Summer 7-30-2021

An Exploratory Epidemiological Analysis Investigating the Representativeness of Children with Likely Autism Served by State Early Intervention and Special Education Systems

Jia Wang

doi: https://doi.org/10.57709/24095731

AN EXPLORATORY EPIDEMIOLOGICAL ANALYSIS INVESTIGATING THE REPRESENTATIVENESS OF CHILDREN WITH LIKELY AUTISM SERVED BY STATE EARLY INTERVENTION AND SPECIAL EDUCATION SYSTEMS

by

JIA WANG

B.M.S., Inner Mongolia Medical University

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

TABLE OF CONTENTS

List of Tables iii
List of Figures iv
ACKNOWLEDGMENTS vi
INTRODUCTION 1
1.1 Background 1
1.2 Purpose of Study 2
1.3 Research Questions 2
REVIEW OF THE LITERATURE 4
2.1 Autism Spectrum Disorder (ASD) 4
2.2 ASD identification 4
2.3 ASD diagnostic systems 6
2.4 Evaluating prevalence of ASD 7
2.5 Application of ASD prevalence 8
METHODS 12
3.1 Data Source and Preparation 12
3.2 Outcome Variable 12
3.3 Covariates 13
3.4 Descriptive Statistics 17
3.5 Correlation 20
3.6 Statistical Analysis 31
RESULTS 33
4.1 Associations between likely percentage missing of ASD and relevant county level socio-demographic covariates using simple and multiple linear regression 33
Bivariate Regression Models 33
Regression Models with Continuous and State Predictors 35
Regression Models with Interactions Between Continuous and State Predictors 37
DISCUSSION 45
5.1 Discussion of Results 45
5.2 Limitation and Future Directions 51
5.3 Conclusion 52
References 53
APPENDIX 59

List of Tables

Table 1.Characteristics of continuous variables from 2020 County Health Rankings Annual Report (N=184) 19

Table 2. Spearman correlations between county level predictor variables and likely percentage missing of ASD in IDEA data 22

Table 3. Spearman correlations between the independent variables with each other in 2020 County Health Rankings annual report 27

Table 4. Spearman correlations between independent variables with each other in 2020 County Health Rankings annual report in CA 28

Table 5. Spearman correlations between the predictors with each other in 2020 County Health Rankings annual report in ID 29

Table 6. Spearman correlations between independent variables with each other in 2020 County Health Rankings annual report in MS 30

Table 7. Comparison of linear regression models between bivariate regression models (model 1), regression models with continuous and state predictors (model 2) and regression models with interactions between continuous and state predictors (model 3) 41

Table 8. Spearman correlations between independent variables with each other in 2020 County Health Rankings annual report in CA, ID and MS 59

Table 9. Comparison of mixed effects models between Random-intercept model (Model 4) and Random-slope model (Model 5) 61

List of Figures

Figure 1. Detail on the Function 32

APPROVAL PAGE

AN EXPLORATORY EPIDEMIOLOGICAL ANALYSIS INVESTIGATING THE REPRESENTATIVENESS OF CHILDREN WITH LIKELY AUTISM SERVED BY STATE EARLY INTERVENTION AND SPECIAL EDUCATION SYSTEMS

by

JIA WANG

Approved:

___Brian Barger, Ph.D.______
Committee Chair

__Ruiyan Luo, Ph.D. _____
Committee Member

__July 30th, 2021 ______________
Date

ACKNOWLEDGMENTS

I would like to acknowledge everyone who played a role in my academic accomplishments. First of all, thanks to my committee for their patient advice and continued support: Dr. Brian Barger, my committee chair, and Dr. Ruiyan Luo. Secondly, thanks to Dr. Ami Klin, the director of Marcus Autism Center, Dr. Teal Benevides and Ladson, Kiana from Augusta University for sharing their data and work. Finally, thanks to my parents, sisters, and my husband, Robert, who supported me with great love and understanding.

