Environ Monit Assess (2019) 191(Suppl 2): 332
https://doi.org/10.1007/s10661-019-7423-2

Using machine learning to examine the relationship between asthma and absenteeism

Maria-Anna Lary · Leslie Allsopp · David J. Lary · David A. Sterling

Received: 23 March 2017 / Accepted: 20 March 2019

Abstract In this study, we found that machine learning was able to effectively estimate student learning outcomes geo-spatially across all the campuses in a large, urban, independent school district. The machine learning showed that key factors in estimating the student learning outcomes included the number of days students were absent from school. In turn, one of the most important factors in estimating the number of days a student was absent was whether or not the student had asthma. This highlights the importance of environmental public health for student learning outcomes.

Keywords Asthma · Learning outcomes · Absenteeism · Machine learning · Environmental & Public Health

This article is part of the Topical Collection on *Geospatial Technology in Environmental Health Applications*

L. Allsopp · D. A. Sterling
Department of Biostatistics and Epidemiology, School of Public Health, University of North Texas Health Science Center, Fort Worth, TX 76107, USA

D. J. Lary (✉)
William B. Hanson Center for Space Sciences, The University of Texas at Dallas, 800 W. Campbell Road WT-15, Richardson, TX 75080, USA
e-mail: davidlary@utdallas.edu

M.-A. Lary
Department of Biostatistics and Epidemiology, School of Public Health, University of North Texas Health Science Center, Fort Worth, TX 75082, USA

Introduction

Graduating high school in our day and age is considered a necessity in developing a successful career. Childhood disability has overwhelming impacts on children, the education system, and the healthcare system (Newacheck and Halfon 1998). Chronically absent students, defined as missing 10% or more of a school year for any reason, are those who are at the greatest risk of failing all of their classes and dropping out. Chronic absenteeism has shown to be a predictor of not completing high school (Allensworth et al. 2014). Low-income students represent fully 78% of the chronically absent students (Balfanz and Byrnes 2012; Coelho et al. 2015). Absenteeism appears to increase during the middle school years, and continues to increase through the 12th grade (Balfanz and Byrnes 2012). Regular attendance helps students lay the fundamental building blocks for the development of more complex skills (Coelho et al. 2015). To address chronic absenteeism, one study suggests to first produce a ranked list that orders students according to their risk of not graduating on time and to then predict when they will go off track to help schools prioritize the urgency of the interventions (Aguiar et al. 2015). Many students in urban schools become disengaged at the start of the middle school, which reduces the odds that they will eventually graduate by an exponential amount. Fortunately, by combining effective whole-school reforms with attendance, behavioral, and extra-help interventions, graduation rates can be significantly increased (Balfanz et al. 2007).

Good education predicts good health, and discrepancies in health and in education achievement are closely linked. It has been demonstrated that the higher the education level, the better an individual's health is likely to be. Although education is highly correlated with income and occupation, evidence suggests that education exerts the strongest influence on health (Freudenberg and Ruglis 2007). According to the United States Centers for Disease Control and Prevention, the interrelationship of health-risk behaviors is inversely related to academic achievement. It is important that leaders in education and health work together to make wise investments in our nation's schools to benefit the present and future population of America. A system that addresses both health behavior and academic achievement would have reciprocal and synergistic effects on the health and academic achievement not only of children and adolescents, but also of adults in the USA (Bradley and Greene 2013).

An article in the *Journal of the American Academy of Pediatrics* supports that a missing element from the current national health policy is a focus on child development and achievement (Fiscella and Kitzman 2009). The less education people have from school, the higher their levels of risky health behaviors such as smoking, being overweight, or having a low level of physical activity. Seldom have health education professionals been in a better position to work together to achieve common goals and rarely has a single problem—high school dropout rates—contributed to so many adverse social, economic, and health conditions. By bringing together programs to improve health and school achievement and by making reducing school dropout rates a public health, educational, and human rights priority, public health professionals have the opportunity to make a lasting contribution to promoting population health and social justice (Freudenberg and Ruglis 2007). Educational attainment has a remarkably strong association with indicators of health, life expectancy, and reduced illness and disability (Spittel et al. 2015). Public health is well positioned to address school truancy in an evidential, comprehensive way, especially in areas of chronic disease and mental health, substance use, and parental engagement. Through partnerships with schools, healthcare providers, community-based organizations, law enforcement, and the courts, public health can provide valuable insights into this social problem by integrating health elements into the proposed solutions for combating this phenomenon affecting communities (Gase et al. 2014). When young children and adolescents miss school, they miss work, and in turn begin to fail their classes. Schools can improve the health and learning of students by supporting opportunities to learn about and practice healthy behaviors, providing school health services, creating safe and positive school environments, and engaging families and community (Michael et al. 2015). By focusing on children and youth as students, addressing critical education and health outcomes, organizing collaborative actions and initiatives that support students, and strongly engaging community resources, we offer important opportunities that will improve educational attainment and healthy development for students (Lewallen et al. 2015).

