5.85	0.137**
Tract ranking Poverty200	240.66	9.87	239.10	9.28	0.072
Tract ranking FHF	234.86	10.90	238.49	13.03	0.14**
Tract ranking MaleLF	238.90	20.61	233.00	20.98	0.127*
Tract ranking Vacant	242.20	9.74	241.02	9.17	0.055
Tract ranking NoVehicle	235.19	23.76	227.66	25.41	0.138*
Tract ranking Dropout_12	132.46	131.23	100.46	127.74	0.109*
RankSuburbs_Sum	1574.22	190.80	1527.94	185.53	0.109*

Table 3. Bivariate association between attendance group and predictors—Grade 6-8 (N=160)

	Non-Chronic absent 73.10%		Chronic absent 26.90%		Correlation
	Proportion		Proportion		
Female	0.48		0.33		.137
old for grade	0.12		0.30		.216**
School Move	0.05		0.07		.036
CYF	0.28		0.60		.296***
DA	0.00		0.00		na
MH	0.28		0.35		.065
EI	0.09		0.07		.038
FSC	0.22		0.14		.092
HH	0.03		0.07		.077
MR	0.02		0.02		.020
JPO	0.01		0.02		.059
DPW	0.39		0.42		.023
HACP	0.02		0.02		.020
homestead	0.28		0.19		.097
	M	SD	M	SD	
Sum of Total Delinquency in 2013	49054.99	21465.89	39513.32	20143.74	0.198*
Number of parcels-structure built before 1914	114.39	69.72	93.05	79.36	0.13
Medium house sales price in 2013	15602.56	1379.56	15104.65	1514.05	0.155*
% below poverty level	30.16	6.26	27.40	5.74	0.197*
% black	0.09	0.02	0.08	0.02	0.18*
Tract ranking Poverty	249.74	5.65	247.70	5.22	0.162*
Tract ranking Poverty200	240.75	9.73	236.44	9.30	0.196*
Tract ranking FHF	235.55	11.49	237.35	12.28	0.069
Tract ranking MaleLF	238.39	20.95	229.98	18.93	0.181*
Tract ranking Vacant	242.37	9.57	238.21	9.44	0.191*
Tract ranking NoVehicle	234.32	24.49	225.58	22.68	0.16*
Tract ranking Dropout_12	130.88	131.56	74.12	118.91	0.194*
RankSuburbs_Sum	1571.98	191.20	1489.37	172.89	0.194*

Table 4. Bivariate association between attendance group and predictors—Grade 9-12 (N=181)

	Non-Chronic absent 60.20%		Chronic absent 39.80%		Correlation
	Proportion		Proportion		
Female	0.46		0.56		.095
old for grade	0.06		0.36		.393***
School Move	0.01		0.11		.230**
CYF	0.30		0.49		.185*
DA	0.00		0.06		.185*
MH	0.21		0.33		.137
EI	0.01		0.06		.139
FSC	0.13		0.13		.005
HH	0.08		0.21		.182*
MR	0.04		0.01		.068
JPO	0.03		0.14		.211**
DPW	0.32		0.53		.206**
HACP	0.02		0.01		.017
homestead	0.33		0.22		.117
	M	SD	M	SD	
Sum of Total Delinquency in 2013	49546.29	21009.10	43388.17	43388.17	0.14
Number of parcels-structure built before 1914	120.13	68.01	92.89	92.89	0.185*
Medium house sales price in 2013	15701.83	1331.80	15166.67	15166.67	0.185*
% below poverty level	30.24	6.19	28.66	28.66	0.124
% black	0.09	0.02	0.08	0.08	0.174*
Tract ranking Poverty	249.53	5.85	249.11	249.11	0.037
Tract ranking Poverty200	241.05	9.47	238.06	238.06	0.15*
Tract ranking FHF	236.45	12.15	234.72	234.72	0.074
Tract ranking MaleLF	237.97	21.36	235.00	235.00	0.07
Tract ranking Vacant	242.76	9.25	239.56	239.56	0.162*
Tract ranking NoVehicle	233.40	25.38	231.72	231.72	0.034
Tract ranking Dropout_12	130.80	131.60	102.89	102.89	0.105
RankSuburbs_Sum	1571.96	191.16	1531.06	1531.06	0.106

