

Healthy **Bodies**, Bright **Minds**

The Impact of **Healthcare Access**
on **Children's Academic performance**

Ayomikun Adeniran

Nick Castillo

Samara Chamoun

Anthony Kling

Edward Varvak

Glenn Young



THE ERDŐS INSTITUTE

Helping PhDs get jobs they love.

Helping you hire the PhDs you need.

PROBLEM AND DATA

We are interested in exploring the effect of **healthcare access** on **academic performance**.

Research shows that **school absences** have **negative** impact on grades and **students' academic achievement**.

The National Survey of Children's Health (NSCH) dataset



- For the health data, we use **NSCH** dataset to generate our "**access to healthcare**" **features** (e.g. children's current health care coverage, how often the child is allowed to see providers).
- Model will include other **health related features**

Predicting variables: access to healthcare features, other health-related features

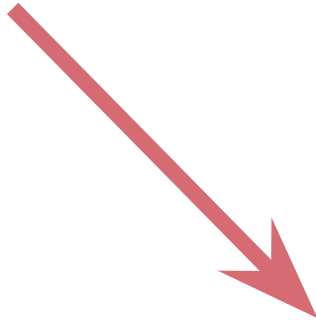


- For the educational data, we use "**days missed in school**" from **NSCH** as a metric of student **educational outcome**, to avoid introducing bias by gluing separate datasets.

Target variable: days missed in school

EDA AND IMPUTING

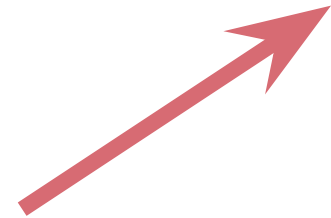
Dropped NaNs from target variable (days_missed)



Replaced conditional NaN values with 0

(e.g. healthcare cost = \$0, then costs_reasonable = NaN)

We used a random forest classifier to **impute** missing values in the training data



FEATURE SELECTION: 3 METHODS

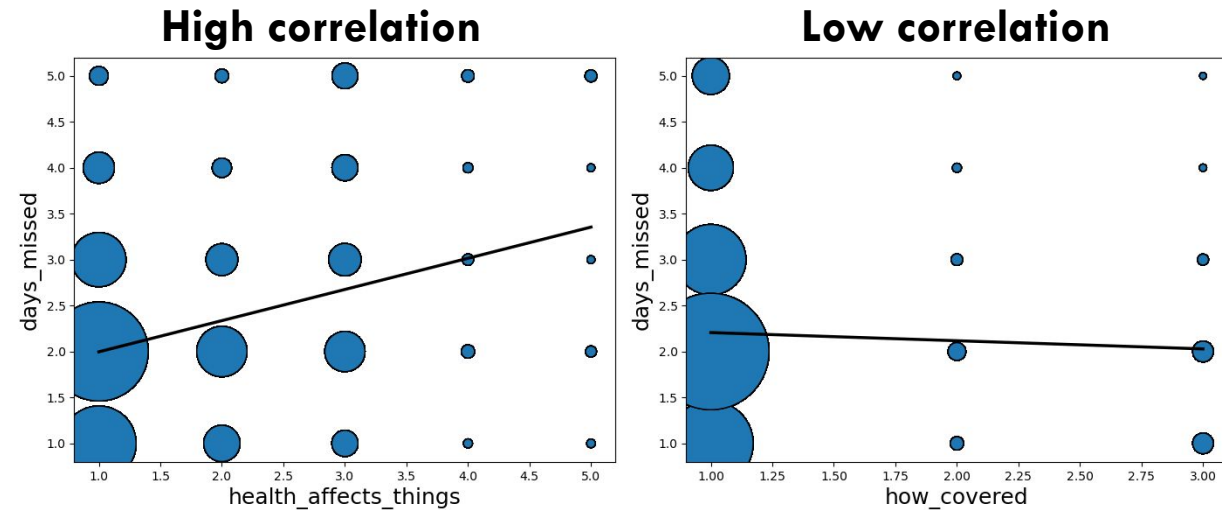


1. **Handpick:** We parsed through the **447 features** in the NSCH dataset, picking any **related to healthcare access**



