The University of Chicago

# Upward Mobility:
# Optimizing Information to Improve Graduation Rates for America's 407,000 Youth in Foster Care

By: Summer Long

A thesis submitted for partial fulfillment of the requirements for the degree of
BACHELOR OF ARTS in PUBLIC POLICY STUDIES
at THE UNIVERSITY OF CHICAGO

Public Policy Studies Preceptor: Maria Bautista
Second Reader: Amy Dworsky

April 17, 2023

**Abstract**

In the United States, there are 407,000 youth in foster care. Of those, nearly 30,000 reach the age of adulthood (age 18 to 21) and are discharged—disconnected from the vital resources that stand-in for familial structures and inter-generational wealth. High school completion is a powerful and accessible avenue to achieve social mobility that is unconditional on one's socioeconomic background or foster care status, yet literature does not exist addressing the topic of high school completion broadly for youth in foster care due to the ethical concerns of qualitative studies and lack of accessible, unified administrative data for quantitative studies. This paper attempts to fill this gap by utilizing the National Youth in Transition Database's (NYTD) survey data and the three completed Cohorts: Cohort 2011, Cohort 2014, and Cohort 2017.

Findings suggest that youth in foster care complete high school at rates far lower than their peers—this paper provides an optimistic estimate of at least 20 percentage points lower. Those who complete high school while in foster care follow similar trends to national graduation rates based on demographic characteristics, and there is evidence of a statistically significant association between all but one characteristic captured in the NYTD data and high school graduation. There is evidence that some characteristics captured in the NYTD data are significant predictors for high school graduation by age 19. However, the attempted deployment of machine learning models and evaluation of their efficacy of predicting high school graduation suggests that data collection standards, data quality, and types of data collected should be revisited by both federal and state governments.

**Acknowledgements**

I would like to thank those who have been supportive in my journey of writing this thesis, as it has led to a great deal of academic, professional, and personal growth. I'd like to thank my preceptor, Maria Bautista, for challenging me and encouraging me to pursue this topic despite the slight hesitancy from others. I am also extremely grateful for Amy Dworsky, who has always been incredibly responsive and supportive—her wisdom, knowledge, and quality of work in the field is extremely inspiring and it is a pleasure to have someone so qualified as my second reader. A final thank you to my friends and family, who have always believed in me.

**Table of Contents**

## Introduction

Today's youth are tomorrow's leaders—and there are 73 million Americans under age 18 who fit the bill (Office of Juvenile Justice and Delinquency Prevention, 2020). However, not all youth grow up in a stable environment. Some endure traumas such as abuse and neglect. Others are not adequately cared for because their families are in poor socioeconomic status—not given the correct resources to improve their children's quality of life. When these scenarios happen, youth are placed into foster care.

Foster care encompasses not only certified homes with the youth, but multiple alternative living spaces. These can include group homes, residential care facilities, emergency shelters, and supervised independent living. Furthermore, youth who are placed into foster care may live with another relative or a trusted adult in their life, such as a coach or teacher (Child Welfare Information Gateway). Regardless of placement arrangement, foster care is a disruptive force in a youth's life that can cause severe physical and mental effects.

407,000 youth were in the United States foster care system in 2020 (Administration for Children & Families, 2021). With a lack of intergenerational wealth, resources, and guidance once they age out of the system, youth in foster care are uniquely disadvantaged as they enter adulthood. One way to mitigate this is by opening pathways to social mobility. Increasing the amount of post-secondary opportunities available as they enter adulthood is critical.

A vital prerequisite to a vast majority of these opportunities is the completion of secondary education—either through the attainment of a high school diploma or an equivalent such as a GED. Increased earnings, employment, college admittance, vocational education, better health outcomes, and lower incarceration rates are all locked behind high school completion (Hahn et al., 2015, Sum et al., 2009). While youth in foster care lose vital support around the time when they graduate high school in many states, having a high school education provides them with an essential springboard into adult success.

Increasing high school completion should be a key goal in the United States for all youth. However, this paper will evaluate specifically youth in foster care for a key reason: they face unique challenges in completing high school compared to their peers. They face higher rates of school disruption in three ways: suspensions, school change, and incarceration. Firstly, they have higher suspension rates from school compared to their peers (Scherr, 2007). Secondly, youth in foster care experience high rates of school mobility, resulting in disruptions in education continuity that can hinder performance (Clemens et al., 2016). Finally, they face higher rates of incarceration: about 70% of youth who exit foster care as legal adults are arrested at least once by age 26 (Courtney et al., 2011).

The unique background of youth in foster care that facilitate their transition into the system, as well as their experiences living within it, typically means that youth in foster care have negative childhood experiences and associated trauma. The lack of positive childhood experience, particularly with parental attachment, can lead to hindrance of cognitive development that makes the playing field uneven (West et al., 2013).

Given that youth in foster care face unique challenges compared to their peers of similar demographic and socioeconomic backgrounds, it is critical to evaluate the question of high school graduation specifically for this population. Researchers currently focus on post-secondary attainment due to the lack of ethical barriers with an adult population as well as the ease of information tracking through the FAFSA, leaving this particular topic disproportionately under-researched. Papers that have explored this topic focus on one particular facet or within a particular state, leaving the others undiscussed. This paper will explore what factors contribute to the attainment of a high school diploma or equivalent for youth in foster care by age 19 overall.

**Literature Review**

The Importance of a High School Diploma

There is a strong need for a high school diploma or equivalent in the United States. The median yearly earnings for one without a high school diploma is $32,641.40—while adults who hold a highschool diploma's median yearly earnings is $42,183.60 (Bureau of Labor Statistics, 2021). Employment opportunities increase when one attains a high school diploma: the unemployment rate for those without a high school diploma was 8.3%, opposed to 6.2% for those who only have one (Bureau of Labor Statistics, 2021). In short, having a high school diploma is a powerful way to access social mobility, as it is not dependent on one's unchangeable history and circumstances.

In addition to allowing one to access social mobility, a high school diploma leads to better health outcomes. Those with a high school diploma are less likely to have a chronic illness (Vaughn, Salas-Wright, & Maynard, 2014). Conversely, high school dropouts are more likely to experience poor health and premature death (Hahn et al., 2015). If the rate of high school dropout was slashed in half, the United States would save $7.3 billion on Medicaid spending annually (DeBaun & Roc, 2013). Dropping out of high school is also correlated to higher rates of incarceration. One out of ten young, male high school dropouts are incarcerated compared to one out of thirty-five high school graduates (Sum et al., 2009).

While obtaining a GED is considered an equivalent to a high school diploma, the positive effects of having one are less than having a diploma. GED holders earn less on average than those who hold a high school diploma—but still far greater than those without either

(Census.gov, 2012). While on-time high school graduation produces the best outcomes, obtaining a GED is still a viable method to mitigate negative effects in the long-term—as many studies consider the two to be equivalent in measuring outcomes.

<u>The Lack of Current Studies for Youth in Foster Care Broadly</u>

When evaluating descriptive statistics, such as the rate of youth in foster care that have received a high school diploma or equivalent, there is not much literature evaluating the topic broadly. While high school or equivalent completion is a prerequisite for the large majority of higher education institutions, there is a significantly higher amount of literature pertaining to higher education matriculation, retention, and completion.

One potential reason behind this is that youth and their higher educational outcomes are easily traced through Question 52 on the FAFSA, which asks if one has a history in foster care to determine Independent status. The National School Clearinghouse and its opt-out system of reporting yields high-to-near complete yield. Additionally, the numerous higher education interventions at the state level—such as tuition vouchers and stipends—and the amount of monetary investment involved incentivizes states and researchers alike to explore if these interventions are effective.

Qualitatively, studies on youth in foster care and high school completion pose unique challenges compared to higher education. The majority of youth in high school are minors, meaning there are ethical barriers to asking insightful yet highly sensitive questions. However, in higher education, these youth are now adults and have more autonomy in their participation. One qualitative study that explores the high school environment for youth in foster care is retrospectively administered while the cohort is in college (Sandh et al., 2020)

The studies that have provided descriptive statistics on the graduation rates for youth in foster care focus at the state level using administrative data from the respective state. In Colorado's foster care system, the on-time graduation rate for youth ranged from 27.5% in Academic Year 2012-2013 to 30% in Academic Year 2014-2015 (Clemens et al., 2016). In Washington, the on-time graduation rate for youth in foster care was 50% in Academic Year 2010-2011 (Burley, 2013). However, existing literature tends to address one factor in isolation instead of multiple.

<u>The Unique Challenges of Youth in Foster Care in Completing High School</u>

Literature suggests a few reasons why youth in foster care graduate at a lower rate than their peers who are not in care. Firstly, youth in foster care experience a higher rate of school mobility—defined as the amount of school changes a youth experiences in their K-12 education. For example, youth in Colorado had 3.46 school changes on average during high school (Clemens et al., 2016). Odds of graduating high school were 39% lower with one additional school change, and African American youth experienced the highest number of school changes on average (Clemens et al., 2016). School changes can lead to educational interruptions due to delay periods in re-enrollment into a new school and change in requirements and prerequisites compared to previous institutions.

Similarly to school changes, suspensions were another source of disruption that were significant. Students are less likely to graduate from high school when they experience out of school suspensions compared to their peers—even if it is just once (Lenderman & Hawkins, 2021). Given that 24% of youth in foster care have been suspended at least once, it is a common school disruption for the population (Scherr, 2007).

Furthermore, youth in foster care disproportionately attend lower-quality schools. These schools are often identified by metrics such as the amount of children achieving satisfactory results on exams or the number of youth that dropout. One analysis of Chicago Public Schools revealed that of high schools with 20 or more students enrolled that are in foster care, 94% had 4-year dropout rates for freshmen that were 30% or above (Smithgall et al., 2004). Comparatively, only 48% of other CPS schools have dropout rates exceeding 30%—a 46% difference!

Experiences prior to entering foster care can also play a role in whether or not a youth will graduate high school on-time or at all. Youth who have experienced confirmed cases of sexual abuse are less likely to graduate compared to their peers in foster care who have not (Okpych et al., 2017). In order to enter foster care, youth have demonstrated less positive parental involvement and attachment. Parental involvement and attachment are important for developing childrens' cognitive skills, as children are able to build these skills without being hindered by insecurity in their environment (West et al., 2013). With cognitive skill development hindered, lower high school graduation rate follows. Notably, youth struggling with substance abuse or had ever received a referral to explore if they have a substance abuse disorder—something that can hinder cognitive skill development—were also less likely to graduate high school (Okpych et al., 2017).

**Historic and Current Measures**

Federal support for accessible and supportive K-12 education for all youth, regardless of background and circumstance, flourished in the mid 1900's. The Richard B. Russell National School Lunch Act was signed into law in 1946, establishing the National School Lunch Program. This program provides free or reduced-cost lunch to students who qualify through subsidies to schools. The Child Nutrition Act in 1966 established the School Breakfast Program, an extension of the National School Lunch Program applying to breakfast. The most notable measure for accessibility of public education would come in 1965 in the form of the Elementary and Secondary Education Act (ESEA).

The ESEA provides federal funding for primary and secondary funding to schools, which otherwise saw funding from their state governments. The Title I program sought to provide targeted funding for schools that had a large sum of low-income students by calculating the percentage of children ages 5 to 17 that lived with families below the “low-income” threshold. Despite needing extra resources due to their unique circumstances, youth in foster care were not considered a category that warranted additional funding. They are not inherently grouped into the “low-income” category, as not all foster care homes were “low-income.”

Amendments to the ESEA were made over the years aimed at extending the bill and improving the quality of it. However, the foster care population was never explicitly mentioned until 1988 when the Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments were signed into law. These amendments reauthorized the ESEA and expanded the definition of children to be counted for Title I funding to include children residing in foster homes. This potentially allowed schools with higher concentrations of foster youth who otherwise did not have Title I funding to now access it. This was the first federal bill that implicitly aimed to improve the quality of education for youth in foster care in particular.

In 1994, the Improving America’s Schools Act (IASA) reauthorized and redefined the goal of the ESEA. It addressed the unique circumstances of foster youth navigating school systems. In Title XI (“Coordinated Services”), the stated purpose is combatting the “growing numbers of children… negatively affected by influences outside of the classroom which increase such children's risk of academic failure." The influences outlined in the bill include unsafe living conditions, physical and sexual abuse, family and gang violence, substance abuse, and poor nutrition—circumstances that apply to the vast majority of youth in foster care. It establishes 'Coordinated Services Projects', aimed at improving access to social, health, and education services necessary for students to succeed in school.

Title XI dictates that during the 'Project Development Plan' for a 'Coordinated Services Project', an assessment must be performed that details the economic, social, and health barriers to educational achievement experienced by youth—including foster youth. Furthermore, a separate assessment of the needs to overcome these barriers of foster youth must be included. When detailing services for families at large to help youth succeed, the bill always states "including foster children and their foster families." Finally, when describing a book distribution initiative to improve literacy, foster youth are outlined as a target population in addition to ESL and low-income students.

The ESEA was once again reauthorized by the No Child Left Behind Act (NCLB) in 2001. However, NCLB did not include any additional measures that specified youth in foster care as a beneficiary. In fact, NCLB does not require reporting of academic achievement for youth in foster care, despite requiring reporting by gender, race, disability, migrant status, English proficiency, and economic status.

NCLB's proposed interventions for improving educational quality are less effective for foster youth. Firstly, NCLB allows students who are in schools that have not made 'adequate yearly progress' for 2 years in a row to transfer to a school that has. However, for foster youth, it is unclear who would both receive notice or make a decision. Furthermore, these notices (in addition to others) are sent only once a year, which can create issues since foster youth are highly mobile. A notice could easily go to a previous address, and never reach the youth and their current foster family.

When NCLB expired in 2007, a few laws were passed before the reauthorization of ESEA in 2015 that targeted educational interventions for youth in foster care. Firstly, the Fostering Connections to Success and Increasing Adoptions Act (FCSIA) in 2008 codified educational stability as a consideration for placement into a foster home. It also dictated that if remaining in a school was not in a child's best interest, then the child would be moved to a different school and be 'immediate[ly] and appropriate[ly]' enrolled. Secondly, the Uninterrupted Scholars Act (USA) in 2013 created a new exception under FERPA to allow schools to release educational records to child welfare agencies without parental consent, so that students' educational records could be transferred and potentially prevent educational disruption.