Figure 1. Effect sizes of demographic variables by grades

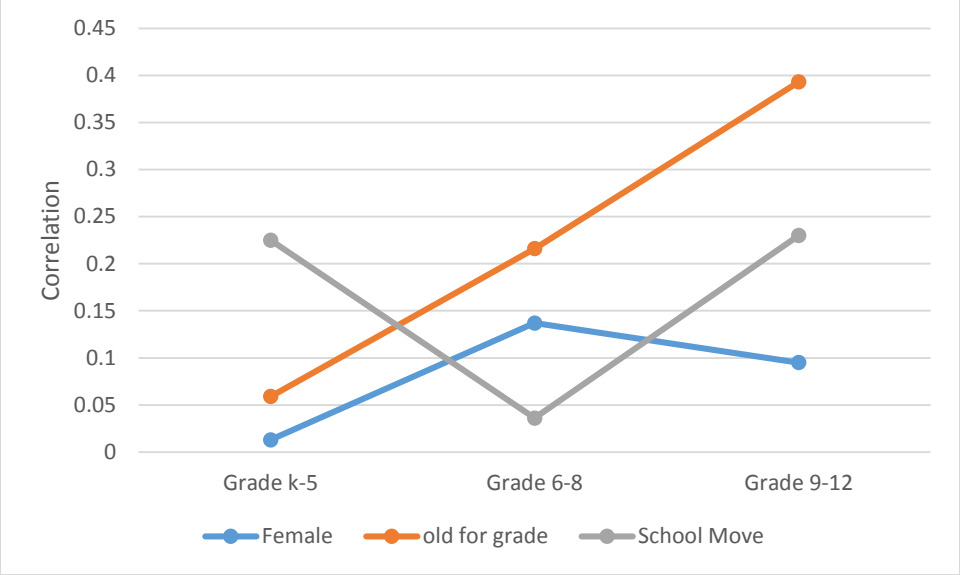
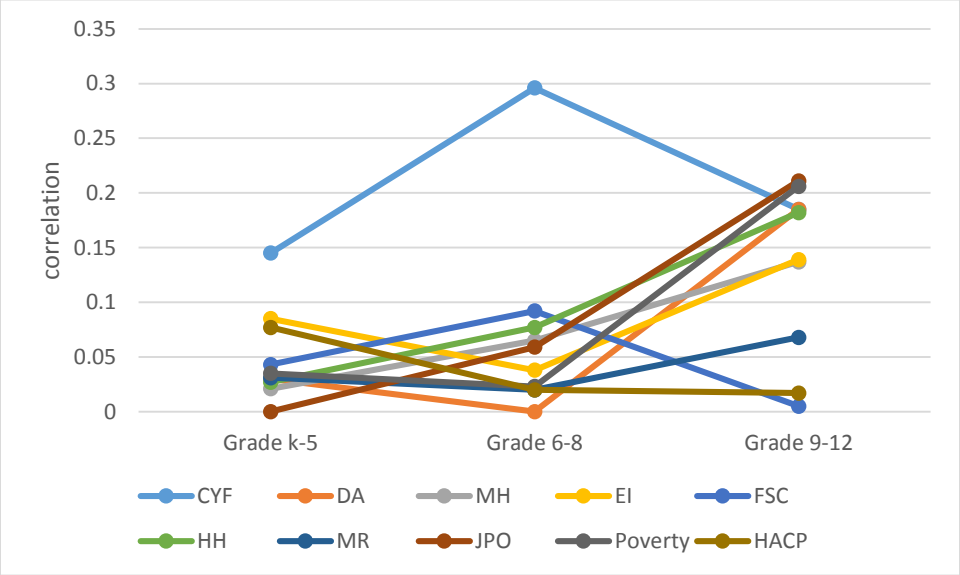
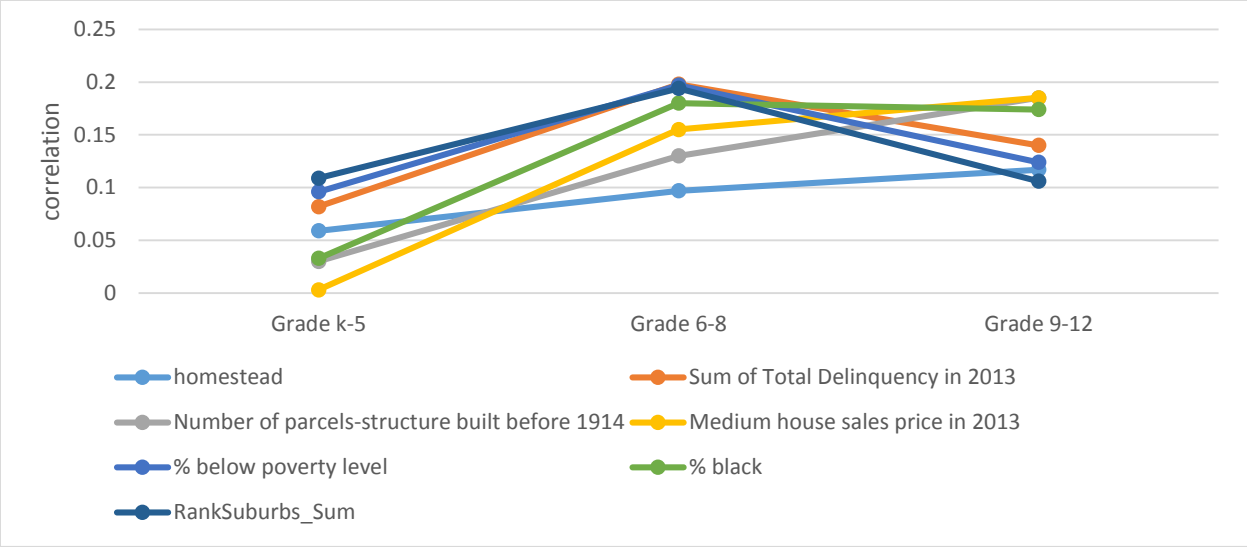