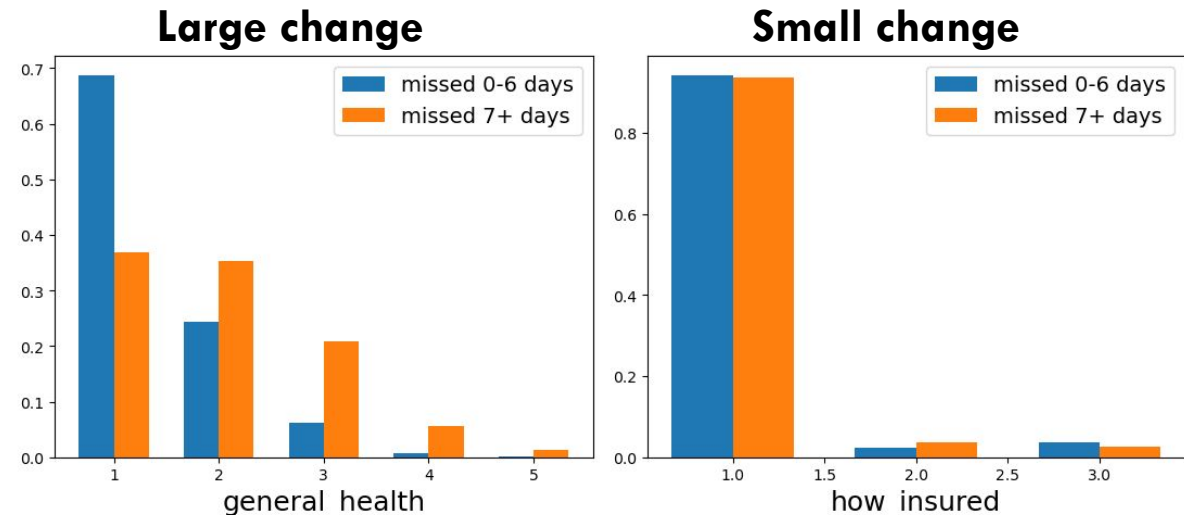
2. **Correlation analysis:**

We computed the **linear correlation** between **each feature** and the **number of days missed**, keeping features with high correlation



3. **Histogram analysis:**

For **each feature**, we measured the change in histogram shape among children with **low** and with **high absenteeism**, keeping features with sufficiently different histograms



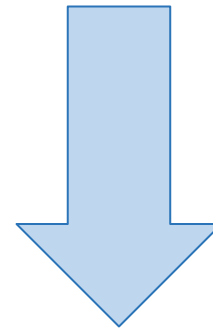
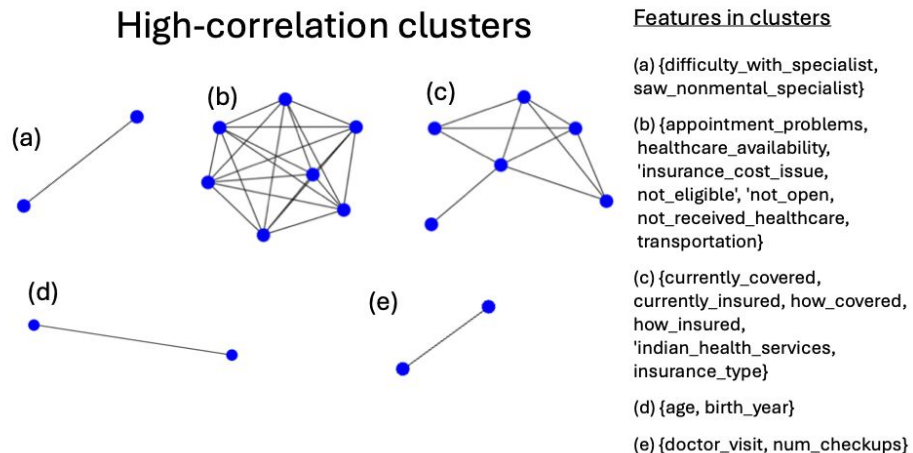
MODEL AND FEATURE SELECTION 2.0

Model selection

- We trained and evaluated a **logistic regression** classifier, a **random forest** classifier, a **support vector** classifier, and a **KNN classifier** to predict whether children will miss more than 7 school days.
- Of these, logistic regression performed best on the metrics of **recall** and **average precision score**.

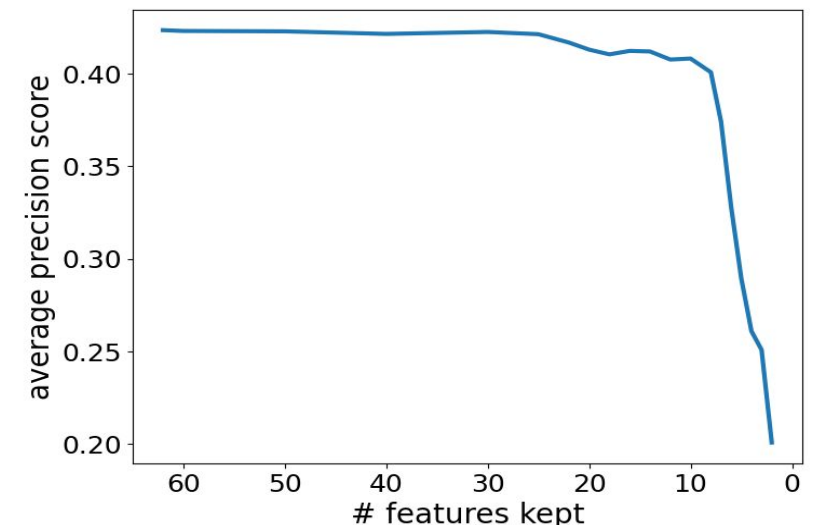
Feature selection 2.0

- We first checked for collinearity and eliminated all but one feature from each highly co-linear “cluster” found