In 2015, the Every Student Succeeds Act (ESSA) was passed and remains in effect. It is the sixth reauthorization of ESEA. ESSA dictated that youth entering foster care would remain at their school prior to entering if it was determined to be in their best interest—expanding on FCSIA by not only taking the child's previous school into account when deciding a foster care placement, but mandating that the child remain in their current school if its in their best interest. Building off USA's theme of continuity regardless of record attainment difficulties, then the child must be enrolled in their new school immediately regardless of if record requirements are met if remaining in the school of origin was determined to not be in the child's best interest. The burden of record transfer is placed upon the school of origin and the new one, not the foster family. Furthermore, transportation must be supplied regardless of additional incurred cost—and the child welfare agency, educational agency, or both must pay it. ESSA also explicitly codified the necessity for states to collect four-year adjusted cohort graduation rates.

**National Youth in Transition Database**

The ESSA requires states to collect four-year adjusted cohort graduation rates, but does not specify data collection around the circumstances youth in foster care individually face. More robustly, the Foster Care Independence Act of 1999 established what is now known as the Chafee Foster Care Program for Successful Transition to Adulthood (Chafee Program). The law requires the Administration for Children and Families (ACF) to develop a detailed data collection system. This established the National Youth in Transition Database (NYTD). It is required for states to gather and report data on 11 broad categories: budget and financial management; housing education and home management training; health education and risk prevention; family support and healthy marriage education; mentoring; supervising independent living, and financial assistance (education, room and board). In order to collect this information, the NYTD is a collection of survey responses aimed at gathering information on these categories. Some of these questions include:[1]

- "Have you ever referred yourself or has someone else referred you for an alcohol or drug abuse assessment or counseling?"
- "Currently are you receiving social security payments (Supplemental Security Income (SSI), Social Security Disability Insurance (SSDI), or dependents' payments)?"
- "Currently is there at least one adult in your life, other than your caseworker, to whom you can go for advice or emotional support?"
- "What is the highest educational degree or certification that you have received?"

The NYTD survey is conducted on "cohorts" of youth at three distinct points: the Age 17 Baseline Survey, Age 19 Follow-up Survey, and Age 21 Follow-up Survey. Each cohort is

[1] A full list of questions can be found in Appendix A.

identified by the baseline year in which the survey is administered, and a new cohort is created every three years. For instance, a baseline year of 2011 would correspond to Cohort 2011.

States must contact all foster care youth who turn 17 in foster care or who enter foster care within 45 days of their 17th birthday in a baseline year to administer the survey. All youth who fit this criteria are in the baseline population for that respective cohort. There are three surveys in total: the Age 17 Baseline Survey, Age 19 Follow-up Survey, and Age 21 Follow-up Survey. After the age 17 Baseline Survey, states may choose to conduct sampling and offer the survey only to foster youth who are probabilistically sampled. The Age 19 and Age 21 Follow-up Surveys are offered to all youth who responded to at least one survey question in the age 17 Baseline Survey—and only to those who are sampled in sampling states.

The 'Cohort' population is defined as youth who responded to at least one survey question and are in the sample for states that chose to make one. In total, there are three completed cohorts: Cohort 2011, Cohort 2014, and Cohort 2017. Although those in the cohort must have responded to at least one question in the age 17 Baseline Survey, their demographic information is still reported in the NYTD. Therefore, it is possible to evaluate if the cohort is representative of the baseline population by sex, race, and state.

<u>Representativeness of Cohorts</u>

**Figure 1: Cohort vs. Baseline Representation of Males**

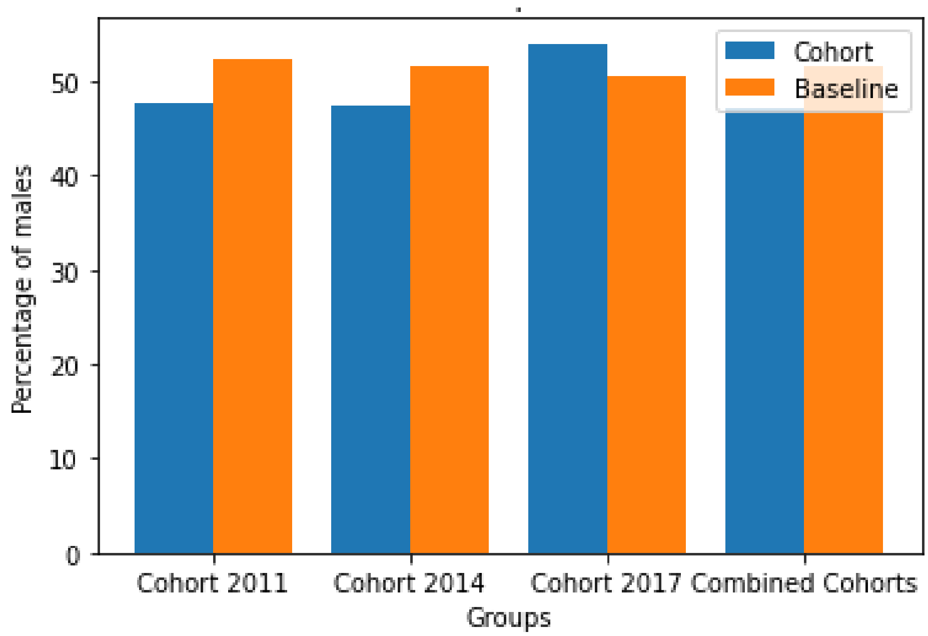



**Figure 2: Combined Cohorts vs. Baseline Representation of Races**

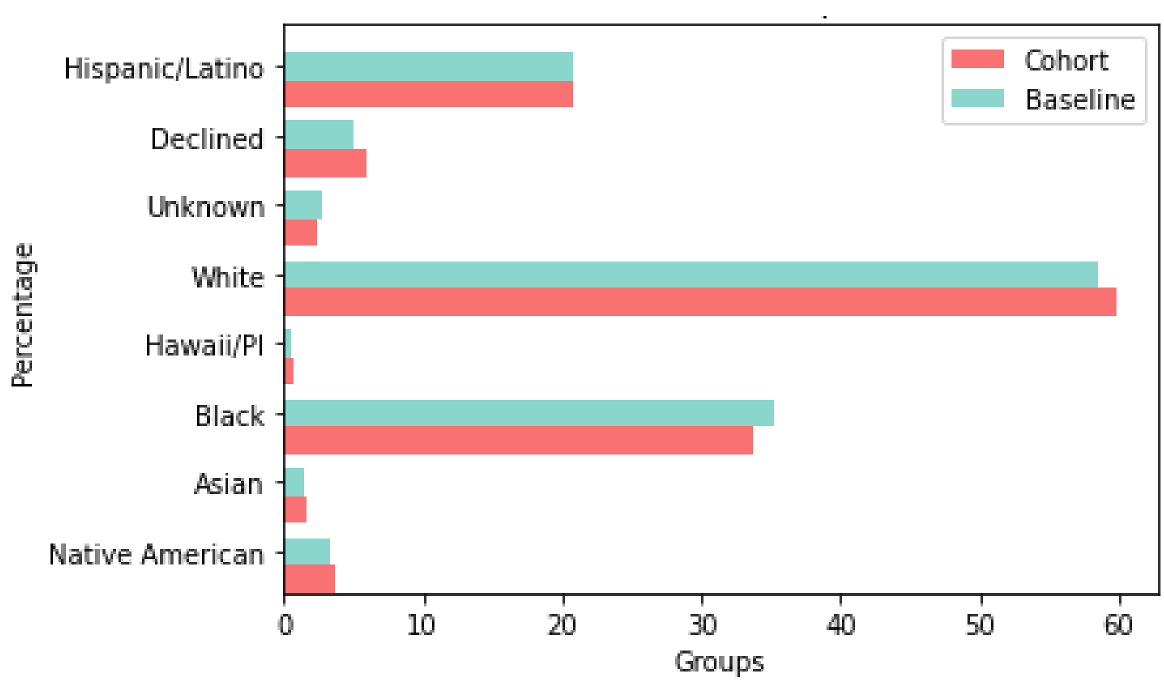


**Figure 3: Combined Cohorts Percentage of Baseline Population Participation at Age 19**

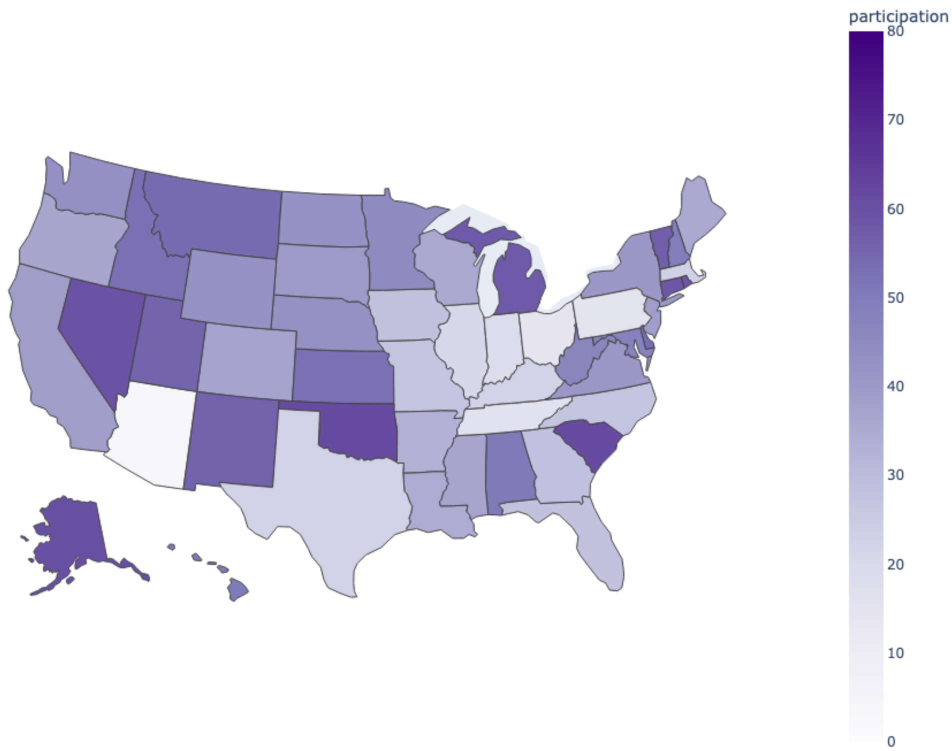



Figure 1 displays the gender representation in the Cohorts. Compared to the baseline populations to the baseline population, there are more females in each Cohort. However, this difference is not large. Overall, the Cohort is representative of the baseline population by gender. Similarly, Figure 2 shows the representation by race in the Cohorts. There are some observed differences in racial composition of the cohort compared to the baseline population. But, these differences are relatively small, and the cohort is fairly representative of the baseline population by race.

However, Figure 3 shows a different picture of representation. The data is not representative by state, as some states have higher rates of participation than others. Since states are allowed to choose the method they administer the survey, it is expected that there are drastic differences in representation at the state level.

The overall response rate for the three cohorts were varied: for the first survey, the response rates were 66.5% for Cohort 2017, 69% for Cohort 2014, and 54% for Cohort 2011 from the baseline population. Only those who responded to the first survey were invited to participate in the Age 19 and Age 21 surveys. Therefore, low response rates for the first wave translate to even lower response rates in the second wave. Overall, Cohorts 2017, 2014, and 2011 had response rates 36%, 37%, and 26% respectively from the baseline population during the age 19 survey.

Since participation in the survey is entirely voluntary, this sample of youth may not entirely reflect the Baseline population. While the dataset may overly represent high school graduation rates, since those who have graduated high school have better outcomes and therefore better means to respond to follow-up surveys, the characteristics of those who do graduate high school are not necessarily overrepresented within the graduate population. This dataset may overrepresent the baseline population's characteristics more positively, but there isn't a reason to believe that the baseline population's characteristics between the youth who graduated high school & those who did not are overrepresented. Therefore, this dataset can be used in a meaningful way to understand factors and predict high school graduation.

Information Not Captured

There are two broad categories of information that are not captured in the dataset that are traditionally used to evaluate high school graduation outcomes: academic and environmental. Academic information includes the youth's academic progress—which can include if they are on track to graduate high school, standardized test scores, and GPA. It can also encompass if the youth is receiving an Individualized Education Plan (IEP). Environmental factors are factors that can shape the youth's experience, but are not direct attributes of the youth themselves. This can include school quality, socioeconomic status of the foster home, and socioeconomic status of the neighborhood in which the school resides.

In all, the dataset includes 26 distinct variables that can be evaluated. Exploring those variables, this paper will investigate the factors that contribute to the attainment of a high school diploma or equivalent for youth in foster care by age 19 using the completed Cohorts. While the focus of this paper will be on high school completion, this dataset can also be used in the future to explore other questions.

**Methodology**

<u>Capturing High School Graduation Status</u>

The datasets do not include a variable for high school graduation, instead containing a categorical variable describing a youth's highest educational attainment at the time of survey, called "highedcert". A dummy variable was encoded ("Graduated") with values 0 or 1 if the youth indicated they had completed no educational milestone or if they had achieved a high school diploma / GED and beyond. The assumption was made that if a youth had achieved a degree that was beyond a high school diploma, they did indeed complete high school or equivalent. This assumption is founded in the fact that colleges, universities, and vocational certifications require high school or equivalent prior to matriculation at their institution. Some youth opted to leave this question blank ('blank') or declined to answer it ('declined'), and these answers were initially preserved in the "Graduated" variable: meaning the variable took on four values: (0, 1, 'blank', 'declined').

Another dummy variable, "Gradby19", was introduced to represent if a youth had reported graduating high school during the follow-up survey at age 19. This variable was created by taking the set of youth who reported at both age 17 and age 19: then, for each youth, the value that “Graduated” at age 19 took was mapped onto the “Gradby19” variable for each survey report for that youth. If the youth did not complete a survey in the age 19 survey, then ‘None’ was filled in. In creating this variable, it is possible to see what traits the youth had at age 17 that could contribute to their outcome at age 19.

Note that some states performed sampling. Other states reported youth, even if they did not complete the survey at all. The datasets were filtered to include only youth that actively participated in the survey and who were eligible to take the age 19 survey. The final datasets contained the survey responses of youth at age 17 who were eligible for and completed the age 19 survey. These final datasets were used to create summary statistics. Youth who responded ‘blank’ or ‘declined’ for their graduation status at age 19 only accounted for ~3% of all youth, so these observations were dropped when creating graduation visualizations. A few responses were erroneously input at ‘78.0’, which did not correspond to any responses on the survey. Other responses were null within the dataframe. These were also dropped. Errors were only present in Cohort 2011 and Cohort 2014.