Absenteeism and asthma

A large factor that contributes to chronic absenteeism is asthma. Major steps have been made to increase the quality, quantity, and geographic availability of asthma surveillance data. However, although innovative approaches and broad public health programs have focused on minimizing the impact of asthma, the disease burden remains high. Given that the primary causes of developing asthma are only partially understood, research, prevention, and intervention efforts aimed at reducing the burden of childhood asthma remain as important as ever (Akinbami et al. 2009). However, although the causes of asthma are poorly understood, we can document that asthma disproportionately affects minority children and children with family incomes below the poverty level (US Department of Health & Human Services 2004). Practitioners in low-income communities where there are varying levels of resources and disease severity can tailor interventions to each child's needs and make substantial gains in outcomes across a range of risk profiles (Mansfield et al. 2015). Children with asthma experience more absenteeism from school compared with their non-asthmatic peers. Excessive absenteeism is related to lower student grades, and philological, social, and educational adjustment. Those with persistent asthma show a trend of performing worse on MAP standardized test scores and have more absence days compared with other students (Moonie et al. 2008). Young adults with chronic illness are significantly more likely than their healthy peers to receive public assistance and less likely to achieve important educational and vocational milestones (Maslow et al. 2011). Due to chronic conditions, health and social welfare systems are unprepared for the rapid growth in demands that will arise from these epidemics, and major increases in public expenditures should be planned for in the near term. For the longer term, the epidemic growth of childhood chronic conditions calls for major efforts to understand causation and means of prevention (Perrin et al. 2007).

Absenteeism and demographic data

A large urban, independent school district (ISD) in Texas provided fully deidentified school records for 4th through 8th grade students. Academic years in the database included 2013–2014 and 2014–2015. The database included records for all children in the grade range who were enrolled in 108 elementary and middle schools for one or both study years. A total of 55,221 complete records were available. Of these, 35,124 records were from students enrolled both years, and the remaining records were from students enrolled in only one study year. In Texas, standardized testing is referred to as the State of Texas Assessment of Academic Readiness (STAAR) test (TEA 2014a). The STAAR tests are administered to all enrolled

Table 1 The variables used in this study

Variables alphabetized within group

Group	Variable name	Variable definition	Acceptable values	Variable type
Health	Asthma	Asthma status	0 = no health alert; 1= health alert; 9 = no data collected	Categorical
Student	Age	Age of student	9 thru 17	Continuous
Student	AY	Academic year of observation	0 = 2013−2014; 1 = 2014−2015	Binary
Student	Campus	Campus Name	N/A	Categorical
Student	Gender	Gender	0 = female; 1 = male	Binary
Student	Grade	Grade Level	4 5 6 7 8	Category
Student	Race	Parent-/guardian-reported race/ethnicity	0 = White; 1 = Black; 2 = Hispanic; 3 = Other	Categorical
Student	SESbin	SES based on Free or Reduced Lunch Status	0 = no benefit (>185% poverty) 1= Free or Reduced (<185% poverty)	Binary
Student	Student	Random, generated ID	Identifier	N/A
Absence	DaysA	Total days absent	0 to 106	Continuous
Absence	DaysP	Total days present	13 thru 177	Continuous
Absence	Enroll	Total days enrolled (days present + days absent)	22 thru 177	Continuous
Absence	Penroll	Proportion of school year the student was enrolled (days enrolled/177)	0.124 thru 1	Continuous
Absence	Setbin	Binary consecutive sets	0 = 0 or 1 sets; 1 = 2 or greater sets	Binary
Absence	SetNum	Total number of sets	0 thru 24	Continuous
Academic	Mscale	Math scale score	931 thru 4359	Continuous
Academic	Mstd	Math grade level standard pass	pass = 0 fail = 1	Binary
Academic	Rscale	Reading scale Score	884 thru 3659	Continuous
Academic	Rstd	Reading grade level standard	pass = 0 fail = 1	Binary

students annually beginning in the 4th grade. Additional data made available included asthma status, race/ethnicity, Free and Reduced Lunch (FRL) status, grade level, school attended, total and date-specific absences, and total days present.