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Jia Wang
Signature of Author

Abstract

Autism Spectrum Disorder (ASD) is a developmental disability that occurs among people across different socio-demographic groups. According to the Centers for Disease Control and Prevention (CDC) estimate, nationally 1.9% of children in the USA were diagnosed with ASD in 2016. However, the diagnostic and identification of ASD vary greatly across states. Differences across states are likely to impact the found prevalence for children with ASD in those states, which can cause the potential number of missed ASD cases to vary. The purpose of this study is to develop a *potentially missed ASD case metric* from available school data and investigate the relationship between the missed case metrics and relevant county-level socio-demographic covariates. The study focuses on the relationship between *potentially missed ASD* and ten variables include *states* (Idaho, Mississippi, and California), *primary care physicians per 100,000 (2017), mental health providers per 100,000 (2019), children in poverty per 100,000 (2018), uninsured children per 100,000 (2017), residential segregation rate between black and white (2014-2018), high school graduation per 100,000 (2016-2017), median household income (2018), child mortality per 100,000 (2015-2018), and the percent of rural population* based on Census Population Estimates (2010). Results: Simple correlation and regression models displayed significant relationships between *potential missing ASD* and most predictors. However, after including states, many predictors are not significant anymore, suggesting that individual states are an important source of variance to consider for analysis of *missed ASD*. By adding interactions between continuous predictors and states into the multiple linear regression models, *uninsured children* and *percent of rural populations* show significant differences between states with predictors of the relationship to missing ASD cases. Future studies should consider linear mixed models for the analysis of missing *ASD.*

Chapter 1
INTRODUCTION

1.1 Background

Autism Spectrum Disorders (ASD) are a set of neurodevelopmental disorders, and individuals who suffer from ASD are mainly affected in their communication, behavior, and social performance (*American Psychiatric Association*, 2013.) A child is considered to have ASD when diagnosed by both clinical criteria and educational classification systems. In the U.S., clinical diagnosis is consistent with the Diagnostic and Statistical Manual of Mental Disorders (DSM) (Randall et al., 2018). Furthermore, there is an educational classification of ASD based on the presence of ASD symptoms leading to negative impacts on children's academic achievement (Barton et al., 2016). Specifically, ASD children who are three to 21 years of age and in the public school system can receive special education services based on the federal special education law of Individuals with Disabilities Education Act (IDEA), though the criteria to identify ASD in the public school system are different across states' Departments of Public Health and Departments of Education (Gist & Stein, 2014) .The prevalence of ASD and other education–system information about ASD in IDEA systems rely on each state's Department of Education. The IDEA reports the prevalence to Congress every academic year (Mandell & Palmer, 2005). The official ASD prevalence data is derived from the U.S. Centers for Disease Control and Prevention's (CDC) Autism and Developmental Disabilities Monitoring (ADDM) Network (Nevison et al., 2018). In the past, ASD was considered a rare disorder that only affected about one in 2000 individuals (Rice et al., 2012). However, the reported prevalence of ASD has risen rapidly in recent years (Fombonne, 2018). According to the estimate by the CDC, one in every 88 children had ASD in 2008, but by 2016, the ASD prevalence reached one in 54 (Maenner, 2020). ASD prevalence increased by 71% from 2008 to 2016. It is unknown if the increases in ASD prevalence are due to a true increase in the risk of developing ASD or different identification standards across states' education departments (Rice et al., 2012). However, it is notable that more children are diagnosed with ASD now than in the past. Due to the high increase in the prevalence, it is important to understand what socio-demographic factors relate to ASD, and the percentage of children with ASD at county-level potentially missed from screening and identification.

1.2 Purpose of Study

The purpose of this study is to develop an initial metric of county level *missed ASD cases* and see how *missed autism case rate*s might relate to relevant county-level socio-demographic covariates.