Since the survey questions do not change between years, the three cohorts were concatenated together to create one, larger sample.

**Table 1: Number of Youth in Age 17 Survey that Participated in Age 19 Survey**

| Cohort | Including Blank/Declined/78.0/nan | Excluding Blank/Declined/78.0/nan | Percent Retained |
|---|---|---|---|
| Cohort 2011 | 7852 | 7392 | 94.14% |
| Cohort 2014 | 8906 | 8680 | 97.46% |
| Cohort 2017 | 8971 | 8787 | 97.95% |
| Combined Cohorts | 25729 | 24859 | 96.62% |

Youth who responded ‘blank’ or ‘declined’ regarding their highest educational attainment were highly correlated with responding ‘blank’ or ‘declined’ on other survey questions. When combining the three cohorts and dropping blank responses, declined responses, and erroneously entered information, the total sample size utilized is $n$ = 24859.

**An Overview of Graduation Rates**

Overall Graduation Rates

Graduation rates by age 19 were drastically worse for youth in all three Cohorts. Comparison is done with the Academic Year that the Cohort would have turned 18, as this is generally when youth in the United States graduate high school. For example, when examining Cohort 2011, the graduation rate is compared to the federal graduation rate in Academic Year 2011-2012 from the National Center for Education Statistics. Table 2 displays the results below.

**Table 2: Overall Graduation Rate by Age 19 by Cohort**

| **Cohort** | **Grad Rate by 19** | **Federal Grad Rate**[2] | **Difference** |
|---|---|---|---|
| Cohort 2011 | 59.28% | 80% | 20.72% |
| Cohort 2014 | 59.86% | 83% | 23.14% |
| Cohort 2017 | 62.66% | 85% | 22.34% |
| Combined Cohorts | 60.68% | N/A | N/A |

While the graduation rate did improve with each new cohort and grow at a faster rate than the federal graduation rate, the percentage point gap remained over 20%. Re-emphasizing that this is an optimistic estimate due to the skew from survey response, this should be noted as the overall graduation rate for the Combined Cohorts and could potentially not represent the true

[2] This measurement uses Adjusted-Cohort Graduation Rate (ACGR). It is defined as “number of students who graduate from high school in four years with a regular high school diploma, divided by the number of students who form the adjusted cohort for the graduating class.”

population graduation rate. The percentage point gap is also an optimistic estimate, as the graduation rate by age 19—which could be the 5 or 6 year graduation rate—is being compared to the 4 year federal high school graduation rate that, unlike the definition used in this paper, excludes GED attainment. Since the government does not collect and publish the 5 and 6 year graduation rates, this is the most feasible comparison possible. The true percentage point gap is likely higher.

Evaluation of Graduation Rate Representativeness

While it is not possible to evaluate the representativeness of the Combined Cohorts' graduation rates, it is possible to evaluate the representativeness in two different avenues. One such way is by evaluating the reported graduation rate by the Cohorts with reported data at the state level to the federal government. This kind of comparison can be done with Cohort 2017. It is the first Cohort in which the ESSA's data collection mandate was in effect, meaning that states were required to report the state-wide graduation rate for youth in foster care. Despite this, 9 states and Puerto Rico did not report their data. Of those, 7 states are labeled as 'Not Available', 2 states are labeled as 'suppressed due to concerns with data quality', and Puerto Rico is labeled as 'suppressed to protect the confidentiality of individual student data.' Despite this, this can be contrasted with the Cohort Graduation Rate by age 19 by state to get a sense of the graduation rate representativeness for the baseline population.

**Table 3: Graduation Rate by Age 19 for Cohort 2017 vs. Area Reported Graduation Rate**

| State/ Area | Grad Rate for Cohort 2017 | Reported State Grad Rate | State/ Area | Grad Rate for Cohort 2017 | Reported State Grad Rate | State/ Area | Grad Rate for Cohort 2017 | Reported State Grad Rate |
|---|---|---|---|---|---|---|---|---|
| AK** | 60.53% | 54% | LA* | 36.58% | 54% | OK | 57.50% | 58% |
| AL* | 50.47% | 67% | MA | 53.77% | 58% | OR | 70.39% | N.R. |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| AR | 61.76% | 65% | MD** | 64.34% | 50% | PA | 55.06% | 56% |
| AZ** | 57.14% | 45% | ME | 55.56% | 53% | PR | 88.50% | N.R. |
| CA** | 72.36% | 58.2% | MI** | 52.99% | 40% | RI* | 50.62% | 57% |
| CO** | 62.62% | 31% | MN | 63.89% | N.R. | SC** | 70.59% | 44% |
| CT** | 68.35% | 47% | MO* | 64.71% | 69% | SD** | 58.54% | 43% |
| DC | 52.63% | 53% | MS** | 77.12% | 65% | TN | 61.43% | 60% |
| DE* | 43.24% | 74% | MT | 73.33% | 71% | TX | 50.98% | N.R. |
| FL | 60.70% | 57% | NC | 57.67% | 57% | UT | 72.73% | N.R. |
| GA | 39.29% | N.R. | ND | 75.00% | 73% | VA** | 68.15% | 54% |
| HI** | 86.36% | 69% | NE** | 60.80% | 55% | VT | 64.71% | N.R. |
| IA | 67.74% | 64% | NH** | 78.26% | 43% | WA** | 56.86% | 50% |
| ID** | 55.36% | 40% | NJ** | 65.62% | 55% | WI** | 69.48% | 60% |
| IL | 63.70% | N.R. | NM** | 59.38% | 39% | WV | 68.18% | N.R. |
| IN | 64.81% | 67% | NV** | 70.59% | 50% | WY | 72.92% | N.R. |
| KS** | 70.37% | 62% | NY* | 43.58% | 57% | | | |
| KY | 69.09% | N.R. | OH | 52.56% | 57% | | | |

* : Cohort graduation rate is lower than state-reported graduation rate by 5%+
**: Cohort graduation rate is higher than state-reported graduation rate by 5%+
N.R.: Not reported

Out of the 41 states and District of Columbia that had reported state-wide graduation rates, 21 states had over-represented graduation rates in Cohort 2017 compared to the reported state-wide graduation rates. 6 states had under-represented graduation rates. This means Cohort 2017 only represented the state-wide graduation rate within 5% in 14 out of 42 reporting areas. This is unsurprising given the expectation that graduation rates would be over-represented.

However, there are still potential explanations for why this “over-representativeness” is present. Large percentages of youth could be graduating at the 5 and 6 year marks instead of 4 years. Youth in care are more likely to be held back than other students, and sometimes their credits do not fully transfer and align with their new school when transferring. Furthermore, youth in foster care are more likely to attain a GED than their peers: in Colorado, 13.8% of Colorado students in foster care earned a GED within 5 years compared to 2.7% for their peers (Colorado Department of Education, 2015). In spite of this, 6 states had Cohort 2017 reporting graduation rates by age 19 below the 4 year ACGR. Further exploration of survey collection methods in these 6 states is necessary to understand this phenomenon.

Another avenue of comparison is to longitudinal studies conducted in the past with similar criteria to be eligible and have a representative sample of youth. Two of these studies that have been conducted that report the graduation rates of youth in foster care are the Midwest Evaluation and the CALYOUTH study. These studies are similar in that the youth were surveyed at multiple points throughout their lives (at points age 17, age 19, and age 21) and eligibility requirements were slightly different—the Midwest Evaluation required youth to be in foster care during their 16th birthday and remain in care until their 17th birthday, while the CalYOUTH study required youth to be between 16.75 and 17.75 years of age and in foster care for at least 6 months. These studies ensured their cohorts were representative of youth in foster care through methods such as stratified random sampling.

From these studies, comparison is possible for four states: California, Illinois, Wisconsin, and Iowa. In California, the reported graduation rate at age 19 from the CalYOUTH study is 75.7% (Courtney et al., 2016). The best cohort to compare this to is Cohort 2011, as the study was initiated in 2012: the reported graduation rate is 74.46% in California. This signals that, at least in California for Cohort 2011, the sample is most likely representative of graduation rate.

The Midwest Evaluation resulted in the publication of 3 reports detailing the circumstances of youth at age 19 in each state, reporting the graduation rate at age 19 for each. Since these interviews were conducted in 2004, the closest Cohort to compare these to out of those available is Cohort 2011. This is because the graduation rate has been climbing historically year-to-year, so choosing the earliest year available yields the fairest comparison. In Illinois, the graduation rate at age 19 was 62.2%, compared to 65.28% for Cohort 2011 (Courtney & Dworsky, 2006a). In Wisconsin, the graduation rate at age 19 was 64.6%, compared to 69.09% for Cohort 2011 (Courtney & Dworsky, 2006c). In Iowa, the graduation rate at age 19 was 73.50%, compared to 73.44% for Cohort 2011 (Courtney & Dworsky, 2006b). Comparing and accounting for graduation rates steadily growing over time, I argue that the Cohorts are most likely representative of the graduation rate, at least for Illinois, Iowa, and Wisconsin for Cohort 2011. Broadly, I will assume the Cohorts are representative of the graduation rate by age 19.

<u>Graduation Rate by State</u>

**Figure 4: Combined Cohorts Graduation Rate by Age 19**

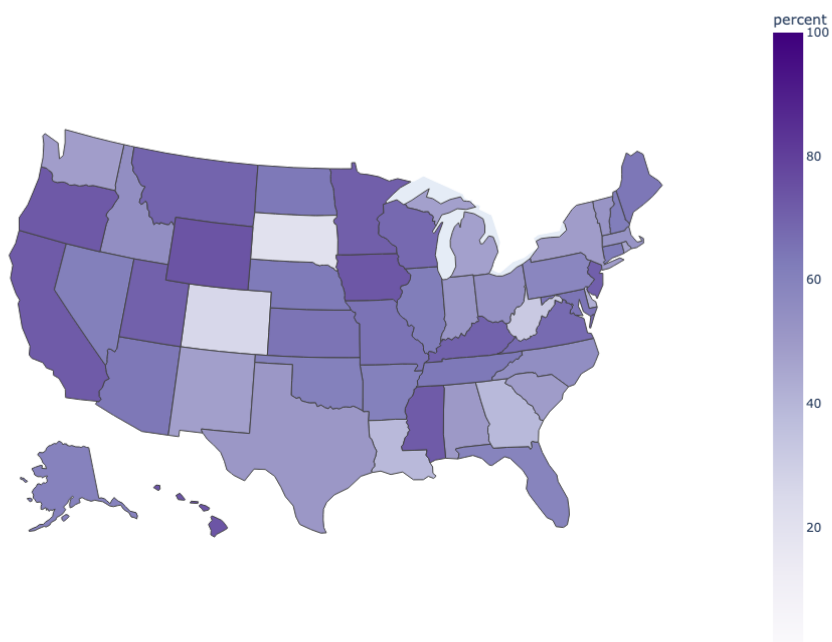


Figure 4 depicts the graduation rate by age 19 for the Combined Cohorts. The chi-square test statistic for state is 1486.16 with a p-value < .000001, meaning that there is a significant association between state and the graduation rate. This makes sense, but could be due to either state-variation in foster care systems or due to survey administration techniques utilized by states favoring one population over another. It is key to note when looking at the state-level graduation rate that the erroneous data entry for highest educational certification was concentrated in 3 states: South Dakota, Colorado, and West Virginia—and these states reported less than 10% for graduation rate in Cohort 2011 and 2014.

Graduation Rate by Gender

**Figure 5: Graduation Rates by Age 19 for Each Cohort by Gender**

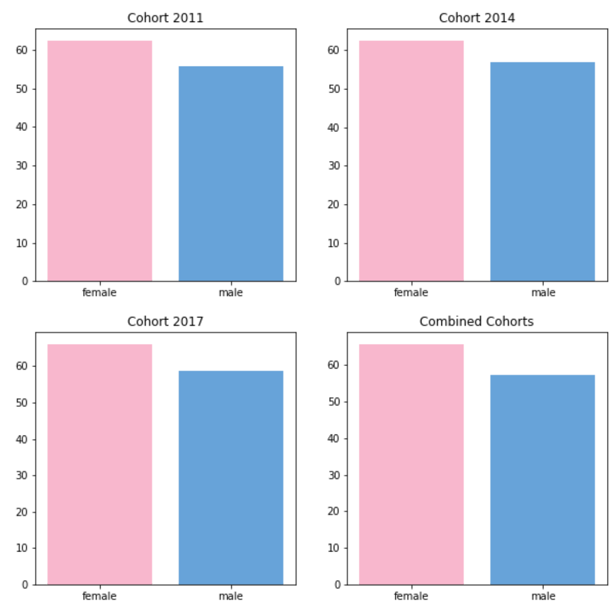



Figure 5 displays the graduation rate by age 19 by gender for each Cohort. Females in the Cohorts graduate at higher rates than males, with a chi-squared test statistic of 110.84 and p-value < .000001 for the Combined Cohorts. This is a trend that is not unique to youth in foster care: in data collected across 37 states' Department of Education examining ACGR between males and females for Academic Year 2017-2018, females graduated at higher rates than males in every state (Reeves et al., 2022). Further analysis is necessary to deduce if foster care status exasperates the gender graduation gap.