Figure 2. Effect sizes of DHS involvement variables by grades



Among tract ranking variables, the sum track ranking presents similar bivariate relationship as the other tract ranking variables, and thus was selected as a census tract variable and presented in Figure 3.

Figure 3. Effect sizes of house and census tract variables by grades


0.25
0.2
0.15
0.1
0.05
0
correlation
Grade k-5
Grade 6-8
Grade 9-12
homestead
Sum of Total Delinquency in 2013
Number of parcels-structure built before 1914
Medium house sales price in 2013
% below poverty level
% black
RankSuburbs_Sum

2. Logistic Regression

Logistic regression was used to predict chronic absence for each grade group. As there are only three census tracts, all census tract variables cannot be included as predictors. Instead, census tract is used as a categorical predictor represented by two indicator-coded variables (Indicator 1 compares tract 492700 to 492900 and indicator 2 compares tract 492800 to 492900). The three census tracts differ in their chronic absence rate for each grade group (p=.016, Cramer's v=.145 for grades k-5; p=.044, Cramer's v=.198 for grades 6-8; p=.044, Cramer's v=.186 for grades 9-12).

Table 5 presents the chronic absence rate for each tract of three grade groups respectively.

Table 5. Cross-tabulation of census tracts by chronic absence

				chronic		Total
				0	1	
Grade K-5	TRACTCE10	492700	Count	51	33	84
			% within TRACTCE10	60.70%	39.30%	100.00%
		492800	Count	142	41	183
			% within TRACTCE10	77.60%	22.40%	100.00%
		492900	Count	90	34	124
			% within TRACTCE10	72.60%	27.40%	100.00%
	Total		Count	283	108	391
			% within TRACTCE10	72.40%	27.60%	100.00%
Grade 6-8	TRACTCE10	492700	Count	24	11	35
			% within TRACTCE10	68.60%	31.40%	100.00%
		492800	Count	58	12	70
			% within TRACTCE10	82.90%	17.10%	100.00%
		492900	Count	35	20	55
			% within TRACTCE10	63.60%	36.40%	100.00%
	Total		Count	117	43	160
			% within TRACTCE10	73.10%	26.90%	100.00%
Grade 9-12	TRACTCE10	492700	Count	26	12	38
			% within TRACTCE10	68.40%	31.60%	100.00%
		492800	Count	54	28	82
			% within TRACTCE10	65.90%	34.10%	100.00%
		492900	Count	29	32	61
			% within TRACTCE10	47.50%	52.50%	100.00%
	Total		Count	109	72	181
			% within TRACTCE10	60.20%	39.80%	100.00%

Table 6 presents odds ratio and statistical significance for grades k-5, 6-8, and 9-12.