The analysis includes three attendance-related variables: (1) total absences for each student, (2) proportion of the school year a student was enrolled, and (3) sets of consecutive absence. Total absence and enrollment days are based on attendance record processes required by the state (TEA 2014b). The proportion of the school year enrolled was included as an indicator of mobility, as transfer between schools during the academic year has been identified as negatively affecting academics. This variable was calculated by adding the number of days absent and present to determine the total number of days enrolled. The result was divided by 177, the maximum number of enrollment days for the school year. Sets of consecutive absence are a novel attendance variable included to improve understanding of absence-related risk, as it has been hypothesized that both lost instruction time and learning disruptions may independently contribute to poor academic outcomes and may suggest distinct interventions to reduce risk (Goodman 2014). The consecutive sets included Monday and Friday absences that span a weekend. A binary variable for sets of consecutive days was created, grouping those with 0 or 1 set of consecutive days, and those with 2 or more sets. This allowed comparison of those with no or low exposure to sets of consecutive absences, to those with moderate to high exposure (Table 1).

Table 2 Twenty different machine learning approaches were tried to estimate the learning outcomes

Method	R^2
Random forest iterative improvement	0.94
Squared exponential Gaussian process regression	0.77
Rational quadratic Gaussian process regression	0.77
Matern 5/2 Gaussian process regression	0.77
Quadratic SVM	0.76
Coarse Gaussian SVM	0.76
Robust linear regression	0.76
Simple tree	0.76
Ensemble of boosted trees	0.76
Linear regression	0.76
Exponential Gaussian process regression	0.76
Ensemble bagged trees	0.76
Linear regression with linear interactions	0.76
Stepwise linear regression	0.76
Medium tree	0.75
Medium Gaussian SVM	0.75
Linear SVM	0.75
Complex tree	0.72
Fine Gaussian SVM	0.54

This table sorts the approaches based on their R^2 value

Demographic variables include race, family income, and grade level. FRL was used as a proxy for family income, as has been done in numerous previous studies. Free lunch benefits are extended to families with an income < 130% of poverty, and reduced-price lunch benefits are extended to those with a family income between 130 and 185%. In Texas, families receive information and an application for the program at school registration. The Free and Reduced benefit categories were combined to create a binary category indicating a family income < 185% of poverty. Race/ethnicity categories in the study were based on parent/guardian report, and identified students as Black/African American, Hispanic, White, or other. This last category included those reported to be American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, and two or more races. These last four race/ethnicity categories were collapsed due to small numbers. Grade level was categorized into a binary variable of elementary (grades 4 and 5) or middle school (grades 6–8). Based on parent-/guardian-reported race/ethnicity, the population was predominantly Hispanic (64.7%) with a substantial Black minority (23.2%), a small proportion of White students (9.3%), and the remaining 2.8% of students identifying as "Other."

In future studies, we hope to extend the analysis to include suspension data and variables describing the environment (e.g., on air quality).

Methods

A literature search for absenteeism and asthma in the Web of Science leads to around 300 articles. However, not one of these has used machine learning to try and build predictive models of the absenteeism that can also help provide insights into the underlying causes.

Consequently, in this study, we use machine learning to estimate two learning outcomes, the math scale score (Mscale) and reading scale score (Rscale). The Texas Education Administration (TEA) converts raw test scores into scaled scores that take into account the difficulty of each specific set of test questions, as difficulty levels may vary across test administrations. The resulting scaled scores allow for direct comparison of student performance within each grade level (TEA 2014a, b). Machine learning is useful in this context for at least two reasons. First, it provides an accurate empirical model. Second, it provides some insight into the relative importance of the available factors.

A previous study has also used machine learning to select and prioritize students who may be at risk of not graduating high school on time (Aguiar et al. 2015). However, this study did not include asthma status or other health information into the analysis.