Through this process, we constructed out a model using **10 total features**

- We then used recursive feature elimination (RFE) to narrow down our features until model performance was affected

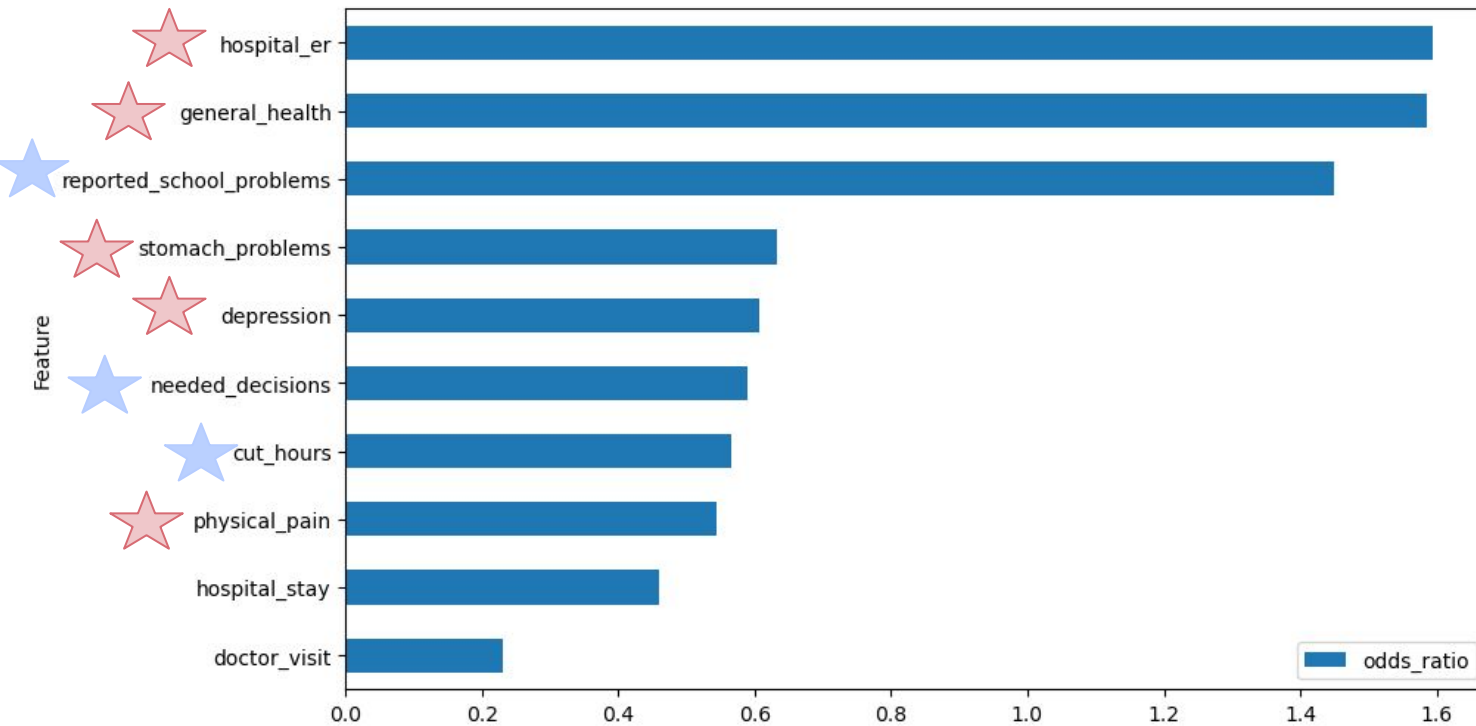


RESULTS



Overall, we found that **poor health** was strongly related to **absenteeism**

We computed the **odds-ratios** of our ten-parameter model



★ Specifically, we found that a **higher number of missed days** was predicted by **poorer general health** and more **time spent in the hospital**, as well as the presence of **depression, chronic physical pain, and digestive problems**

★ Additionally, children who reported having **problems at school**, needed **healthcare-related decisions** made on their behalf, or experienced **health problems for which their family needed to cut work hours** were also found to be more likely to miss school



Many features
related to health
affect absenteeism!

But, how about the
features related to
health care access?

CONCLUSIONS AND FUTURE DIRECTIONS



shutterstock.com · 113198836



- The preliminary results suggest that **access to health care is not the strongest predictor** of **child absenteeism**
- It is possible that the relationship between access to health care and absenteeism was **drowned out by the more potent predictors of missed days**, such as the general health of the child. A **future study** could **control for predictors** which are more related to access to healthcare
- Likewise, it is possible that absenteeism is a poor metric for education outcomes; **future work** could **try other metrics, such as grades or scores on standardized tests**

ACKNOWLEDGMENT

Thank you to:

- Roman Holowinsky
- Alec Clott
- Steven Gubkin
- Amalya Lehmann
- Our mentor: Evelyn Huszar
- All the Erdős Institute

thank you



THE ERDŐS INSTITUTE

Helping PhDs get jobs they love.
Helping you hire the PhDs you need.

HRSA
Health Resources & Services Administration