1.3 Research Questions

The following questions aimed to be investigated by the study:

1. Can we create a potentially missed case metric that could prove to be a sensitive measure for studies determining if the prevalence of ASD eligibility status in school systems is similar to CDC estimates?
2. Does the potential missed ASD case metric correlates with socio-demographic covariates the literature indicates are associated with ASD prevalence?
3. Does the missed ASD case metric display discriminant validity (i.e., is it reasonable, but not perfectly, correlated with other important variables)?
4. Do states differ in their potential missed ASD cases?
5. Which socio-demographic covariates are most predictive of potentially missed cases?
6. From these initial analyses does a linear regression or linear mixed regression approach seem to be the most appropriate for data sets comprising the county-level data from the

entire U.S.? Justify whether a linear mixed regression approach is justified and outline considerations for fixed and random effects based on these preliminary data.

Chapter 2

REVIEW OF THE LITERATURE

2.1 Autism Spectrum Disorder (ASD)

Autism Spectrum Disorder (ASD) is a developmental disability. ASD's main characteristics include delayed communication development, difficulty with social interaction, repetitive behavior patterns, and limited interests or activities (*American Psychiatric Association*, 2013). The symptoms of ASD are heterogeneous among people (Jones & Klin, 2009). Most children will express the core symptoms throughout their lives, and symptoms can be mild or severe (Jones & Klin, 2009). Some children show their ASD symptoms as early as a few months, but many other children have their symptoms noticed by their family, teachers or physicians after a few years (CDC, 2021). In general, many children do not receive a final diagnosis until they are much older (Sheldrick et al., 2017). The American Psychiatric Association recommends that the average age at diagnosis in the U.S.is 4 years old (*2016*). According to the Centers for Disease Control and Prevention (CDC), 1 in every 54 (1.9%) children was diagnosed with ASD in 2016. ASD occurs among all gender, racial, ethnic, and socioeconomic groups, and is 3-4 times more commonly identified in boys than in girls (*NIMH, 2020*). Data for ASD from the Autism and Developmental Disabilities Monitoring (ADDM) Network found a similar prevalence of ASD among non-Hispanic black children as compared to non-Hispanic white children (CDC, 2020). However, the number of Hispanic children identified with ASD has a lower frequency than non-Hispanic white or black children (Becerra et al., 2014).

2.2 ASD identification

The earlier children with ASD are identified, the more promptly early intervention (EI) services and supports can be delivered, which can benefit their personal development and quality of life in the future (Landa, 2018). In the U.S, children from birth through age three can have access to many state-mandated EI services, including Part C of IDEA, Child Find, EI Eligibility Assessment, and Individualized Family Service Plan (IFSP) systems. These services identify and serve children who meet the state-defined development delays or disabilities (Barger et al., 2018a). Child Find systems located in state Departments of Education are tasked with locating and evaluating children whose disabilities impact their educational performance (Ennis et al., 2017). Part C is administered in different systems across states, most commonly Departments of Public Health and Departments of Education (Bricker et al., 2013). Furthermore, children ages three to 21 identified with physical, developmental, and mental health conditions that impact their educational performance can then receive special education services under IDEA Part B (MacFarlane & Kanaya, 2009).

In addition to formal identification systems routing children to services, successful early identification requires accurate screening tools and informal monitoring approaches (Barger et al., 2018b). Screening tools are most commonly studied and are typically brief caretaker rating scales that are scored to indicate the presence of particular developmental problems that might require further assessment. Screening measures are classified as Level 1 and Level 2 instruments (Petrocchi et al., 2020). Level 1 screening measures have been applied to the general population to identify children with ASD or other developmental disorders. Level 2 screening measures have been applied to identify children at a higher risk of developmental issues or family members with ASD (Petrocchi et al., 2020). Developmental monitoring refers to how children's development progresses over time and whether they are meeting their milestones of development at rates similar to other children their age (Lipkin et al., 2020). Recently, research shows that children who simultaneously received developmental monitoring and screening together are more likely to get EI or community-based treatment compared to children receiving screening or monitoring alone, or no screening and monitoring at all (Barger et al., 2018b). Furthermore, recent research suggests that screening and monitoring received together is more strongly associated with early ASD identification than either received alone (Barger et al., 2018b).