Nearly 20 different machine learning approaches were tried. The one which performed best was a random forest with iterative correction described in further detail in the next section (Table 2).

Random forest

A random forest is an ensemble statistical machine learning approach, consisting of an ensemble of decision trees (Breiman 1984, 2001; Ho 1998). Random forests have proved to be a very useful multivariate, non-linear, non-parametric approach for both regression and supervised classification. Ensemble methods are less prone to over-learning the noise of the data and typically provide better generalization. A random forest also provides a useful ranking of the relative importance of the predictors, an example of which is shown in Fig. 3 for estimating the real valued learning outcomes. To decide how many trees we should use in our random forest, we examined how the error decreased as the number of trees in the ensemble (forest) is increased (Fig. 2).

To ensure that we can independently validate our estimates of the learning outcomes, we randomly split up the available data into a training dataset and an independent validation dataset. The random forest training uses only the training data; the independent validation dataset is held back and used for a completely independent validation. As can be seen in the scatter diagrams shown in Fig. 1, our approach performs very well indeed.

A random forest can facilitate the estimation of a continuous variable, v, such as the learning outcomes, as a multi-variate, non-parametric function of N input variables, i.e.:

$$v = f_{RandomForest}(x_1, \ldots, x_N) \quad (1)$$

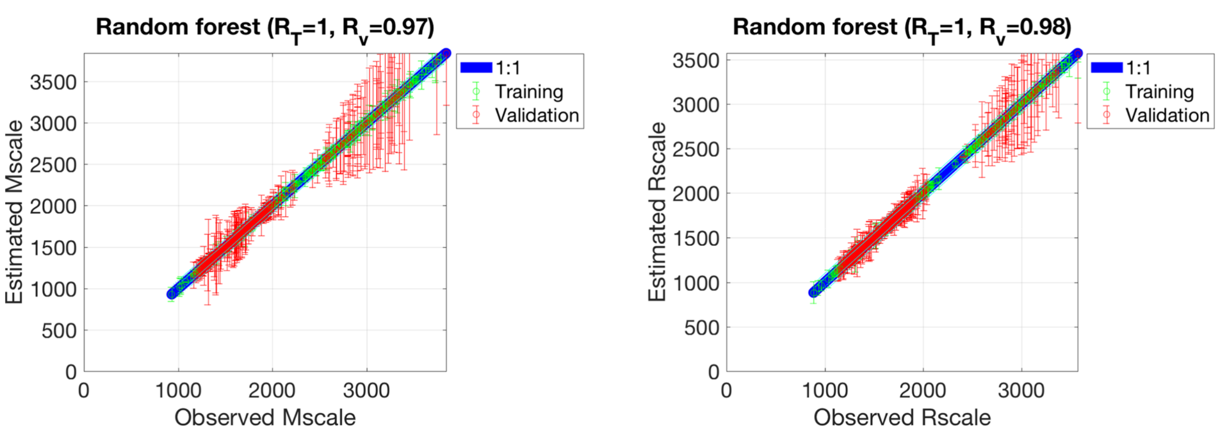

Fig. 1 Scatter diagrams for the two learning outcomes we estimated for all students in a school district over 2 years, Mscale (left) and Rscale (right). In each case, the actual learning outcome is plotted horizontally, and the estimated learning outcome using machine learning (4) is plotted vertically. The perfect 1:1 line, with a slope of 1 and an intercept of 0, is the thick, dark blue line. To ensure that there is independent validation of the estimated learning outcomes, we randomly split up the available data into a training dataset (green points) which is used in the training of the random forests and an independent validation dataset (red points) that is not used in the training of the random forests. The error bars are the estimated uncertainties ϵ_{RF} (2). The random forest training uses only the training data; the independent validation dataset is held back and used for a completely independent validation. In each case, the plot title shows the correlation coefficient for the training dataset (R_T) and the independent validation dataset (R_V). As can be seen, the approach performs well

where $x_1, \ldots, x_N$ are the N input parameters (listed in Table 2).

Error estimates and error correction

Two enhancements were then made that allowed both an improvement of the performance and provided an estimated error, ϵ, for each learning outcome. The enhancement was inspired by Newton-Raphson iteration.