Table 6. Odds ratio of logistic regression predicting chronic absence for k to 5th grade (* p < .05, ** p < .01, *** p < .001)

	Grade K-5		Grade 6-8		Grade 9-12	
Female	0.796		0.438		1.293	
old for grade	1.070		3.089	*	10.086	***
School Move	5.973	***	1.116		16.969	*
CYF	2.120	**	4.483	***	0.96	
DA	0.000		na		5E+16	
MH	1.056		1.002		1.031	
EI	0.469		0.355		2.736	
FSC	0.702		0.373		1.634	
HH	0.904		2.4		2.38	
MR	0.000		2.628		0	
JPO	na		2.582		7.16	*
DPW	1.055		0.499		2.261	
HACP	0.000		6.57		0.057	
homestead	0.656		0.468		0.857	
Tract	Wald χ^2=6.402	*	Wald χ^2=3.504		Wald χ^2=3.571	
492700 vs 492900	1.952	*	1.05		0.445	
492800 vs 492900	0.941		0.438		0.47	
Pseudo R^2	0.155		0.28		0.43	

3. Classification and Regression Tree Analysis

3.1 Grade K-5

- N = 391
- 4 predictors contributed to the classification of chronic absence. CART produced 6 child nodes.
- The predictors are: level-1, school move; level-2, tract; level-3, CYF; level-4, Tract.
- Node 2 had the highest levels of chronic absence (65.4%) which contained 6.6% of students who moved.
- Node 9 had the lowest level of chronic absence (14.2%) which contained 27.1% of students who were did not move, lived in tract 482800, were not in CYF, and non-homestead.
- Overall 74.4% of students were correctly classified.

Table 7. Importance table (Kindergarten-Grade 5)

Independent Variable	Importance	Normalized Importance
School Move or not (1: Yes)	0.02	100.00%
TRACTCE10	0.007	35.70%
CYF (EVER IN Children, Youth, and Families Children)	0.004	18.60%
homestead	0.002	8.60%

Table 8. Chronic absence by node (kindergarten-Grade 5).

Move	Tract	CYF	Homestead	Node	N	% Total Sample	% Chronic absence
Y				2	26	6.60%	65.40%
N	492700			3	77	19.70%	36.40%
N	492800;492900	Y		6	84	21.50%	29.80%
N	492900	N		8	72	18.40%	22.20%
N	482800	N	N	9	106	27.10%	14.20%
N	482800	N	Y	10	26	6.60%	26.90%

Figure 4. Chronic absence by node (kindergarten-Grade 5).