2.3 ASD diagnostic systems

Once children are identified via screening and/or monitoring there is typically a lengthy diagnostic period wherein a (preferably) multi-disciplinary team of clinicians seek to determine the presence of developmental conditions (Randall et al., 2018). From a medical perspective, the clinical diagnosis of ASD is primarily rendered using one of two classification frameworks—the International Classification of Diseases and the Diagnostic (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM) (Randall et al., 2018). ICD shows a high similarity to DSM. However, they differ in scale and international reach: ICD is a global categorization system for physical and mental illness published by the World Health Organization (WHO); whereas DSM is primarily used for diagnosis and billing purposes in the U.S.(Clark et al., 2017). The ICD-11 is applied to similar areas as DSM, like clinical areas and public health, but it is used as a global system to diagnose ASD (Clark et al., 2017). Generally, the DSM and ICD codes align.

In the U.S., early identification of ASD is complicated by an educational eligibility system that uses ASD as a classification code (Maenner, 2020). While the medical diagnoses of ASD are typically based on the standard criteria (*American Psychiatric Association*, 2013), an educational classification is based on the children's academic achievement under the influence of their disability like ASD (Barton et al., 2016). In 1990, the federal special education law, IDEA, qualified ASD as a separate condition, and eligible children with ASD qualified for special education without a medical diagnosis (Zirkel, 2011). IDEA indicates that each state can establish its own educational evaluation criteria for children's eligibility to access special education services as long as each state's ASD eligibility criteria meets the essential requirement by the federal regulations of ASD, which do not necessarily always comport to the DSM classification criteria (Barton et al., 2016). Critically, to receive an IDEA eligibility determination of ASD a child must not only display core symptoms of ASD, but must also provide evidence that the symptoms have a negative impact on educational outcomes.

Data collection and management systems allowing for analyzable ASD outcomes data differ greatly between medical and educational systems (Barnard-Brak, 2019). While both systems have their limitations, the educational system has a relatively strong state level organization with federally mandated reporting practices. Compared to the education system, the medical system does not have ASD specific reporting requirements and data systems are typically medical system, not state, specific. Furthermore, data are often privately owned or only accessible through memorandums of agreement with medical systems, which can be quite costly (Wang et al., 2013).

2.4 Evaluating prevalence of ASD

Considering that there are two major classification systems in the U.S. for children with an ASD makes determining the prevalence of ASD challenging. The official ASD prevalence data for the U.S. is derived from data compiled from both DSM medical and IDEA education systems via the ADDM Network (Nevison et al., 2018). ADDM Network is an active surveillance program funded by the CDC to both tracks the prevalence and monitor characteristics of children living with ASD or other developmental disabilities during 2016 (CDC, 2020). The ADDM Network data collects from health, education, and other service provider records of children from 11 sites across the U.S. using the same methods, which are modeled by CDC's Metropolitan Atlanta Developmental Disabilities Surveillance Program (MADDSP) (CDC, 2020). Children are included in the ADDM Network using a standardized ASD surveillance case definition with DSM-V criteria that is applied to all available records by their research team to ensure that cases have a high likelihood of clinical ASD status (Shaw, 2020). In contrast, the IDEA systems relied on each state's Department of Education to compile the prevalence and associated characteristics of ASD data from early childhood into adulthood (Mandell & Palmer, 2005). On behalf of the IDEA, local and state education systems have tracked children ages 3 to 21 years that receive special education services (Nevison & Zahorodny, 2019). The Department of Education reports annually to Congress on implementing the prevalence and education-system characteristics of ASD from IDEA for each state by academic year (Guerin, 2004).