A series of iterations were executed; for each iteration, a random forest was used to estimate the learning outcome variable, v_{RF}, as indicated in Eq. 1. Then, the learning outcome variable, v_{RF}, was compared with the actual learning outcome from the training dataset, v_{Actual}, to calculate the error of the random forest estimate, ϵ_{RF}, i.e.:

$$\epsilon_{RF} = v_{Actual} - v_{RF} \tag{2}$$

Then, an additional random forest was used to learn this error, ϵ_{RF}, as a function of the N input variables and the random forest estimate of the learning outcome variable, v_{RF}, estimated in Eq. 1, i.e.:

$$\epsilon_{RF} = f_{RandomForest}(x_1, \ldots, x_N, v_{RF}) \tag{3}$$

Then, this estimated error, ϵ_{RF}, is used to correct the initial random forest estimate of the learning outcome, v_{RF}, i.e., by rearranging (2) to give an improved estimate of the learning outcome, $v_{Improved\ Estimate}$, i.e.:

$$v_{Improved\ Estimate} = v_{RF} + \epsilon_{RF} \tag{4}$$

This was then repeated for a set of n iterations (in this case, n=2 was sufficient). After each iteration, the estimated learning outcome and estimated learning outcome error were added as additional input variables for the next iteration. This considerably improved the reliability of our estimated learning outcomes, as can be seen in Fig. 1.

Results and discussion

Figure 1 shows scatter diagrams for the two learning outcomes we estimated, Mscale (left) and Rscale (right). In each case, the actual learning outcome is plotted horizontally, and the estimated learning outcome using machine learning (4) is plotted vertically. The perfect 1:1 line, with a slope of 1 and an intercept of 0, is the thick, dark blue line. To ensure that there is independent validation of the estimated learning outcomes, we randomly split up the available data into a training dataset (green points) which is used in the training of the random forests and an independent validation dataset (red points) that is not used in the training of the random forests. The error bars are the estimated uncertainties ϵ_{RF} (2). The random forest training uses only the training data; the independent validation dataset is held back and used for a completely independent validation. In each case, the plot title shows the correlation coefficient for the training dataset (R_T) and the independent validation dataset (R_V). As can be seen, the approach performs well.

Number of required ensemble members

As mentioned earlier, a random forest is an ensemble statistical learning approach, consisting of an ensemble of decision trees (Breiman 1984, 2001; Ho 1998). Ensemble methods are less prone to overlearning the noise of the data and typically provide better generalization. To decide how many trees we should use in our random forest, we examined how the error decreased as the number of trees in the ensemble (forest) is increased. Figure 2 shows the total error as a function of the number of trees in the random forest for estimating the learning outcomes Mscale (left) and Rscale (right). We used 50 independent estimators (trees in our random forest) as increasing the number of independent estimators beyond 50 increases the computational expense but does not yield any significant additional error reduction.

Variable importance

A random forest provides a useful ranking of the relative importance of the predictors for estimating the real valued learning outcomes. Figure 3 shows the relative importance of the input variables for estimating the learning outcomes Mscale (left) and Rscale (right). The random forests show that the three most important parameters in estimating Mscale are the student's age, the number of days absent, and which campus they attend. The random forests show that the three most important parameters in estimating Rscale are the student's age, which campus they attend, and the number of days absent.