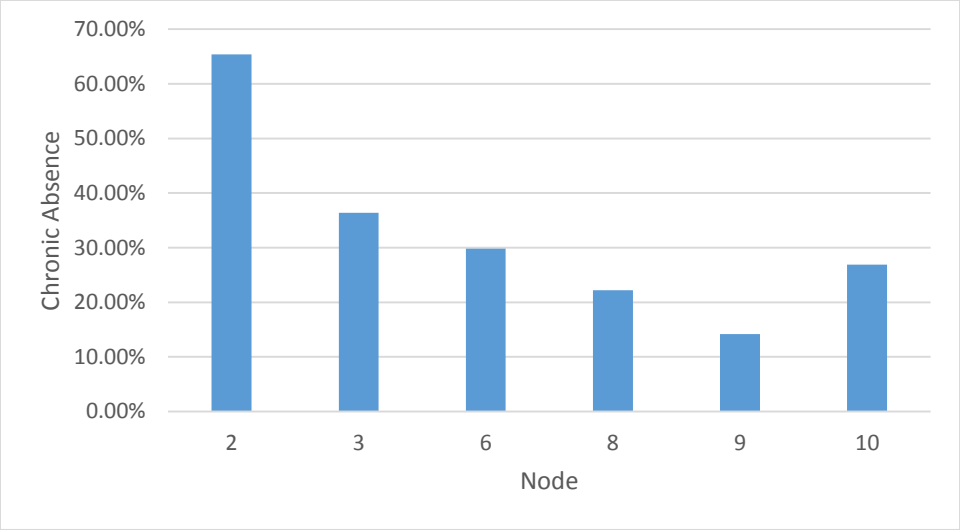
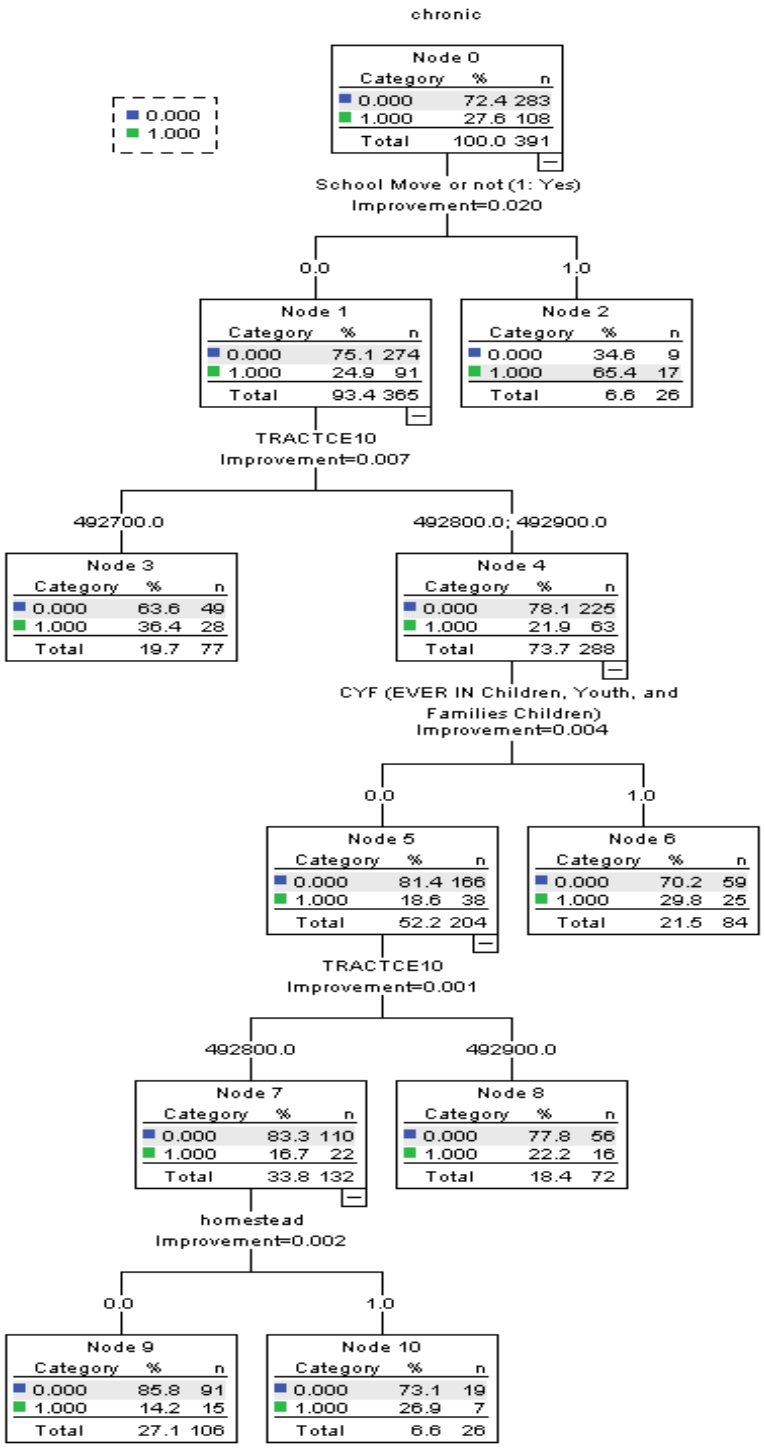


Figure 5. Tree Diagram (kindergarten-Grade 5).



3.2 Grade 6-8

- N = 160
- 2 predictors contributed to the classification of chronic absence. CART produced 3 child nodes.
- The predictors are: level-1, CYF; level-2, tract.
- Node 1 had the highest levels of chronic absence (44.1%). Node 1 contained 36.9% of students who were in CYF.
- Node 4 had the lowest level of chronic absence (10.7%) which contained 46.9% of students who were not in CYF, and lived in tract 492700 and 492800.
- Overall 73.1% of students were correctly classified.

Table 9. Importance table (Grade 6-8)

Independent Variable	Importance	Normalized Importance
CYF (EVER IN Children, Youth, and Families Children)	0.035	100.00%
TRACTCE10	0.023	67.40%
old for grade	0.018	53.20%
HH (EVER IN Hunger and Homeless)	0.002	6.80%
MH (EVER IN Mental Health)	0.002	4.80%
HACP (EVER IN Housing Authority of City of Pittsburgh)	0.001	3.00%
DPW: if any of Department of Public Welfare-FS, SSI, and TNF	0	0.60%
School Move or not (1: Yes)	6.60E-05	0.20%

Table 10. Chronic absence by node (Grade 6-8).