<u>Graduation Rate by Race</u>

**Figure 6: Graduation Rates by Age 19 for Each Cohort by Race**

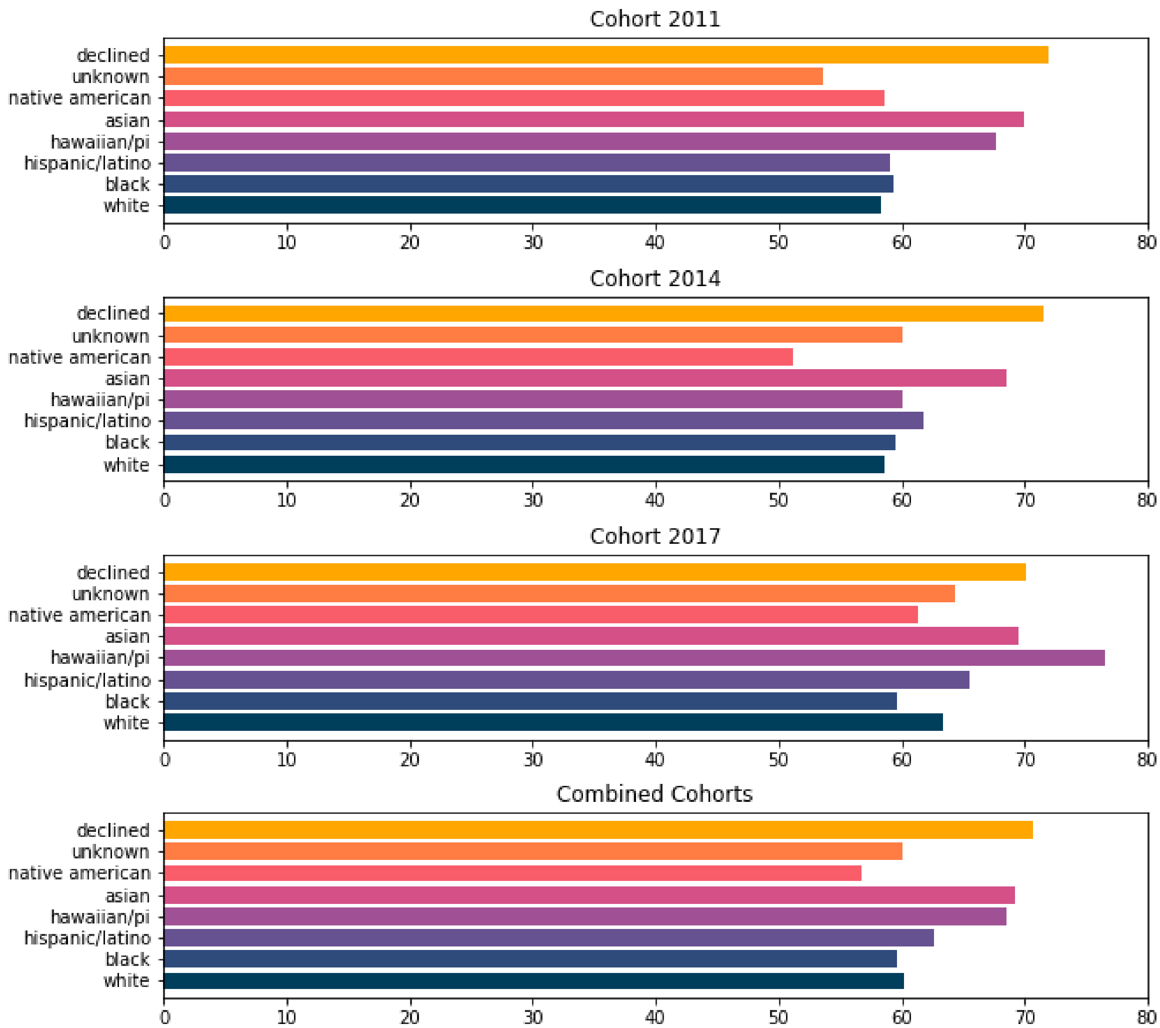



Figure 6 displays the graduation rate by age 19 for each Cohort. Graduation rates vary across racial groups, and the distributions of rates also vary across Cohorts. Unlike gender, there is not a consistent group that graduates at the highest rate when evaluating all 3 Cohorts and the

Combined Cohorts. Interestingly, in Cohort 2011, those who declined to state their race graduated at the highest rates. This trend persisted across all cohorts except Cohort 2017, where they ranked second. It is hard to deduce who could have opted to decline to state their race—one theory is that youth who did not see their racial identity specifically listed as an option opted to decline instead. The chi-square statistics are all p-value < .000001 for the Combined Cohorts.

Similar to gender, it is difficult to deduce whether the racial gap is due to foster care or due to pre-existing inequalities in the United States. For example, the 4 year ACGR for school year 2017-2018 is 89% for white students, 79% for black students, 81% for hispanic students, 73% for Native American students, and 92% for Asian students, Hawaiian Native students, and Pacific Islander students combined (National Center for Education Statistics). Despite youth in foster care having lower graduation rates across the board, the gaps in percentage points between the different racial groups are proportionally larger. Further analysis is necessary to deduce if foster care status exasperates the racial group graduation gap.

Graduation Rate by Life History

Life history information describes past experiences that a youth has had. These factors include incarceration, homelessness, substance abuse counseling referrals, having a child, and if they were married when they had a child (if applicable). The questions in the survey capture the youth's lifetime experience at age 17.

**Table 4: Graduation Rate by Age 19 of Life History Characteristics with Chi-square Test**

| **Characteristic** | **Graduation Rate** | **$\chi 2$ Test Statistic** | **P-Value** |
|---|---|---|---|
| **History of homelessness?** | | | |
| Yes | 58.54% | 14.3978 | .0061 |
| No | 61.27% | | |

| | | | |
|---|---|---|---|
| Declined | 58.00% | | |
| Blank | 56.41% | | |
| **History of substance abuse counseling referral?** | | | |
| Yes | 55.39% | 99.1130 | < .0000001 |
| No | 62.52% | | |
| Declined | 56.92% | | |
| Blank | 52.94% | | |
| **History of incarceration?** | | | |
| Yes | 53.75% | 207.4771 | < .0000001 |
| No | 63.64% | | |
| Declined | 56.75% | | |
| Blank | 52.50% | | |
| **Has children?** | | | |
| Yes | 49.79% | 82.0789 | < .0000001 |
| No | 61.49% | | |
| Declined | 56.77% | | |
| Blank | 48.45% | | |
| **Is married, if had a child?** | | | |
| Yes | 66.67% | 73.9010 | < .0000001 |
| No | 50.59% | | |
| Declined | 71.58% | | |
| Blank | 51.47% | | |
| Not Applicable | 61.23% | | |

The null hypothesis for each test is that there is no significant association or independence between the characteristic and high school graduation. When evaluating all characteristics at the $\alpha = .05$ level, all null hypotheses are rejected. There is evidence of a statistically significant relationship between history of homelessness, history of substance abuse counseling, history of incarceration, having a child, and being married if one has a child with high school graduation.

<u>Graduation Rate by Resources</u>

Resources information details the resources the youth is currently a recipient of and is the majority of the information from the survey. This includes connection to an adult, Social Security, educational assistance, public financial assistance, public food assistance, public housing assistance, other financial assistance, Medicaid, other health insurance, medical insurance (if other health insurance), mental health insurance (if other health insurance), and prescription health insurance (if other health insurance). Some of these resources have conditions for the youth to be able to receive them, and could be a proxy for other information. For example, receiving educational assistance if you are enrolled in high school could be a proxy for educational quality, since non-selective public high schools are free. Another example is that receiving Social Security could be a proxy for personal or familial disability, which without proper accommodation can affect your ability to pursue a high school diploma.

**Table 5: Graduation Rate by Age 19 of Resource Characteristics with Chi-squared Test**

| **Characteristic** | **Graduation Rate** | **$\chi 2$ Test Statistic** | **P-Value** |
|---|---|---|---|
| **Connection to an adult?** | | | |
| Yes | 60.88% | 8.9828 | .0295 |
| No | 57.16% | | |

| | | | |
|---|---|---|---|
| Declined | 61.99% | | |
| Blank | 52.38% | | |
| **Receives Social Security?** | | | |
| Yes | 54.35% | 56.8820 | < .0000001 |
| No | 61.59% | | |
| Declined | 59.79% | | |
| Blank | 53.57% | | |
| **Receives Educational Aid?** | | | |
| Yes | 73.14% | 53.1221 | < .0000001 |
| No | 60.32% | | |
| Declined | 60.20% | | |
| Blank | 57.58% | | |
| **Public financial Assistance?** | | | |
| Yes | 61.51% | 53.4222 | < .0000001 |
| No | 59.65% | | |
| Declined | 54.55% | | |
| Blank | 54.09% | | |
| Not Applicable | 61.78% | | |
| **Public food assistance?** | | | |
| Yes | 72.08% | 75.8460 | < .0000001 |
| No | 58.51% | | |
| Declined | 55.77% | | |
| Blank | 54.98% | | |

| | | | |
|---|---|---|---|
| Not Applicable | 61.78% | | |
| **Public housing assistance?** | | | |
| Yes | 58.06% | 46.2268 | < .0000001 |
| No | 59.88% | | |
| Declined | 58.06% | | |
| Blank | 54.12% | | |
| Not Applicable | 61.71% | | |
| **Other financial assistance?** | | | |
| Yes | 61.64% | 16.8934 | .0007 |
| No | 60.57% | | |
| Declined | 62.47% | | |
| Blank | 46.88% | | |
| **Is on Medicaid?** | | | |
| Yes | 60.67% | 11.6050 | .0205 |
| No | 60.09% | | |
| Declined | 60.59% | | |
| Blank | 50.50% | | |
| Do not know | 62.71% | | |
| **Has other health insurance, if not in foster care?** | | | |
| Yes | 63.83% | 33.2680 | < .0000001 |
| No | 60.03% | | |
| Declined | 62.50% | | |
| Blank | 50.44% | | |

| | | | |
|---|---|---|---|
| Do not know | 61.18% | | |
| Not applicable | 55.56% | | |
| **Has medical insurance, if has other health insurance?** | | | |
| Yes | 63.37% | 19.0843 | .0019 |
| No | 66.90% | | |
| Declined | 55.00% | | |
| Blank | 55.38% | | |
| Do not know | 61.35% | | |
| Not applicable | 60.17% | | |
| **Has mental health insurance, if has other health insurance?** | | | |
| Yes | 63.37% | 21.3019 | .0007 |
| No | 71.58% | | |
| Declined | 70.27% | | |
| Blank | 54.10% | | |
| Do not know | 64.80% | | |
| Not applicable | 60.19% | | |
| **Has prescription insurance, if has other health insurance?** | | | |
| Yes | 63.58% | 19.0105 | .0019 |
| No | 64.29% | | |
| Declined | 67.86% | | |

| | | | |
|---|---|---|---|
| Blank | 54.10% | | |
| Do not know | 66.55% | | |
| Not applicable | 60.19% | | |

The null hypothesis for each test is that there is no significant association or independence between the characteristic and high school graduation. When evaluating all characteristics at the $\alpha = .05$ level, all null hypotheses are rejected. There is evidence of a statistically significant relationship between connection to an adult, Social Security, educational assistance, public financial assistance, public food assistance, public housing assistance, other financial assistance, Medicaid, other health insurance, medical insurance (if other health insurance), mental health insurance (if other health insurance), and prescription health insurance (if other health insurance) with high school graduation.

Graduation Rate by Human Capital Development

Human capital development information includes what the youth is actively pursuing, or has gained from the active pursuit of something. Full-time employment, part-time employment, employed either part-time or full-time, employment skills, and current enrollment are all included in this category.

**Table 6: Graduation Rate by Age 19 of Human Capital Development Characteristics & Chi-Square Test**

| Characteristic | Graduation Rate | χ2 Test Statistic | P-Value |
|---|---|---|---|
| **Current full-time employment?** | | | |
| Yes | 65.50% | 6.6254 | .0848 |
| No | 60.57% | | |

| | | | |
|---|---|---|---|
| Declined | 65.52% | | |
| Blank | 52.38% | | |
| **Current part-time employment?** | | | |
| Yes | 71.90% | 226.2841 | < .0000001 |
| No | 58.72% | | |
| Declined | 62.06% | | |
| Blank | 53.57% | | |
| **Employment (full or part time)?** | | | |
| Yes | 71.22% | 221.4767 | < .0000001 |
| No | 58.63% | | |
| Did not answer both | 61.72% | | |
| **Has employment skills?** | | | |
| Yes | 66.87% | 114.0639 | < .0000001 |
| No | 58.98% | | |
| Declined | 59.61% | | |
| Blank | 55.26% | | |
| **Currently enrolled?** | | | |
| Yes | 60.99% | 21.6250 | < .0000001 |
| No | 54.88% | | |
| Declined | 63.60% | | |
| Blank | 52.38% | | |

The null hypothesis for each test is that there is no significant association or independence between the characteristic and high school graduation. When evaluating all characteristics at the $\alpha = .05$ level, all null hypotheses are rejected except one. Full-time employment was not significantly different between those who graduated and those who did not in the chi-squared test at the $\alpha = .05$ level, and was uniquely the only variable that was not significantly different. There is evidence of a statistically significant relationship between part-time employment, employed either part-time or full-time, employment skills, and current enrollment with high school graduation.

**Modeling**

In order to understand the significance of predictors beyond correlation, I plan on using regression and machine learning. The goal of regression analysis is to understand the relationship between circumstances and resources in the youth's life and high school completion, and whether they have statistical significance. The goal of machine learning is to produce a model that can predict, given the characteristics of a youth at age 17, whether they will graduate high school by age 19. While logistic regression can be argued to be a form of machine learning as it predicts a binary outcome, it is still a type of regression analysis that is specifically used for binary outcomes, as it provides coefficients and corresponding p-values that help us understand the particular relationship and magnitude between each predictor and high school graduation.

<u>Machine Learning's Ethicality in Child Welfare</u>

The application of machine learning in the child welfare sector raises ethical concerns. Previous and current deployments of machine learning include using multi-class classification algorithms to characterize youth in a gradient between 'low' to 'high' risk, which could determine their investigation and subsequent removal from the home. However, these models are built on decades of data in which there was—and still is—socioeconomic and racial discrimination when making a decision about what family is reported and investigated, as well as if an investigation escalates to removal from the home.

Some states have attempted to deploy machine learning and other artificial intelligence tools with the intention of improving efficiency of their child welfare systems. However, these systems have received critique due to their inherent bias. In 2017, Illinois ended a program that used machine learning to predict children at risk of serious injury or death after it wildly misclassified youth (Government Technology). In June 2022, Oregon announced it would be concluding its use of predictive models to determine which families are investigated following reports of black families being flagged at disproportionately high rates (Associated Press).

This discussion is particularly timely, as the Associated Press published an article in March 2023 describing a potentially discriminatory case of artificial intelligence being deployed to select a child's removal from the home. Two parents described as having developmental disabilities took their daughter to the hospital due to her refusal to eat, and after being discharged the Pittsburgh Department of Children and Family Services showed up to remove the child from the home on the basis of neglect. While the particular family questioned whether they were flagged due to having developmental disabilities and if that created a high enough 'risk score' for removal, the county refused to disclose what attributes the algorithm included to make an assertion. The U.S. Department of Justice is now investigating the county's welfare system to determine if the county's algorithm is discriminating against families with disabilities.