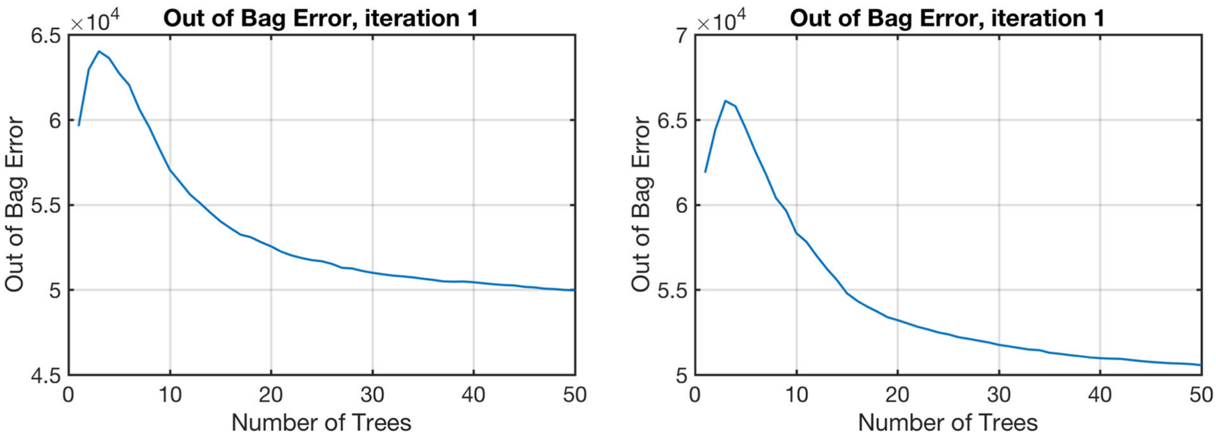

Fig. 2 The total error as a function of the number of trees in the random forest for estimating the learning outcomes: Mscale (left) and Rscale (right). We used 50 independent estimators (trees in our random forest) as increasing the number of independent estimators beyond 50 increases the computational expense but did not yield a significant additional error reduction

Estimating absenteeism

Having seen above that the number of days that a student is absent is one of the most important factors in determining the learning outcomes, let us now turn our attention to see which factors are the most important in estimating how many days a student is absent. We did this by using an additional random forest to estimate the number of days that a student is absent. The left plot of Fig. 4 shows the relative importance of the input variables for estimating the number of days a student was absent. The random forests show that one of the most important parameters in estimating the number of days a student was absent was whether or not they had asthma.

The right plot of Fig. 4 is a scatter diagram for the estimation of the number of days a student was absent. The actual number of days a student was absent is plotted horizontally, and the estimated number of days a student was absent using machine learning is plotted vertically. The perfect 1:1 line, with a slope of 1 and an intercept of 0, is the thick, dark blue line. To

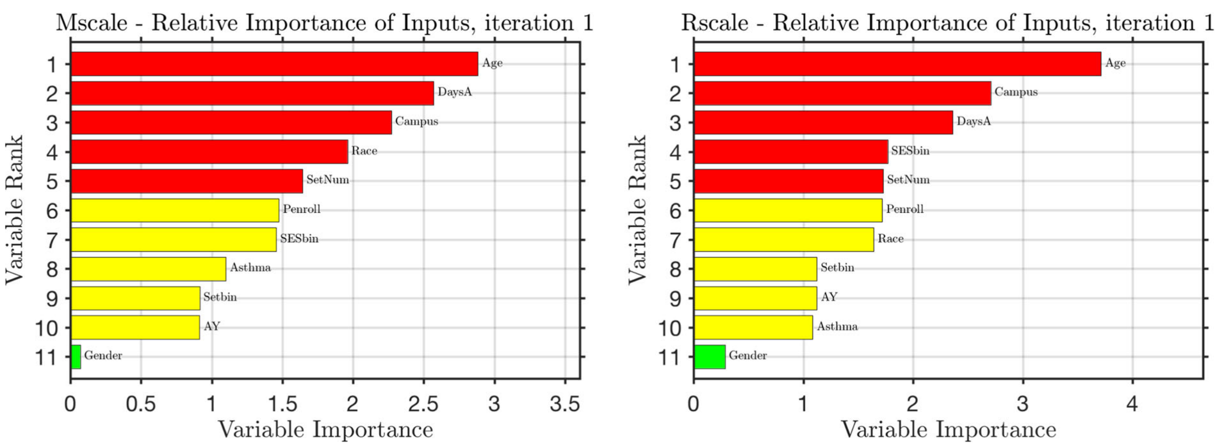

Fig. 3 The relative importance of the input variables for estimating the learning outcomes: Mscale (left) and Rscale (right). The random forests show that the three most important parameters in estimating Mscale are the student's age, the number of days absent, and which campus they attend. The random forests show that the three most important parameters in estimating Rscale are the student's age, which campus they attend, and the number of days absent