CYF	Tract	Node	N	% Total Sample	% Chronic absence
Y		1	59	36.90%	44.10%
N	492900	3	26	16.20%	34.60%
N	492700;492800	4	75	46.90%	10.70%

Figure 6. Chronic absence by node (Grade 6-8).

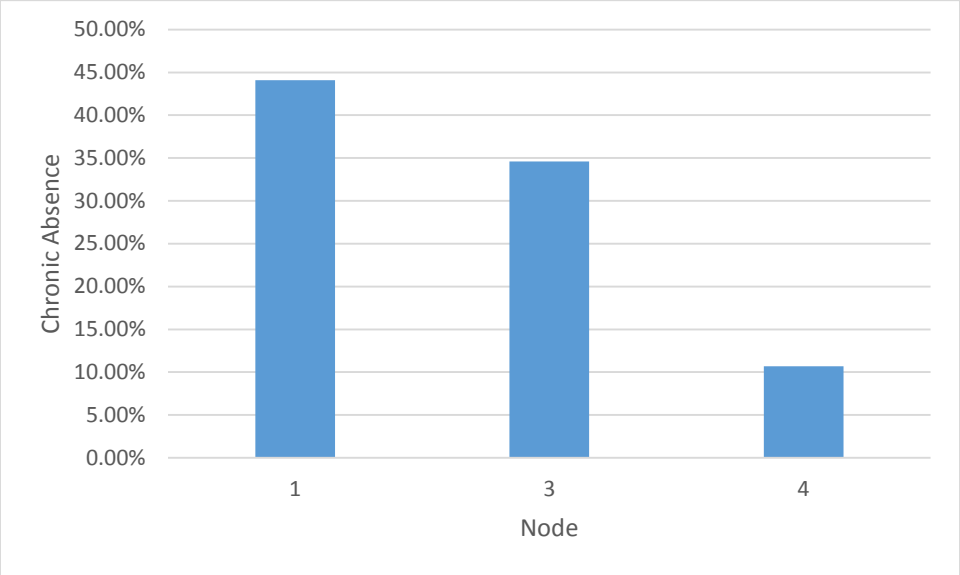
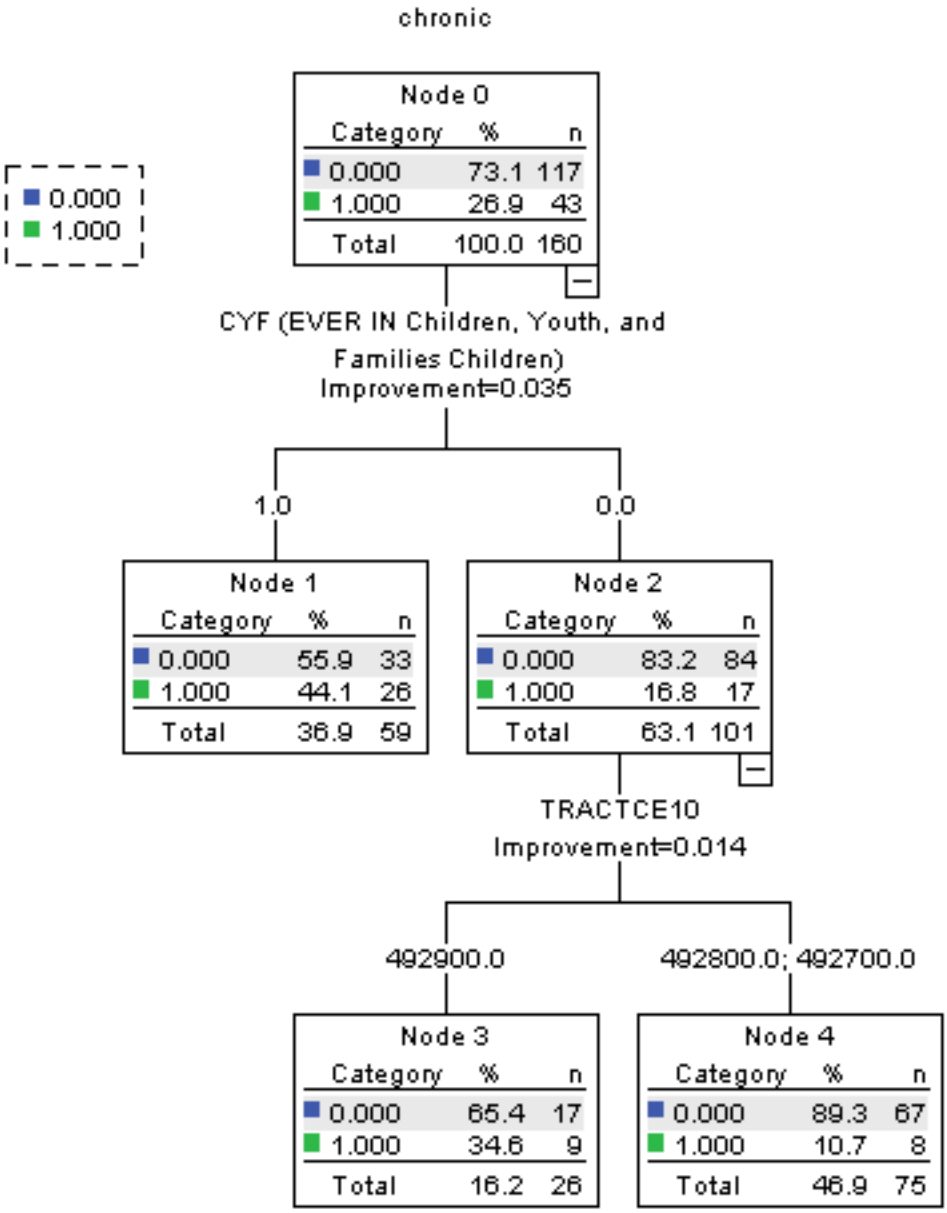