However, I argue that machine learning can be ethically used to predict high school graduation because the algorithm's goal in this problem is resource allocation and efficiency—not making high-stakes decisions that are disruptive to homes. It is about adding additional resources if necessary instead of taking away resources. A framework created jointly by the The Alan Turing Institute and the Rees Centre at University of Oxford guides this approach to ensure this research is in line with the ethical standards of both machine learning and social work (2020).

Machine Learning Algorithms to Investigate

In order to fully investigate the predictive power of the National Youth in Transition Database's information on high school graduation rates, 4 different machine learning algorithms will be employed: Logistic Regression, Artificial Neural Network, XGBoost, and Random Forest. These are implemented with scikit-learn's functions for each respective model and evaluated, when appropriate, with statsmodel functions.

Feature Transformation and Exclusion

In machine learning, features refer to the input variables used to train a model. During the training phase, the model learns the relationship between the inputted features and the target variable. Using this learned relationship, it is able to classify new data.

In order to utilize the variables in the dataset as features, they must be converted to binary dummy variables with values 0 or 1. This is achieved using one-hot encoding. For example, the characteristic 'History of incarceration' has possible responses 'Yes', 'No', 'Declined', and 'Blank'. One-hot encoding creates columns 'History of incarceration Yes', 'History of incarceration No', 'History of Incarceration Declined', and 'History of Incarceration Blank'. One of the four columns is dropped in creation in order to prevent the 'dummy variable trap', which leads to multicollinearity. For example, in dropping 'History of Incarceration Blank', its value can be assumed to be '1' if the others are '0'. Furthermore, in order to prevent erroneous data entry from creating noise or confusion for the model training, observations that are from

Colorado, South Dakota, West Virginia, or contain ‘78.0’ in any response are dropped. This leaves $n$ = 24013.

**Table 7: Example of One-Hot Encoded Columns**

| History in Incarceration | History of Incarceration Yes | History of Incarceration No | History of Incarceration Declined |
|---|---|---|---|
| Yes | 1 | 0 | 0 |
| No | 0 | 1 | 0 |
| Declined | 0 | 0 | 1 |
| Blank | 0 | 0 | 0 |

Feature selection

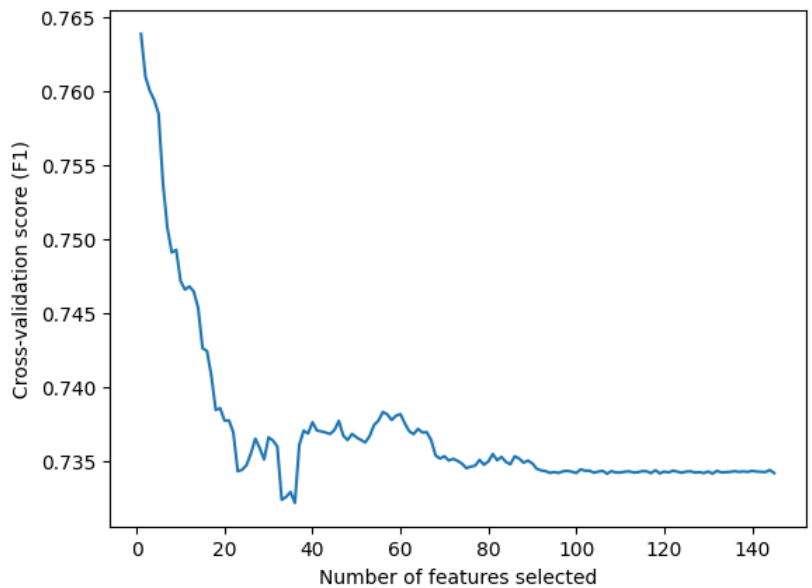
In order to select features for the models, a few methods can be employed: using the variables that are significant with the chi-squared test, using variables that are assessed as important from domain knowledge, and using features selected from recursive feature elimination. This means for each of the 4 machine learning algorithms, 4 models with 3 different sets of features will be made. In total, 12 models will be explored and the best model will be chosen from each to explore further.

Using the chi-square test significant variables, all features, one-hot encoded, will be included except full-time employment. This yields models with the highest dimensionality since they will contain the most features compared to the other feature selection methods. While this could run the risk of complicating the models and leading to poorer accuracy, it is possible it will yield the highest accuracy and is worthwhile to investigate. These features are going to be referred to as the ‘Chi-square’ set for short, and models built with them will be referred to as ‘Chi-Square (Type of Algorithm).’

Selection from domain knowledge can be described as falling into three categories: selected as controls, selected from the literature review, and selected from summary statistics with discretion. Sex, state, and race were selected as controls since these are immutable characteristics of the youth and capture disparities that exist. From the literature review, history of incarceration, history of substance abuse, history of homelessness, having a connection to an adult, and having children were selected. These were described as hindrances in educational attainment. Characteristics selected from reviewing summary statistics and reflecting on underlying factors include part-time employment, food assistance, educational aid, and social security. Notably, these characteristics have not been explored in the context of high school completion for youth in foster care and are novel analysis from this paper. These features are going to be referred to as the ‘Manual’ set for short, and models built with them will be referred to as ‘Manual (Type of Algorithm)’.

Recursive feature elimination is a technique that selects a subset of attributes from a larger set of features. It uses a model to repeatedly train on different subsets of features. The coefficients learned by the model are analyzed, and the least weighted attributes are removed and replaced with new ones until all attributes have been tried.  Recursive feature elimination was performed with logistic regression with an L1 penalty and optimized with F1 score. This is to ensure every feature has a non-zero coefficient.

**Figure 7: Recursive Feature Elimination with Cross Validation Using Logistic Regression**




Typically, the number of features with the best cross-validation score would be selected. However, this would lead to a poor model in practice with the sole feature being if an individual resides in Puerto Rico. Instead, the peak F1-Score at 56 features is used and a Logistic Regression model to view the features is used. Afterwards, features with high correlations are removed to avoid multicollinearity. This leaves 51 features in the RFE subset, found in Appendix B. These features are going to be referred to as the 'RFE' set for short, and models built with them will be referred to as 'RFE (Type of Algorithm).'

<u>Determination of Best Model</u>

In order to determine which model is the best for predicting high school graduation, it must be tested. In order to do this, a subset of the data is not used for the model to learn but instead is used to test the model's effectiveness—this is called 'train-test split'. This is to ensure the model is learning the underlying patterns and relationships, instead of memorizing the original dataset or 'overfitting' it. In this paper, 80% of the data is used as the training data and 20% of the data is used as the testing data.

Since the goal of this machine learning model is to predict the binary outcome of high school graduation, the following attributes will be used to determine the best model: Accuracy, Recall, Precision, and weighted F1-score. Accuracy describes the percentage of those correctly classified as graduating high school or not by age 19 when predicting on the testing data. Recall will have two values: one measures the proportion of actual high school graduates who were correctly classified, and one measures the proportion of actual non-graduates who were correctly classified by the model. Similarly, Precision will also have two values: one is the proportion of predicted high school graduates who actually graduated, and the other is the proportion of predicted non-graduates who actually did not graduate. The F1-score is the harmonic mean of Precision and Recall. It gives a balanced perspective of Precision and Recall. Overall, these metrics will allow us to determine which models are best at identifying graduates and non-graduates—as one model may perform better than another at identifying a particular class.

It is important to note that a trained model should surpass ~60% accuracy, as approximately 60% of youth reported graduating by age 19 in the Combined Cohort. This is because a model would be inferior to a person guessing that a youth will graduate 100% of the time, as this would have ~60% accuracy.

Hyperparameters

Unlike parameters, which are mutable while a model is learning, hyperparameters are set for a model prior to training and are immutable during learning. For example, the max depth of a decision tree is a hyperparameter. Hyperparameters can influence how long models will take to train, and can be strongly influential to the performance of a model. Since appropriate hyperparameters are difficult to determine without the use of a computer, they can be found experimentally by training multiple models with different combinations of hyperparameters. This is achieved with random search in scikit-learn, which can explore a larger space of possible hyperparameters with less computational cost.

Logistic Regression

Logistic regression is a common model for binary classification tasks. Three models were fit. The hyperparameter to fit, *C*, is the regularization term that prevents overfitting. It was searched in the space [$10^{-4}$ and $10^{4}$] using random search and eight points were chosen to test. Table 8 contains the results of the three fit models.

**Table 8: Logistic Regression Models and Results**

| **Features** | **Accuracy** | **Recall** | **Precision** | **F1-Score** |
|---|---|---|---|---|
| **Chi-square** | .6421 | | | |
| Graduates | – | .8440 | .6660 | .7445 |
| Non-graduates | – | .3155 | .5557 | .4025 |
| **RFE** | .6438 | | | |
| Graduates | – | .8430 | .6677 | .7452 |
| Non-graduates | – | .3215 | .5587 | .4082 |
| **Manual** | .6438 | | | |
| Graduates | – | .8501 | .6659 | .7468 |
| Non-graduates | – | .3101 | .5611 | .3994 |

The RFE model and Manual model have the same accuracy, but RFE is chosen as the better model because it performs better for predicting non-graduates. Through random search, it was fit with the best hyperparameter $C = 3.281332398719396$. Table 9 shows the coefficients, z-score, and p-values of the predictors in the model.

**Table 9: Logistic Regression Results for RFE Features**

| Feature | Coefficient | Z-Statistic | P-value |
|---|---|---|---|
| Constant | -0.2764 | -0.601 | .548 |
| Alabama | -0.5593 | -5.252 | .000 |
| California | 0.4682 | 10.421 | .000 |
| Delaware | -0.9776 | -4.778 | .000 |
| Georgia | -1.0923 | -10.063 | .000 |
| Hawaii | 0.0467 | 1.787 | .074 |
| Iowa | 0.5423 | 4.012 | .000 |
| Indiana | -0.3873 | -2.593 | .010 |
| Kentucky | 0.3528 | 2.788 | .005 |
| Louisiana | -0.8704 | -6.552 | .000 |
| Massachusetts | -0.5417 | -5.392 | .000 |
| Maine | 0.3064 | 1.007 | .314 |
| Michigan | -0.6777 | -8.284 | .000 |
| Minnesota | 0.3063 | 2.683 | .007 |
| New Jersey | 0.3519 | 2.632 | .009 |
| New Mexico | -0.4807 | -2.110 | .035 |
| New York | -0.7029 | -8.820 | .000 |
| Ohio | -0.2962 | -2.618 | .009 |
| Oregon | 0.3224 | 2.363 | .018 |
| Pennsylvania | -0.2536 | -2.247 | .025 |
| Puerto Rico | 1.6178 | 8.956 | .000 |
| Rhode Island | -0.7437 | -5.819 | .000 |
| South Carolina | -0.5932 | -5.404 | .000 |

| | | | |
|---|---|---|---|
| Texas | -0.4672 | -5.486 | .000 |
| Utah | 0.3633 | 3.085 | .002 |
| Vermont | -0.5157 | -1.839 | .066 |
| Washington | -0.5166 | -4.872 | .000 |
| Wyoming | 0.4911 | 1.969 | .049 |
| Sex (Male) | -0.2515 | -7.920 | .000 |
| Hispanic Origin (Declined) | -0.3632 | -1.086 | .278 |
| Current part-time employment (Declined) | 0.0754 | 0.470 | .530 |
| Current part-time employment (Yes) | 0.5802 | 12.188 | .000 |
| Employment skills (Yes) | 0.2907 | 7.341 | .000 |
| Social security (Yes) | -0.2545 | -5.324 | .000 |
| Educational aid (Declined) | 0.5323 | 1.112 | .266 |
| Educational aid (No) | 0.5077 | 1.102 | .270 |
| Educational aid (Yes) | 0.9849 | 2.093 | .036 |
| Public financial assistance (Declined) | -0.2594 | -1.120 | .263 |
| Public food assistance (Declined) | -0.4908 | -1.574 | .116 |
| Other financial assistance (Declined) | 0.4313 | 1.705 | .088 |
| Other financial assistance (No) | 0.3451 | 1.585 | .113 |
| Other financial assistance (Yes) | 0.3810 | 1.700 | .089 |

| | | | |
|---|---|---|---|
| Connection to adult (No) | -0.2815 | -4.098 | .000 |
| History of substance abuse counseling referral (Yes) | -0.2499 | -6.501 | .000 |
| History of incarceration (Yes) | -0.2488 | -5.728 | .000 |
| Has children (Declined) | 0.4464 | 2.501 | .012 |
| Has children (No) | 0.4541 | 2.964 | .003 |
| If has children, is married (No) | -0.0844 | -0.515 | .607 |
| Has other health insurance (Yes) | 0.0092 | 0.083 | .934 |
| If has other health insurance, has mental health insurance (Do not know) | -0.2777 | -1.300 | .194 |
| If has other health insurance, has mental health insurance (Yes) | -0.2753 | -1.452 | .147 |
| If has other health insurance, has prescription insurance (Not applicable) | -0.3408 | -1.754 | .079 |

The model's $R^2$ is .06. However, the log-likelihood ratio test is statistically significant ($p$-value < .00000001). The selected features are effective predictors of the outcome variable, but do not fully explain the variability in the graduation by age 19.

Some states were statistically significant in predicting the outcome variable in the model. Residence in Alabama, Delaware, Georgia, Indiana, Louisiana, Massachusetts, Michigan, New

Mexico, New York, Ohio, Pennsylvania, Rhode Island, South Carolina, or Washington was associated with a decrease in the log odds of graduation by Age 19. Conversely, residence in California, Iowa, Kentucky, Minnesota, New Jersey, Oregon, Puerto Rico, Utah, or Wyoming was associated with an increase in the log odds of graduation by age 19.

Unsurprisingly, gender was statistically significant. Being male was associated with a decrease in the log odds of graduation by age 19. Furthermore, having part-time employment, employment skills, receiving educational aid, declining to state if one has children, and having no children were all associated with an increase in the log odds of graduation by age 19. Having no connection to an adult, having a history of incarceration, having a history of substance abuse referral, and being on Social Security were all associated with a decrease in the log odds of graduation by age 19.

In the best performing model, no features identifying an individual's race were included—except whether one declined to state if they were of Hispanic origin. This sole feature was not statistically significant. This is surprising, considering that racial disparities and systemic barriers are well documented in the United States at large. Furthermore, features that pertained to health insurance coverage also were not statistically significant.

Overall, the model is missing features that would explain a lot of the variability that is not available in this dataset—especially since Recursive Feature Elimination was performed on the logistic regression model. The selected features are effective predictors of the outcome variable, but do not fully explain the variability in the graduation by age 19.