2.5 Application of ASD prevalence

CDC ASD prevalence data is historically provided to highlight the central tendency and variability of ASD across ages, races, and socio-demographics in diverse communities (Maenner, 2020). The ADDM Network first began its ASD prevalence estimates in 2007. There are 11 ADDM sites[1] reporting both 4- and 8-year-olds' prevalence of ASD from the year 2018. The early ADDM Network tracks prevalence and monitors early identification of ASD among four-year-olds, which benefits the understanding of the characteristics of children with ASD. The early identification among younger ASD children leads to a previous diagnosis and earlier ASD intervention and causes a better developmental outcome (Shaw, 2020). In 2016, the overall ASD

[1] Arizona, Arkansas, Colorado, Georgia, Maryland, Minnesota, Missouri, New Jersey, North Carolina, Tennessee, Wisconsin

prevalence of 6 sites[2] was 15.6 per 1,000 (one in 64) children aged four years for Early ADDM Network sites. Prevalence varied state by state; for instance, Missouri recorded 8.8 per 1,000 while New Jersey recorded 25.3 per 1,000 (Shaw, 2020). The ADDM Network set 8 year-olds as a "constant-age tracking" method to estimate the prevalence of ASD by assuming the most children are diagnosed with ASD by about eight years old (Nevison et al., 2018). For 2016, the average ASD prevalence among children aged eight years across 11[3] sites was 18.5 per 1,000 (one in 54) children aged eight years, in which ASD prevalence varied by location, ranging from 13.1 (Colorado) to 31.4 (New Jersey) (Maenner, 2020). Based on the Morbidity and Mortality Weekly Report (MMWR) of ASD, the prevalence of ASD has a higher rate among boys than among girls at every site. Further, these studies showed no overall difference in ASD prevalence between black and white children aged eight; the disparity in ASD prevalence among aged four has decreased between white and black children. However, Hispanic children continue to be identified as having ASD less frequently than white or black children (Shaw, 2020; Maenner, 2020).

Children with ASD need various services in both the health care and education systems to meet their developmental needs (Bilaver et al., 2016). The use of prevalence data of ASD in public health promotes early and equitable identification of ASD and timely enrollment in services across race and socioeconomic groups (Maenner, 2020). Theoretically, the prevalence of ASD can be used to allocate the resources on education spending associated with the special education service, evaluate student-teacher ratio, and build school-based health services centers (Guerin, 2004). However, states and counties differ in eligibility criteria, policies, and

[2] Arizona, Colorado, Missouri, New Jersey, North Carolina, Wisconsin
[3] Arizona, Arkansas, Colorado, Georgia, Maryland, Minnesota, Missouri, New Jersey, North Carolina, Tennessee, Wisconsin

procedures regarding identifying ASD at the school-level; these differences are likely to impact the found prevalence and age of identification for children with ASD across different sites (Barton et al., 2016). A recent evaluation of publications related to autism in school psychology journals found a lack of publications related to the development and psychometric properties of assessments for the educational placement of students with (Mckenney et al., 2015). Even though there was a significant increase in the number of students identified under the IDEA ASD category over the past twenty years, many research findings indicate that a large number of students with ASD are under-identified or misclassified under other special education categories. This suggests a great deal of variability in the prevalence estimates across school systems that may impact the provision of services (Barton et al., 2016).

It is known that the estimated prevalence of ASD differs between clinically informed and education only identification approaches. For example, Barnard-Brake et al. (2019) found that educational eligibility categories differ from CDC clinically informed ASD case ascertainment approaches; prevalence rate estimates of clinical diagnoses of ASD were systematically higher than educational eligibility categories due to the distinction of ASD academic eligibility criteria and policies across the states for special education needed by IDEA (Sullivan, 2013), though there is some variation in relation to gender and race (Barnard-Brak, 2019). Variance across state educational systems in ASD categorizations likely reflect local educational and health services policies for children with ASD. In the long run, a challenge for the field is to determine whether the prevalence variance is due to methodological factors, diagnostic systems, or real differences in population parameters (Fombonne, 2018).