Figure 7. Tree Diagram (Grade 6-8).



3.3 Grade 9-12

- N = 181
- Only 1 predictor, Old for Grade, is significant, resulting in two nodes.
- Overall 71.3% of students were correctly classified.

Table 11. Importance table (Grade 9-12)

Independent Variable	Importance	Normalized Importance
old for grade	0.074	100.00%

Table 12. Chronic absence by node (Grade 9-12).

Old for grade	Node	N	% Total Sample	% Chronic absence
N	1	149	82.30%	30.90%
Y	2	32	17.70%	81.30%

Figure 8. Chronic absence by node (Grade 9-12).

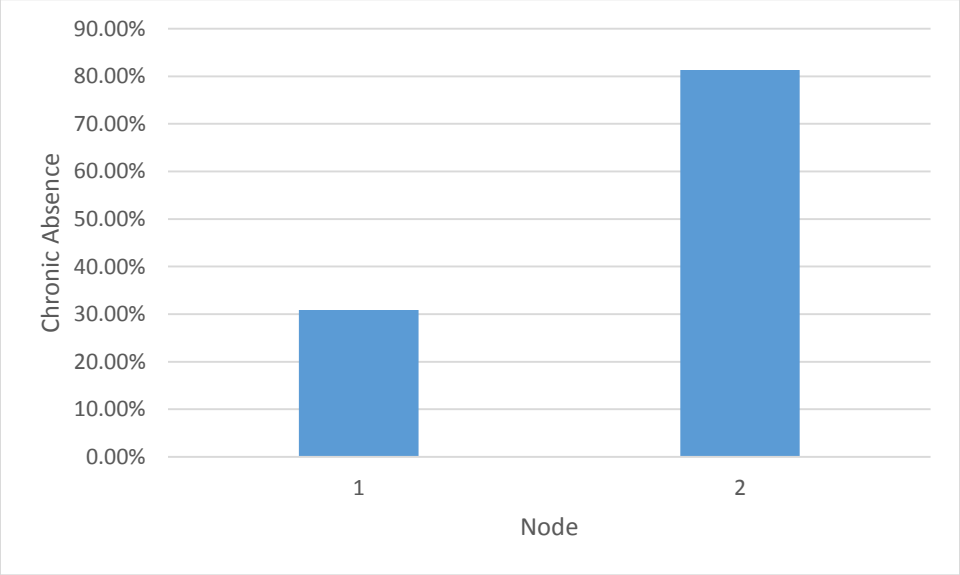


Figure 9. Tree Diagram (Grade 9-12).