Artificial Neural Network

Artificial Neural Networks are a commonly used algorithm due to their ability to learn complex patterns or relationships in data, including non-linear functions. It is called an Artificial

Neural Network because its structure conceptually represents the human brain. There are four hyperparameters to fit: the dropout rate, L1 regularizer, learning rate, and number of neurons. The dropout rate dictates how many neurons in a layer are set to zero (‘dropped out’) so that the remaining neurons must learn more generalizable features to prevent overfitting. The L1 regularizer is another method to prevent overfitting by helping the model simplify. The learning rate determines how much the weights of the neural network are updated in each iteration and must be balanced between instability and slow convergence. The number of neurons affects the model’s ability to detect complex patterns, but can be an avenue for overfitting if too many are used. Table 10 displays the results of fitting the three feature sets.

**Table 10: Artificial Neural Network Models and Results**

| **Features** | **Accuracy** | **Recall** | **Precision** | **F1-Score** |
|---|---|---|---|---|
| **Chi-square** | .6398 | | | |
| Graduates | – | .8423 | .6654 | .7492 |
| Non-graduates | – | .3123 | .5504 | .3985 |
| **RFE** | .6396 | | | |
| Graduates | – | .8154 | .6717 | .7366 |
| Non-graduates | – | .3553 | .5433 | .4297 |
| **Manual** | .6442 | | | |
| Graduates | – | .8767 | .6596 | .7528 |
| Non-graduates | – | .2681 | .5734 | .3654 |

The manually selected features model performed the best, with accuracy of .6442. The hyperparameters selected for this model were a dropout rate of $4.209 \times 10^{-5}$, L1 regularizer of $5.229 \times 10^{-4}$, learning rate of $6.513 \times 10^{-4}$, and 23 neurons. In order to assess the impact of features on the model's performance, permutation importance is computed. Permutation importance is measured by randomly permuting the values of each feature and measuring the change in model performance. Changes that result in larger disparities in model performance lead to higher feature importance. Table 11 lists the computed feature importance in descending order of magnitude for features with positive, non-zero importance. Features with negative or zero importance are not reported, as these are not improving the predictive power.

**Table 11: Permutation Feature Importance of Manual Artificial Neural Network Model**

| **Feature** | **Importance** |
|---|---|
| Georgia | 0.0075 |
| California | 0.0072 |
| Current part-time employment (Yes) | 0.005 |
| New York | 0.0045 |
| Michigan | 0.0045 |
| Employment skills (Yes) | 0.0043 |
| Louisiana | 0.0034 |
| Gender | 0.0027 |
| Washington | 0.0023 |
| Social Security (No) | 0.0023 |
| Rhode Island | 0.0019 |
| Massachusetts | 0.0019 |
| Hispanic Origin (Yes) | 0.0018 |
| Alabama | 0.0017 |
| Delaware | 0.0015 |
| History of incarceration (Yes) | 0.0014 |
| Public food assistance (Not applicable) | 0.0014 |

| | |
|---|---|
| Connection to adult (No) | 0.0013 |
| South Carolina | 0.0012 |
| Iowa | 0.0012 |
| Indiana | 0.0011 |
| Texas | 0.0011 |
| Puerto Rico | 0.0011 |
| Kentucky | 0.0007 |
| Kansas | 0.0007 |
| Educational aid (Yes) | 0.0007 |
| Race declined (Yes) | 0.0006 |
| Has children (Yes) | 0.0006 |
| History of substance abuse counseling referral (Yes) | 0.0006 |
| Minnesota | 0.0006 |
| Connection to adult (Yes) | 0.0006 |
| Race unknown (No) | 0.0005 |
| Public food assistance (Yes) | 0.0004 |
| Wisconsin | 0.0004 |
| Utah | 0.0004 |
| Wyoming | 0.0004 |
| New Jersey | 0.0004 |
| Tennessee | 0.0003 |
| Virginia | 0.0003 |
| History of substance abuse counseling referral (No) | 0.0003 |
| Has children (No) | 0.0003 |

| | |
|---|---|
| New Mexico | 0.0002 |
| Oregon | 0.0002 |
| Public food assistance (Declined) | 0.0002 |
| Ohio | 0.0002 |
| Social Security (Declined) | 0.0002 |
| Native American (Yes) | 0.0002 |
| Asian (Yes) | 0.0001 |
| Mississippi | 0.0001 |
| History of substance abuse counseling referral (Declined) | 0.0001 |
| Hispanic origin (Unknown) | 0.0001 |
| Nebraska | 0.0001 |
| History of homelessness (Yes) | 0.0001 |
| Hawaiian/Pacific Islander (No) | 0.0001 |
| Race Declined (No) | 0.0001 |
| Hispanic origin (Declined) | 0.0001 |

XGBoost

XGBoost, short for Extreme Gradient Boosting, is a machine learning algorithm that builds an ensemble of decision trees with gradient boosting. While there are ten hyperparameters to fit, five will be searched for: the number of decision trees, the learning rate, the maximum depth of each tree, the minimum child weight, the fraction of trees randomly sampled without replacement at each split, and gamma. The number of decision trees in the ensemble decides the capacity of the model to capture complexity, needing to be balanced with the risk of overfitting. The learning rate controls how much the weights of the models are updated at each iteration. The maximum depth controls how many nodes a particular decision tree can have. The minimum child weight dictates the minimum number of training examples required to split a node further. The fraction of trees randomly sampled without replacement at each split indirectly dictates the size of the subset of features used in each split. The value of gamma controls the minimum loss reduction required to continue splitting a node further.

**Table 12: XGBoost Models and Results**

| Features | Accuracy | Recall | Precision | F1-Score |
|---|---|---|---|---|
| **Chi-squared** | .6417 | | | |
| Graduates | – | .8838 | .6559 | .7530 |
| Non-graduates | – | .2501 | .5709 | .3479 |
| **RFE** | .6390 | | | |
| Graduates | – | .8598 | .6594 | .7464 |
| Non-graduates | – | .2817 | .5541 | .3736 |
| **Manual** | .6444 | | | |
| Graduates | – | .8538 | .6654 | .7479 |
| Non-graduates | – | .3057 | .5638 | .3965 |

The manually selected features model performed the best, with accuracy of .6444. The hyperparameters selected for this model were 300 decision trees, a learning rate of .15, a maximum depth of each tree as 3, a minimum child weight of 2, a fraction of trees randomly sampled without replacement at each split of .90, and a gamma of 0. In order to interpret the model, the feature importances were computed using the gain metric—a way to measure the predictive power of a feature by measuring the resulting reduction in sum of squared errors (SSE) when the feature used to split the node. Table 13 displays the top 30 features and computed feature importances, ordered by importance. To see the full list of features and computed feature importances, refer to Appendix C.

**Table 13: Top 30 Features of Manual XGBoost Model by Importance**

| **Feature** | **Importance** |
|---|---|
| California | 38.5040 |
| Puerto Rico | 24.6258 |
| Current part-time employment (Yes) | 13.6768 |
| Georgia | 10.9762 |
| New York | 8.1693 |
| Louisiana | 7.7550 |
| Has children (Yes) | 7.5240 |
| Current part-time employment (No) | 7.4555 |
| Michigan | 6.3654 |
| History of incarceration (No) | 6.0785 |
| Employment skills (Yes) | 5.5162 |
| Washington | 5.0861 |
| Tennessee | 4.9870 |
| Pennsylvania | 4.6914 |
| Texas | 4.6757 |
| Alabama | 4.6551 |
| New Jersey | 4.6380 |
| History of substance abuse counseling referral (Yes) | 4.5822 |
| Has children (Declined) | 4.4833 |
| Gender | 4.4016 |

| | |
|---|---|
| Iowa | 4.3639 |
| Rhode Island | 4.3143 |
| Massachusetts | 4.1034 |
| South Carolina | 3.7795 |
| Public food assistance (Not applicable) | 3.7674 |
| Social security (No) | 3.7574 |
| Asian (No) | 3.7044 |
| Wisconsin | 3.6850 |
| Educational aid (Yes) | 3.6500 |
| Oregon | 3.5458 |

Random Forest

Random forest is an ensemble method that creates a model by combining multiple decision trees. The hyperparameters to fit are the number of decision trees, the maximum depth of each decision tree, the minimum samples required to split a node in a decision tree, and the minimum number of samples in a node in a decision tree. The number of decision trees can improve the accuracy of the model but must be weighed with the risk of overfitting. The maximum depth of each decision tree limits how many splits can be made in each tree. The minimum number of samples required to split a node dictates the point in which a decision tree will stop growing. The minimum number of samples in a leaf dictates how many nodes will be made—as nodes with fewer samples than the minimum will be grouped together in a different node.

**Table 14: Random Forest Models and Results**

| Features | Accuracy | Recall | Precision | F1-Score |
|---|---|---|---|---|
| **Chi-square** | .6379 | | | |
| Graduates | – | .9111 | .6470 | .7567 |
| Non-graduates | – | .1962 | .5769 | .2928 |
| **RFE** | .6348 | | | |
| Graduates | – | .8949 | .6481 | .7518 |
| Non-graduates | – | .2142 | .5574 | .3094 |
| **Manual** | .6350 | | | |
| Graduates | – | .9043 | .6463 | .7538 |
| Non-graduates | – | .1995 | .5631 | .2946 |

The best model is the one with features selected by the Chi-square test significance, with accuracy .6379. The best hyperparameters were the number of estimates as 800, the minimum samples in split as 8, the minimum samples in leaf as 4, and max depth as 18. The feature importances were computed using the reduction in Gini impurity in order to interpret the model's features. Table 15 displays the top 30 features and their computed feature importances, ordered by importance. The table of all feature importances can be found in Appendix D.

**Table 15: Top 30 Features of Chi-square Random Forest Model by Importance**

| Feature | Importance |
|---|---|
| California | 0.07103 |
| Gender | 0.03326 |
| Georgia | 0.02635 |
| History of incarceration (No) | 0.02632 |

| | |
|---|---|
| History of incarceration (Yes) | 0.02577 |
| Current part-time employment (Yes) | 0.02338 |
| Current part-time employment (No) | 0.02134 |
| Employment (full or part time) (Yes) | 0.02108 |
| Puerto Rico | 0.02068 |
| Employment (full or part time) (No) | 0.02005 |
| Employment skills (Yes) | 0.02002 |
| History of substance abuse counseling referral (No) | 0.01952 |
| Employment skills (No) | 0.01852 |
| Michigan | 0.01779 |
| History of substance abuse counseling referral (Yes) | 0.01658 |
| New York | 0.01626 |
| Social Security (No) | 0.01441 |
| Social Security (Yes) | 0.01339 |
| White (No) | 0.01320 |
| Homeless (No) | 0.01312 |
| White (Yes) | 0.01282 |
| Hispanic origin (No) | 0.01266 |
| History of homelessness (Yes) | 0.01233 |
| Other health insurance (No) | 0.01196 |
| Louisiana | 0.01174 |
| Black (No) | 0.01165 |
| Hispanic origin (Yes) | 0.01152 |
| Black (Yes) | 0.01143 |

| | |
|---|---|
| Has children (No) | 0.01109 |
| Race declined (No) | 0.01081 |

**Summarized Results and Policy Recommendations**

A Quick Note: Confirming Existing Findings

All four models included history of incarceration, history of substance abuse counseling referral, having children, and having a connection with an adult as important and highly ranked predictors. The coefficients' signs in Table 9 align with what other literature has stated on the topic: history of incarceration, history of substance abuse counseling referral, and having children are associated with a lesser likelihood of graduating high school. Having a connection with an adult is associated with a higher likelihood of graduating high school.

State Matters – A Lot: States Must Strive for Improvement

As seen from the best performing logistic regression and the ranked feature importances, state or area of residence is influential in whether a youth will graduate by Age 19. The most influential predictor of graduation in all models was a state. In particular, Puerto Rico, California, and Iowa have large positive coefficients in Table 9 relative to other predictors. Conversely, Georgia, New York, and Louisiana have large negative coefficients. Results from this paper show that state has a significant association with high school graduation.

When considering why these particular states stand out as over or underachieving for their youth in foster care compared to average, programming for transition to adulthood and completion of high school was investigated. A particular example of an overachieving state's programming is Iowa's Aftercare program. Established in 2004, it provides an additional robust safety net of support after youth age out of foster care. Youth are expected to meet with an Aftercare Self-Sufficiency Advocate twice a month who helps in addressing barriers they encounter and formulating plans for adult success (Youth Policy Institute of Iowa, 2020). An emphasis on personal responsibility is held in the program. Youth are required to spend 80 hours a month working or must be attending school to remain in the program (Youth Policy Institute of Iowa). Conversely, special programs could not be found in New York or Louisiana for their youth in foster care. Optimistically, Georgia is currently piloting one to a subset of youth in its state, but its current eligibility for participation is small (Multi-Agency Alliance for Children).

Another potential reason these states stand out is their accessibility of information online. California has a robust dashboard of success of educational outcomes that could provide public-facing pressure for schools and agencies to create environments conducive to better outcomes. Furthermore, the CalYOUTH study was conducted in California and the Midwest Evaluation was conducted in Iowa—these provide robust understandings of the challenges youth in foster care face when transitioning into adulthood. Conversely, lower-performing states may not understand how to serve their youth better to promote higher graduation rates. For instance, Georgia failed to report the graduation rate of their youth in foster care to the federal government in Academic Year 2017-2018, as noted in Table 3. It is unclear if states like Georgia have the information available but choose not to share it, or have a true gap in knowledge of the conditions of their youth in foster care.

Overall, states need to implement more programs that provide youth with stability when they age out of foster care—so that they can complete high school without additional pressures of losing state support. States also need to take more initiative in surveying the individual needs of youth in more in-depth ways themselves, or providing incentives for researchers to do so on their behalf.

School Quality: Influential in Outcomes

From Table 9, the second largest positive coefficient that is statistically significant—and the largest positive coefficient that is not a demographic characteristic—is receiving educational aid. At age 17, those receiving educational aid are most likely enrolled in tuition-based private high schools. This assumption is justified for two reasons: firstly, in the United States, those who graduate high school within their 45 days of their 17th birthday (the criteria to be in the baseline population and Cohort for the data) are outliers, as the youngest one can graduate by without special circumstances is within a few months of their 18th birthday. Secondly, only 98 out of 726 of the youth receiving educational aid in the NYTD data had reported graduating high school in the Age 17 survey, while the rest remained enrolled in school without reporting they had received a high school diploma. In this way, educational aid serves as a proxy for school quality.