To date, educational eligibility and CDC prevalence estimate (research exist) in parallel siloes, with occasional cross-consideration to inform research (Barnard-Brak, 2019). However, both educational and CDC estimates indicate wide variance across states in terms of numbers of children identified and associations with socio-demographic and socio-economic factors. To date, researchers have not investigated the degree to which school ASD eligibility prevalence relates to CDC prevalence estimates. This thesis seeks to develop a county level metric indicating if the number of children served in special education with an ASD is similar to the number expected from ASD population estimates at the county and state level. Additionally, we want to understand whether school reported ASD prevalence is more similar to CDC estimates in some areas, lower in others, and what are the primary drivers of similarity/differences.

Chapter 3

METHODS

3.1 Data Source and Preparation

The data used in this study were from two sources. The dependent variable was calculated based on Idaho (ID), Mississippi (MS), and California (CA) state departments of education’s 2018-2019 school year special education database. All predictor variables used in this study came from the 2020 County Health Rankings annual report. The County Health Rankings is a program aiming to understand nearly every county's health situation in all US states (*2020 County Health Rankings Key Findings Report*, 2020). The health among community measures shows the inequality, disparities, and challenges in community groups. The Robert Wood Johnson Foundation collaborates with the University of Wisconsin Population Health Institute to deliver this program to communities across the nation since 2010. The County Health Rankings evaluated the health of communities by examining five key factors that influence the health of counties: health outcomes, health behaviors, clinical care, social and economic factors, and the physical environment. Based on the 2020 annual report findings, social and economic factors are related to community level health outcomes more than any other predictors (*2020 County Health Rankings Key Findings Report*, 2020.). The predictors of the study were selected from three factors: health outcome, clinical care, social and economic factors.

3.2 Outcome Variable

Potential percentage missing of ASD. The most recent ASD estimate (1/54) was used to determine the percentage of children with ASD potentially missed for each county. First, we took the total N (n_{total}) for each county and developed a metric of expected N with ASD ($n_{expected}$) based on an assumption of 1 in 54. The dependent variable was calculated based on ID, MS, and

CA state departments education's 2018-2019 school year county-level autism data. The state departments of education provided the actual autism value of children whose ages are from 5-18 years old at the county-level. Second, potential miss rate is calculated from the national prevalence of ASD. Expected autism cases were calculated based on the total population of children aged from 5-18 in each county's school system and CDC's estimate of the proportion of children (1/54=1.9%) who were diagnosed with ASD. (Formula: $Expected\ autism\ cases\ = \ the\ population\ aged\ 5-18\ in\ each\ state\ county\ *\ 1.9\%\ CDC\ autism\ prevalence$). The numerator of *potential percentage missing of ASD* was the difference between numbers of the expected autism cases and the actual autism cases reported from special education data, and the denominator is the expected autism cases ($Formula: Potential\ percent\ of\ missing\ ASD = \frac{Expected\ ASD\ case - Actual\ ASD\ cases}{Expected\ ASD\ cases}$*100). The results of percentage missing of ASD can be positive or negative. The negative value means that expected number of autism cases were less than the actual number of autism, which indicates that the county special education system identifies more autism cases than the CDC estimated ASD case number. On the other hand, the positive missing ASD value means that the CDC's estimation is greater than the actual number of observed autism cases. When the absolute value of percentage missing of ASD close to zero, the number of ASD cases that the department of education identify were close to the approximation of expected ASD in a school system per CDC estimates. A bigger absolute value indicates that the observed and expected ASD values are more different.