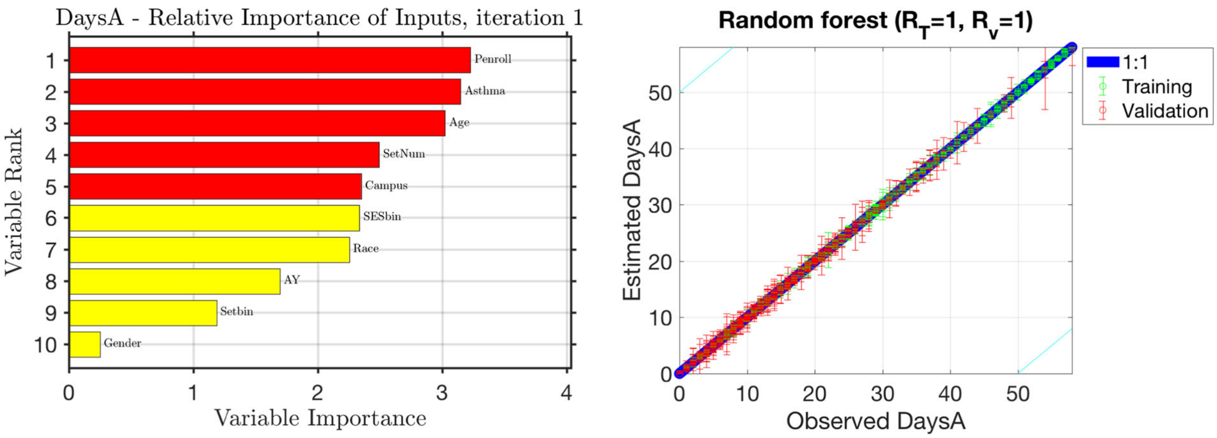

Fig. 4 The left plot shows the relative importance of the input variables for estimating number of days a student was absent for all students in a school district over 2 years. The random forests show that one of the most important parameters in estimating the number of days a student was absent was whether or not they had asthma. The right plot is a scatter diagram for estimation of the number of days a student was absent (DaysA). The actual number of days a student was absent is plotted horizontally, and the estimated number of days a student was absent using machine learning is plotted vertically. The perfect 1:1 line, with a slope of 1 and an intercept of 0, is the thick, dark blue line. To ensure that there is independent validation of the estimated learning outcomes, we randomly split up the available data into a training dataset (green points) which is used in the training of the random forest and an independent validation dataset (red points) that is not used in the training of the random forest. The error bars are the estimated uncertainties. The random forest training uses only the training data; the independent validation dataset is held back and used for a completely independent validation. The plot title shows the correlation coefficient for the training dataset (R_T) and the independent validation dataset (R_V). As can be seen, the approach performs well

ensure that there is independent validation of the estimated learning outcomes, we randomly split up the available data into a training dataset (green points) which is used in the training of the random forests and an independent validation dataset (red points) that is not used in the training of the random forests. The error bars are the estimated uncertainties. The random forest training uses only the training data; the independent validation dataset is held back and used for a completely independent validation. The plot title shows the correlation coefficient for the training dataset (R_T) and the independent validation dataset (R_V). As can be seen, the approach performs well.

Conclusion

In this study, we found that machine learning was able to effectively estimate student learning outcomes geo-spatially across all the campuses in a large, urban, independent school district (ISD) in Texas. The machine learning showed that key factors in estimating the student learning outcomes included the number of days students were absent from school. In turn, one of the most important factors in estimating the number of days a student was absent was whether or not the student had asthma. This highlights the importance of environmental public health for student learning outcomes. These results underscore the potential importance of effective school-based asthma management programs, which have been robustly demonstrated to reduce absenteeism among affected children (Walter et al. 2016).

In future studies, we hope to extend the analysis to include suspension data and variables describing the environment (e.g., on air quality).

Acknowledgements This work was supported with the involvement of the Fort Worth Independent School District.

Funding information This work was supported in part through funding from Glazow Smith Kline and Boehringer Inngleheim and the UNT Health Science Center for Professional and Continuing Education.

Compliance with ethical standards The Institutional Review Board for the UNT Health Science Center reviewed and approved the collection and use of the research method and data (IRB no. 2013-222).

References

Aguiar, E., Lakkaraju, H., Bhanpuri, N., Miller, D., Yuhas, B., Addison, K.L. (2015). Who, when, and why: A machine learning approach to prioritizing students at risk of not graduating high school on time. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 93–102): ACM.

Akinbami, L.J., Moorman, J.E., Garbe, P.L., Sondik, E.J. (2009). Status of childhood asthma in the united states, 1980–2007. *Pediatrics*, *123*(Supplement 3), S131–S145.