Smithgall et al.'s 2004 study highlights that youth in foster care disproportionately attend lower quality schools. Findings from modeling conducted in this paper show that receiving educational aid is associated with a $e^{0.9849} \approx 2.67$ times likelihood of graduating high school by age 19, holding other factors constant. While school mobility and disruption is critical to avoid for youth in foster care, states should explore ways to improve access to quality education for youth in foster care. Holes created by NCLB and remaining in ESSA today that can prevent key decision makers from receiving notification that a youth is eligible to move to a better quality school must be patched, so that youth in foster care have access to the same choices in quality education as their peers.

Part-Time Employment: Consistently Powerful, Positive Predictor

In all models, part-time employment was ranked highly in feature importance. In Table 9, results from logistic regression show that having a part-time job is associated with a $e^{0.5802} \approx$

1.79 times likelihood of graduating high school by age 19, holding other factors constant. While having a part-time job is not studied much in the context of high school graduation and not at all for the foster care population, one study found that one having a part-time job during high school was associated with an 18% greater chance in graduating from school (O'Gorman & Pandey, 2015).

One potential reason is that having a part-time job can give a young person a better idea of the kind of career they want to pursue and motivation to complete a high school diploma if necessary. Furthermore, having control and independence in one's life through going to work and earning one's own finances could help alleviate the stressors youth in foster care face. Youth could form new relationships with their coworkers that also bring a sense of stability.

An alternative theory is that those who choose to work part-time jobs are already more likely to graduate from high school, since they are more driven. I push back on this being the sole explanation. As noted earlier, states that have positive predictive power in graduation rates and high feature importance have mechanisms to encourage youth to develop skills for stability. These states include part-time employment as a goal or a requirement, and in the process capture youth who may not have otherwise worked a part-time job. States with mechanisms not already in place should experiment with interventions that connect youth with part-time employment and evaluate if this is effective.

Receiving Social Security is a Signal: Additional Support is Necessary

In this survey, those who responded to receiving Social Security are either receiving Supplemental Security Income (SSI) payments or Social Security Disability Insurance (SSDI) payments—either directly or as a dependent beneficiary. SSI payments are made to eligible low-income persons with disabilities, meaning the youth or parent is fulfilling this criteria. SSDI payments are made to those with a predetermined length of work history who then develop a disability. Youth can only receive SSDI payments through a parent, as they are too young to have sufficient work history.

In Table 9, receiving Social Security was associated with an $e^{-0.2545} \approx 0.78$ times likelihood of graduating high school by age 19, holding other factors constant. While literature exists discussing the unique challenges of those who receive SSI payments and those who have disabilities at large in completing high school, no existing literature has described this as a predictor in high school graduation for youth in foster care. It is critical that states identify youth with this specific cross-section of experiences, especially since youth who have disabilities are more susceptible to abuse and neglect that lead into placement in foster care (Slayter, 2016).

More Information is Needed: A Critique of Current Data and Collection Mechanisms

While this paper justifies that the NYTD could be a representative sample of sex, race, and graduation rates, the current information collection system is in desperate need of improvement in order to fully understand and adequately predict high school graduation rates. The highest accuracy any model achieved—the XGBoost model with manually selected features (Table 12)—has unsatisfactory performance. With an accuracy of .6444, it is predicting whether a youth graduates from high school by age 19 correctly only 64.44% of the time. Furthermore, all models struggled in particular with correctly identifying those who did not graduate from high school—the recall for non-graduates on the model with highest accuracy was .3057, meaning it only correctly identified 30.57% of the youth who did not graduate from high school. The logistic regression model revealed significant p-values, a significant log-likelihood ratio test, and an abysmal $R^2$ value of .06. Modeling reveals that the information collected contains statistically significant predictors, but is far from adequate for deploying a model in practice.

Resolving this necessitates a three-pronged approach: better questions, better data quality, and better survey administration. Since high school completion is a key tool in social mobility, questions that can be used to identify those at risk of not graduating should be added and prioritized. These questions should include both ones regarding academic performance—which has successfully been used to create high accuracy machine learning models in previous studies of youth broadly—and factors identified in the literature review that have been significant in qualitative and quantitative studies, such as school mobility.

If survey apathy becomes apparent or response rates decline due to the addition of more questions, I propose that questions that are particularly strong predictors of high school graduation be prioritized in the front of the survey and questions about the receipt of resources that are conditional on one's exit from foster care be pushed further back. These questions were not particularly strong predictors of high school graduation—the best model of each algorithm ranked them extremely low or excluded them altogether. These questions should be especially de-prioritized because the number of youth who are not in foster care by age 19 are decreasing, as 48 states plus the District of Columbia now offer extended foster care after age 18 (Child Welfare Information Gateway, 2022). Many of these changes are recent, as many states moved towards extended foster care in light of COVID-19.

In order to improve data quality, the United States government should scrutinize the data that states submit for the NYTD more intensely and penalize states who either fail to comply with the ESSA's provisions on reporting for 4 year AGCR graduation rates or submit erroneous data consistently to the NYTD. 4 states submitted erroneous data regarding educational achievement to the NYTD for two Cohorts, and 3 of those did so at high rates. 9 states did not report their 4 year ACGR graduation rates for Academic Year 2017-2018: 7 are labeled as 'Not

Available' and 2 are labeled as 'suppressed due to concerns with data quality.' Since the states with immense amounts of erroneous data entry in the NYTD data (Colorado, South Dakota, and West Virginia) had to be dropped while modeling, the model would not be as effective in predicting the graduation rate in the three states.

Finally, survey administration must be standardized for NYTD, as states currently have their own discretion in how to administer the survey. Lack of standardization in administration is leading to drastically varied response rates that could fail to represent the characteristics of the baseline population within the Cohorts at the state level. Some states have response rates that are below 20%, while others have response rates upwards of 60% for the Age 19 survey. Improving survey administration practices by standardizing them across states will allow future Cohorts to be balanced in representativeness across states.

**Conclusion**

Youth in foster care graduate at far lower rates than their peers—even when comparing the 4 year ACGR with the graduation rate for youth in foster care by age 19. It is critical to resolve this gap, as high school graduation is the most accessible way for youth to achieve social mobility and open doors for success. While models deployed in this paper were not particularly strong at predicting graduation, there were still statistically significant features that are notable. In the short term, policies should be implemented that enhance resources noted to be significant and assist populations that have cross-sections of identities identified as particularly vulnerable.

In the long term, the United States federal government and state governments must both take drastic action to improve the state of affairs for youth in foster care by improving the data questions, quality, and the mechanism in which it is collected. There is a lack of broad and state-level quantitative studies for high school graduation that may never be addressed until better data is accessible. It should be one of the top priorities for governments looking to serve their youth in foster care and set them up for long-term success.

## Works Cited


Associated Press. (2023). Here's how an AI tool may flag parents with disabilities. Retrieved from https://apnews.com/article/child-protective-services-algorithms-artificial-intelligence-disability-02469a9ad3ed3e9a31ddae68838bc76e

Associated Press. (2022). Oregon dropping AI tool used in child abuse cases. Retrieved from https://apnews.com/article/politics-technology-pennsylvania-child-abuse-1ea160dc5c2c203fdab456e3c2d97930

Burley, M. (2013). Educational outcomes of foster youth-updated benchmarks (document number 13-06-3901). Retrieved from the Washington State Institute for Public Policy website. http://www.wsipp.wa.gov/ReportFile/1341/Wsipp_EducationalOutcomes-of-Foster-Youth-Updated-Benchmarks_Full-Report.pdf

Child Welfare Information Gateway. (2022). Extension of foster care beyond age 18. U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau. https://www.childwelfare. gov/topics/systemwide/laws-policies/statutes/extensionfc/

Child Welfare Information Gateway (n.d.). Foster Care. Retrieved March 12, 2023, from https://www.childwelfare.gov/topics/outofhome/foster-care/

Clemens, E. V., Lalonde, T. L., & Sheesley, A. P. (2016). The relationship between school mobility and students in foster care earning a High School credential. *Children and Youth Services Review*, *68*, 193–201. https://doi.org/10.1016/j.childyouth.2016.07.016

Courtney, M. & Dworsky, A. (2006a). Midwest Evaluation of the Adult Functioning of Former Foster Youth from Illinois: Outcomes at Age 19. Chicago: Chapin Hall at the University of Chicago.

Courtney, M. & Dworsky, A. (2006b). Midwest Evaluation of the Adult Functioning of Former Foster Youth from Iowa: Outcomes at Age 19. Chicago: Chapin Hall at the University of Chicago.

Courtney, M. & Dworsky, A. (2006c). Midwest Evaluation of the Adult Functioning of Former Foster Youth from Wisconsin: Outcomes at Age 19. Chicago: Chapin Hall at the University of Chicago.

Courtney, M. E., Okpych, N. J., Charles, P., Mikell, D., Stevenson, B., Park, K., Kindle, B., Harty, J., & Feng, H. (2016). Findings from the California Youth Transitions to Adulthood Study (CalYOUTH): Conditions of foster youth at age 19. Chicago, IL: Chapin Hall at the University of Chicago.

DeBaun, B. & Roc, M.; Alliance for Excellent Education (2013). Well and Well-Off: Decreasing Medicaid and Health-Care Costs by Increasing Educational Attainment.

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Government Technology. (2017). Illinois Ends Child Abuse Prediction Program. Retrieved from https://www.govtech.com/health/illinois-ends-child-abuse-prediction-program.html

Hahn, R. A., Knopf, J. A., Wilson, S. J., Truman, B. I., Milstein, B., Johnson, R. L., Fielding, J. E., Muntaner, C. J. M., Jones, C. P., Fullilove, M. T., Moss, R. D., Ueffing, E., & Hunt, P. C. (2015). Programs to increase high school completion. *American Journal of Preventive Medicin*e, 48(5).

Lenderman, K. & Hawkins, J. (2021). Out of the classroom and less likely to graduate: The relationship between exclusionary discipline and four-year graduation rates in Texas. Texas Education Review, 9(2), 6-20. http://dx.doi.org/10.26153/tsw/13913

Leslie, L., Holmes, L., Hitrova, C., & Ott, E. (2020). Ethics of machine learning in children's social care [Report]. What Works for Children's Social Care. https://www.turing.ac.uk /news/publications/ethics-machine-learning-childrens-social-care

Marrus E (2015). Education in Black America: Is It the New Jim Crow? *Arkansas Law Review (1968-Present)*, 68(1), 27–54. https://heinonline.org/HOL/LandingPage?handle= hein.journals/arklr68&div=7&id=&page=

Multi-Agency Alliance for Children (n.d.). Learn, Educate, Achieve, Dream, and Succeed. https://www.maac4kids.org/leads/

Office of Juvenile Justice and Delinquency Prevention (2021). Youth (0 to 17) population profile detailed by age, sex, and Race/ethnicity, 2020. OJJDP Statistical Briefing Book. https://www.ojjdp.gov/ojstatbb/population/qa01104.asp?qaDate=2020.

O'Gorman, M. & Pandey, M. (2015). Cultivating the Arctic's Most Valuable Resource: An Analysis of the Barriers to High School Completion Among Youth in Nunavut. Department of Economics, University of Winnipeg.

Okpych, N. J., Courtney, M., & Dennis, K. (2017). Memo from CalYOUTH: Predictors of high school completion and college entry at ages 19/20. Chicago, IL: Chapin Hall at the University of Chicago.

Ploutz, C. (2018). Machine Learning Applications in Graduation Prediction at the University of Nevada, Las Vegas. UNLV Theses, Dissertations, Professional Papers, and Capstones. 3309. http://dx.doi.org/10.34917/13568668

Reeves, R., Buckner, E., & Smith, E. (2022). The unreported gender gap in high school graduation rates. Brookings Institution. https://www.brookings.edu/blog/up-front/ 2021/01/12/the-unreported-gender-gap-in-high-school-graduation-rates/

Sandh, S., Donaldson, V. M., & Katz, C. C. (2020). Students connected to foster care: An overview of high school experiences. *Children and Youth Services Review*, *113*, 104905. https://doi.org/10.1016/j.childyouth.2020.104905

Scherr, T. G. (2007). Educational experiences of children in foster care: Meta-analyses of special education, retention and discipline rates. School Psychology International, 28(4), 419–436. doi:10.1177/0143034307084133

Smithgall, C., Gladden, R., Howard, E., George, R., & Courtney, M. (2004). Educational Experiences of Children in Out-Of-Home Care. Chicago: Chapin Hall at the University of Chicago.

Slayter, E. (2016). Youth with disabilities in the United States. 64, 155-165. doi: 10.1016/j.childyouth.2016.03.012

Sum, A. M., Khatiwada, I., & McLaughlin, J. (2009). The consequences of dropping out of high school. https://repository.library.northeastern.edu/downloads/neu:376324

U.S. Bureau of Labor Statistics. (2022). Education pays, 2021. *Career Outlook*. Retrieved from https://www.bls.gov/careeroutlook/2022/data-on-display/education-pays.htm

U.S. Census Bureau. (2012). “GED Recipients Have Lower Earnings, are Less Likely to Enter College.” Retrieved from https://www.census.gov/newsroom/blogs/random-samplings /2012/02/ged-recipients-have-lower-earnings-are-less-likely-to-enter-college.html

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (n.d.).  Public high school 4-year adjusted cohort graduation rate (ACGR), by selected student characteristics and state: 2010-11 through 2017-18. National Center for Education Statistics (NCES) Home Page, a part of the U.S. Department of Education. Retrieved from https://nces.ed.gov/programs/digest/d19/tables/dt19_219.46.asp

Vaughn, M. G., Salas-Wright, C. P., & Maynard, B. R. (2014). Dropping out of school and chronic disease in the United States. *Zeitschrift fur Gesundheitswissenschaften = Journal of public health*, *22*(3), 265–270. https://doi.org/10.1007/s10389-014-0615-x

West, K. K., Mathews, B. L., & Kerns, K. A. (2013). Mother-child attachment and cognitive performance in middle childhood: An examination of mediating mechanisms. *Early Childhood Research Quarterly*, 28, 259–270. doi:10.1016/j.ecresq.2012.07.005

Youth Policy Institute of Iowa (2020). Aftercare Facts and Information. https://iowaaftercare.org/wp-content/uploads/2020/06/AftercareFactSheet_2019_FINAL_UPDATED_onWEB.pdf.

## Appendix A: Survey Questions

| Question | Condition to receive (if not asked to all) |
|---|---|
| Currently are you employed full-time? | |
| Currently are you employed part-time? | |
| In the past year, did you complete an apprenticeship, internship, or other on-the-job training, either paid or unpaid? | |
| Currently are you receiving social security payments (Supplemental Security Income (SSI, Social Security Disability Insurance (SSDI), or dependents' payments)? | |
| Currently are you using a scholarship, grant, stipend, student loan, voucher, or other type of educational financial aid to cover any educational expenses? | |
| Currently are you receiving any periodic and/ or significant financial resources or support from another source not previously indicated and excluding paid employment? | |
| What is the highest educational degree or certification that you have received? | |
| Currently are you enrolled in and attending high school, GED classes, post-high school vocational training, or college? | |
| Currently is there at least one adult in your life, other than your caseworker, to whom you can go for advice or emotional support? | |
| Have you ever been homeless? | |
| Have you ever referred yourself or has someone else referred you for an alcohol or drug abuse assessment or counseling? | |

| | |
|---|---|
| Have you ever been confined in a jail, prison, correctional facility, or juvenile or community detention facility, in connection with allegedly committing a crime? | |
| Have you ever given birth or fathered any children that were born? | |
| If you responded yes to the previous question, were you married to the child's other parent at the time each child was born? | |
| Currently are you on Medicaid [or use the name of the State's medical assistance program under title XIX]? | |
| Currently do you have health insurance, other than Medicaid? | |
| Does your health insurance include coverage for medical services? | Only asked if responded yes to having other health coverage |
| Does your health insurance include coverage for mental health services? | Only asked if responded yes to having other health coverage |
| Does your health insurance include coverage for prescription drugs? | Only asked if responded yes to having other health coverage |
| Currently are you receiving ongoing welfare payments from the government to support your basic needs? [The State may add and/ or substitute the name(s) of the State's welfare program]. | Only asked if not in foster care |
| Currently are you receiving public food assistance? | Only asked if not in foster care |
| Currently are you receiving any sort of housing assistance from the government, such as living in public housing or receiving a housing voucher? | Only asked if not in foster care |

## Appendix B: One-hot Encoded Features from Recursive Feature Elimination and Pruning

| Feature Category | Selected |
|---|---|
| State | Alabama, California, Delaware, Georgia, Hawaii, Iowa, Indiana, Kentucky, Louisiana, Massachusetts, Maine, Michigan, Minnesota, New Jersey, New Mexico, New York, Ohio, Oregon, Pennsylvania, Puerto Rico, Rhode Island, South Carolina, Texas, Utah, Vermont, Washington, Wyoming |
| Gender | Male |
| Race | Hispanic origin (declined) |
| Part-time employment | Yes, declined |
| Employment skills | Yes |
| Social security | Yes |
| Educational aid | Yes, no, declined |
| Public financial assistance | Declined |
| Public food assistance | Declined |
| Private financial assistance | Yes, no, declined |
| Connection to an adult | No |
| History of substance abuse referral | Yes |
| History of incarceration | Yes |
| Has children | No, declined |
| Married, if have children | No |
| Other health insurance | Yes |
| Mental health insurance, if have other health insurance | Yes, do not know |
| Prescription health insurance, if have other health insurance | Not applicable |

## Appendix C: All Ordered Feature Importance Computed by Gain of Manual XGBoost Model

| Feature | Importance |
|---|---|
| California | 38.5040 |
| Puerto Rico | 24.6258 |
| Current part-time employment (Yes) | 13.6768 |
| Georgia | 10.9762 |
| New York | 8.1693 |
| Louisiana | 7.7550 |
| Has children (Yes) | 7.5240 |
| Current part-time employment (No) | 7.4555 |
| Michigan | 6.3654 |
| History of incarceration (No) | 6.0785 |
| Employment skills (Yes) | 5.5162 |
| Washington | 5.0861 |
| Tennessee | 4.9870 |
| Pennsylvania | 4.6914 |
| Texas | 4.6757 |
| Alabama | 4.6551 |
| New Jersey | 4.6380 |
| History of substance abuse counseling referral (Yes) | 4.5822 |
| Has children (Declined) | 4.4833 |
| Gender | 4.4016 |
| Iowa | 4.3639 |
| Rhode Island | 4.3143 |

| | |
|---|---|
| Massachusetts | 4.1034 |
| South Carolina | 3.7795 |
| Public food assistance (Not applicable) | 3.7674 |
| Social security (No) | 3.7574 |
| Asian (No) | 3.7044 |
| Wisconsin | 3.6850 |
| Educational aid (Yes) | 3.6500 |
| Oregon | 3.5458 |
| History of incarceration (Yes) | 3.5249 |
| North Carolina | 3.3127 |
| Kansas | 3.2198 |
| Delaware | 3.1800 |
| Kentucky | 3.1031 |
| Mississippi | 3.0542 |
| Current part time employment (Declined) | 2.9941 |
| Indiana | 2.9755 |
| Utah | 2.7898 |
| Employment skills (No) | 2.7398 |
| Hispanic origin (Yes) | 2.6848 |
| Virginia | 2.5798 |
| Minnesota | 2.5525 |
| Connecticut | 2.4025 |
| History of substance abuse counseling referral (No) | 2.4188 |
| North Dakota | 2.4078 |
| History of homelessness (Declined) | 2.4001 |

| | |
|---|---|
| Connection to an adult (No) | 2.3198 |
| Has children (No) | 2.3000 |
| Florida | 2.2907 |
| History of homelessness (Yes) | 2.2864 |
| Race unknown (No) | 2.2169 |
| Connection to an adult (Yes) | 2.1911 |
| Social Security (Yes) | 2.1088 |
| Hawaiian/Pacific Islander (Yes) | 2.0965 |
| Nevada | 2.0279 |
| Race Unknown (Yes) | 1.9979 |
| Missouri | 1.8903 |
| White (No) | 1.8479 |
| Native American (No) | 1.8367 |
| Homeless (No) | 1.7739 |
| Race declined (No) | 1.7449 |
| Social Security (Declined) | 1.6882 |
| Hispanic origin (Unknown) | 1.6591 |
| Hispanic origin (No) | 1.6322 |
| Employment skills (Declined) | 1.5567 |
| Oklahoma | 1.5190 |
| History of substance abuse counseling referral (Declined) | 1.5074 |
| Public food assistance (Yes) | 1.5062 |
| Ohio | 1.5039 |
| New Hampshire | 1.4567 |
| Nebraska | 1.4527 |

| | |
|---|---|
| History of incarceration (Declined) | 1.4008 |
| Hawaii | 1.3052 |
| Wyoming | 1.2966 |
| Black (No) | 1.2929 |
| Public food assistance (Declined) | 1.2689 |
| Asian (Yes) | 1.2633 |
| Vermont | 1.2478 |
| Connection to adult (Declined) | 1.1948 |
| Educational aid (No) | 1.1100 |
| Public food assistance (No) | 1.1014 |
| White (Yes) | 1.0640 |
| Black (Yes) | 0.9763 |
| Maryland | 0.8280 |
| New Mexico | 0.8057 |
| Arizona | 0.6614 |
| Montana | 0.6588 |
| Native American (Yes) | 0.6050 |
| Hispanic origin (Declined) | 0.5629 |
| Educational aid (Declined) | 0.5549 |
| Maine | 0.4860 |
| Idaho | 0.4140 |
| Arkansas | 0.3067 |
| District of Columbia | 0.2752 |
| Race declined (Yes) | 0.2511 |
| Hawaiian/Pacific Islander (No) | 0.1442 |

### Appendix D: Ordered Feature Importance of Chi-square Random Forest Model

| Feature | Importance |
|---|---|
| California | 0.07103 |
| Gender | 0.03326 |
| Georgia | 0.02635 |
| History of incarceration (No) | 0.02632 |
| History of incarceration (Yes) | 0.02577 |
| Current part-time employment (Yes) | 0.02338 |
| Current part-time employment (No) | 0.02134 |
| Employment (full or part time) (Yes) | 0.02108 |
| Puerto Rico | 0.02068 |
| Employment (full or part time)? (No) | 0.02005 |
| Employment skills (Yes) | 0.02002 |
| History of substance abuse counseling referral (No) | 0.01952 |
| Employment skills (No) | 0.01852 |
| Michigan | 0.01779 |
| History of substance abuse counseling referral (Yes) | 0.01658 |
| New York | 0.01626 |
| Social Security (No) | 0.01441 |
| Social Security (Yes) | 0.01339 |
| White (No) | 0.01320 |
| Homeless (No) | 0.01312 |
| White (Yes) | 0.01282 |
| Hispanic origin (No) | 0.01266 |

| | |
|---|---|
| History of homelessness (Yes) | 0.01233 |
| Other health insurance (No) | 0.01196 |
| Louisiana | 0.01174 |
| Black (No) | 0.01165 |
| Hispanic origin (Yes) | 0.01152 |
| Black (Yes) | 0.01143 |
| Has children (No) | 0.01109 |
| Race declined (No) | 0.01081 |
| Other financial assistance (No) | 0.01004 |
| Medicaid (Yes) | 0.00999 |
| Other health insurance (Do not know) | 0.00972 |
| Other financial assistance (Yes) | 0.00919 |
| Public food assistance (Not applicable) | 0.00910 |
| Connection to adult (Yes) | 0.00888 |
| Public housing assistance (Not applicable) | 0.00878 |
| Public financial assistance (Not applicable) | 0.00876 |
| If had a child, was married (No) | 0.00848 |
| Has children (Yes) | 0.00817 |
| Connection to adult (No) | 0.00815 |
| Alabama | 0.00754 |
| Texas | 0.00747 |
| Public financial assistance (No) | 0.00732 |
| Currently enrolled (Yes) | 0.00718 |
| Public food assistance (No) | 0.00708 |
| Public housing assistance (No) | 0.00707 |

| | |
|---|---|
| Washington | 0.00706 |
| Married, if has children (Not applicable) | 0.00704 |
| Rhode Island | 0.00693 |
| Race declined (Yes) | 0.00672 |
| Educational aid (Yes) | 0.00669 |
| Currently enrolled (No) | 0.00653 |
| Massachusetts | 0.00649 |
| Medicaid (No) | 0.00648 |
| Pennsylvania | 0.00644 |
| Educational aid (No) | 0.00636 |
| Other health insurance (Yes) | 0.00623 |
| Has medical insurance, if has other health insurance (Not applicable) | 0.00608 |
| Medicaid (Do not know) | 0.00582 |
| South Carolina | 0.00582 |
| Has mental health insurance, if has other health insurance (Not applicable) | 0.00571 |
| Has prescription health insurance, if has other health insurance (Not applicable) | 0.00563 |
| Iowa | 0.00555 |
| Has mental health insurance, if has other health insurance (Yes) | 0.00549 |
| Has prescription health insurance, if has other health insurance (Yes) | 0.00545 |
| Has medical health insurance, if has other health insurance (Yes) | 0.00520 |
| Virginia | 0.00494 |
| Asian (No) | 0.00490 |

| | |
|---|---|
| Minnesota | 0.00488 |
| Florida | 0.00484 |
| Kentucky | 0.00474 |
| Kansas | 0.00473 |
| New Jersey | 0.00450 |
| Wisconsin | 0.00444 |
| Native American (No) | 0.00431 |
| Utah | 0.00410 |
| Indiana | 0.00400 |
| Hawaiian/Pacific Islander (No) | 0.00384 |
| Delaware | 0.00372 |
| Race unknown (No) | 0.00363 |
| Social Security (Declined) | 0.00358 |
| Public food assistance (Yes) | 0.00353 |
| Oregon | 0.00349 |
| Native American (Yes) | 0.00340 |
| Race Unknown (Yes) | 0.00311 |
| Tennessee | 0.00311 |
| Ohio | 0.00303 |
| Hispanic origin (Unknown) | 0.00293 |
| Missouri | 0.00285 |
| Has children (Declined) | 0.00284 |
| Oklahoma | 0.00284 |
| Has mental health insurance, if has other health insurance (Do not know) | 0.00278 |
| History of homelessness (Declined) | 0.00275 |

| | |
|---|---|
| Mississippi | 0.00271 |
| Connecticut | 0.00256 |
| Illinois | 0.00259 |
| History of substance abuse counseling referral (Declined) | 0.00253 |
| Maryland | 0.00232 |
| Nevada | 0.00232 |
| North Carolina | 0.00224 |
| Other financial assistance (Declined) | 0.00221 |
| History of incarceration (Declined) | 0.00220 |
| Nebraska | 0.00215 |
| Medicaid (Declined) | 0.00189 |
| Connection to adult (Declined) | 0.00185 |
| Other health insurance (Declined) | 0.00183 |
| Asian (Yes) | 0.00180 |
| Wyoming | 0.00171 |
| Educational aid (Declined) | 0.00170 |
| Has medical health insurance, if has other health insurance (Do not know) | 0.00170 |
| Employment skills (Declined) | 0.00166 |
| Has prescription insurance, if has other health insurance (Do not know) | 0.00163 |
| Current part-time employment (Declined) | 0.00160 |
| Currently enrolled (Declined) | 0.00154 |
| Public financial assistance (Yes) | 0.00153 |
| Vermont | 0.00137 |
| Arkansas | 0.00131 |

| | |
|---|---|
| Has medical insurance, if has other health insurance (No) | 0.00121 |
| Arizona | 0.00109 |
| Public financial assistance (Declined) | 0.00108 |
| North Dakota | 0.00102 |
| New Hampshire | 0.00092 |
| Montana | 0.00090 |
| New Mexico | 0.00086 |
| Hawaiian/Pacific Islander (Yes) | 0.00083 |
| Idaho | 0.00082 |
| Public housing assistance (Yes) | 0.00082 |
| Married, if has children (Declined) | 0.00071 |
| Public housing assistance (Declined) | 0.00071 |
| DC | 0.00068 |
| Has mental health insurance, if has other health insurance (No) | 0.00061 |
| Hawaii | 0.00055 |
| Public food assistance (Declined) | 0.00054 |
| Has prescription insurance, if has other health insurance (No) | 0.00045 |
| Hispanic origin (Declined) | 0.00022 |
| Maine | 0.00020 |
| Has medical health insurance, if has other health insurance (Declined) | 0.00011 |
| Has mental health insurance, if has other health insurance (Declined) | 0.00000 |
| If had a child, was married (Yes) | 0.00000 |

| | |
|---|---|
| Has prescription insurance, if has other health insurance (Declined) | 0.00000 |
| Has other health insurance (Not applicable) | 0.00000 |