Allensworth, E.M., Gwynne, J.A., Moore, P., de la Torre, M. (2014). Looking forward to high school and college: Middle grade indicators of readiness in chicago public schools. *University of Chicago Consortium on Chicago School Research*.

Balfanz, R., Herzog, L., Mac Iver, D.J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223–235.

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

Bradley, B.J., & Greene, A.C. (2013). Do health and education agencies in the United States share responsibility for academic achievement and health? a review of 25 years of evidence about the relationship of adolescents' academic achievement and health behaviors. *Journal of Adolescent Health*, *52*(5), 523–532.

Breiman, L. (1984). *Classification and regression trees*. Belmont: The Wadsworth Statistics/Probability Series. Wadsworth International Group.

Breiman, L. (2001). *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Coelho, R., Fischer, S., McKnight, F., Matteson, S., Schwartz, T. (2015). The effects of early chronic absenteeism on third-grade academic achievement measures. In *Workshop in Public Affairs*.

Fiscella, K., & Kitzman, H. (2009). Disparities in academic achievement and health: the intersection of child education and health policy. *Pediatrics*, *123*(3), 1073–1080.

Freudenberg, N., & Ruglis, J. (2007). Reframing school dropout as a public health issue.

Gase, L.N., Kuo, T., Coller, K., Guerrero, L.R., Wong, M.D. (2014). Assessing the connection between health and education: Identifying potential leverage points for public health to improve school attendance. *American Journal of Public Health*, *104*(9), e47–e54.

Goodman, J. (2014). Flaking out: Student absences and snow days as disruptions of instructional time. Technical report, National Bureau of Economic Research.

Ho, T.K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(8), 832–844.

Lewallen, T.C., Hunt, H., Potts-Datema, W., Zaza, S., Giles, W. (2015). The whole school, whole community, whole child model: a new approach for improving educational attainment and healthy development for students. *Journal of School Health*, *85*(11), 729–739.

Mansfield, C., Viswanathan, M., Woodell, C., Ohadike, Y.U., Kennedy Lesch, J., Malveaux, F.J. (2015). Improving asthma outcomes for children: many paths to progress. *Pediatric Allergy, Immunology, and Pulmonology*, *28*(2), 80–86.

Maslow, G.R., Haydon, A.A., Ford, C.A., Halpern, C.T. (2011). Young adult outcomes of children growing up with chronic illness: an analysis of the national longitudinal study of adolescent health. *Archives of Pediatrics & Adolescent Medicine*, *165*(3), 256–261.

Michael, S.L., Merlo, C.L., Basch, C.E., Wentzel, K.R., Wechsler, H. (2015). Critical connections: health and academics. *Journal of School Health*, *85*(11), 740–758.

Moonie, S., Sterling, D.A., Figgs, L.W., Castro, M. (2008). The relationship between school absence, academic performance, and asthma status. *Journal of School Health*, *78*(3), 140–148.

Newacheck, P.W., & Halfon, N. (1998). Prevalence and impact of disabling chronic conditions in childhood. *American Journal of Public Health*, *88*(4), 610–617.

Perrin, J.M., Bloom, S.R., Gortmaker, S.L. (2007). The increase of childhood chronic conditions in the united states. *Jama*, *297*(24), 2755–2759.

Spittel, M.L., Riley, W.T., Kaplan, R.M. (2015). Educational attainment and life expectancy: a perspective from the nih office of behavioral and social sciences research. *Social Science & Medicine*, *127*, 203–205.

TEA (2014a). Texas Education Agency. STAAR performance standards. http://tea.texas.gov/student.assessment/staar/performance-standards/. Accessed: 2017-03-21.

TEA (2014b). Texas Education Agency. student attendance accounting handbook 2014-2015. http://tea.texas.gov/Finance_and_Grants/Financial_Compliance/Student_Attendance_Accounting_Handbook/. Accessed: 2017-02-20.

US Department of Health & Human Services, U. (2004). The president's task force on environmental health risks and safety risks to children. Available: [Accessed 13 September 2004].

Walter, H., Sadeque-Iqbal, F., Ulysse, R., Castillo, D., Fitzpatrick, A., Singleton, J. (2016). Effectiveness of school-based family asthma educational programs in quality of life and asthma exacerbations in asthmatic children aged five to 18: a systematic review. *JBI Database of Systematic Reviews and Implementation Reports*, *14*(11), 113–138.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.