3.3 Covariates

This study aims to assess the relationship between potential missed autism cases and county-level sociodemographic data from three states: ID, MS, and CA. The study consisted of

184 records. ID has 44 (23.9%) counties, MS has 82 (44.6%) counties, and CA has 58 (31.5%) counties. The interest variables of three key factors and years of data included as follows:

Health Outcomes

Child mortality. The child mortality rate has a large influence on years of potential life lost (YPLL). The child mortality rate represents the number of deaths among children under age 18 within a county per every 100,000 residents. Data on deaths were provided by National Center for Health Statistics (NCHS) mortality files, 2015-2018.

Clinical Care

Primary care physicians. Primary care physicians were defined as M.D.s. and D.O.s, and obstetrics/gynecology were removed as a primary care physician type (*2020 County Health Rankings Key Findings Report*, 2020). Primary care is the sustenance of the healthcare system, but the healthcare providers vary greatly across states or within states (*2020 County Health Rankings Key Findings Report*, 2020). The *Primary care physicians* represent the number of primary care physicians within a county per every 100,000 residents. The data source is from the American Medical Association (AMA) area health resource file, 2017. For variable primary care physicians, 9 of 184 records are zero. All zeros were re-evaluated with one divided by each county's population, which allows for a rank ordering of counties, keeping very small values close but not equal to zero.

Mental health providers. Mental health providers are defined as psychiatrists, psychologists, licensed clinical social workers, counselors, marriage and family therapists, and mental health providers that treat alcohol and other drug abuse, as well as advanced practice nurses specializing in mental health care (*2020 County Health Rankings Key Findings Report*, 2020). Mental health providers represent the number of mental health providers within a county per every 100,000 residents. The data source is from the Centers for Medicare & Medicaid Services (CMS), national provider identification file, year of 2019. For variable mental health providers, 10 of 184 records are zero. All zeros were re-evaluated to handle the zero values, with one divided by each county's population, which treats zeros as small values but not equal to zero.

Uninsured children. Uninsured children are included in the data because those children are less likely to receive preventive care on time (Murphey, 2017). Uninsured children represent the number of children under age 19 without health insurance coverage within a county per every 100,000 residents. The data source is from the Small Area Health Insurance Estimates (SAHIE), 2017.

Social and Economic Factors

High school graduation. Education is an important predictor of health. Graduating with a high school diploma is associated with health benefits when compared to those that earn a Graduate Equivalency Diploma (GED), where GED earners are about twice as likely to have worse self-reported health (Zajacova & Everett, 2014). Also, it is important to note that as rates of high school and college completion are increasing, there are decreasing race/ethnicity gaps in educational attainment in the last five years (Ma et al., 2016). High school graduation represents the ninth-grade cohort that graduates from high school in four years within a county per every 100,000 residents. The data source is from the Local Education Agency (school district) level from EDFacts data were used for all states, 2016-2017.

Children in poverty. Children in poverty capture an upstream measure of poverty that assesses both current and future health risks. Low-income children are susceptible to more frequent and severe chronic conditions and behavior disorders than children living in high-income households

(McCarty, 2016). Children in poverty represent the number of people under age 18 living in poverty within a county per every 100,000 residents. The data source is from SAHIE, 2018.

Median household income. Median household income is the income amount at which half of the households in a county earn more, and half of the households earn less. Median household income is a well-recognized indicator of income and poverty, which are in turn related to compromised physical and mental health (Galea et al., 2011). The data source is from SAHIE, 2018.

Black/White residential segregation. Racial/ethnic residential segregation refers to the degree to which two or more groups live separately from one another in a geographic area, in this case, between black and white county residents (*2020 County Health Rankings Key Findings Report*, 2020). The residential segregation index ranges from 0 to 100, zero indicates complete integration, 100 indicates complete segregation, which higher values indicate greater residential segregation. The range of this database is from 2.4 to 77.53. The data source is from the American Community Survey (ACS), 2014-2018.

Percent of the rural population. The Census Bureau's urban-rural classification is a delineation of geographic areas. The urban areas represent the densely developed territory and encompass residential, commercial, and other non-residential urban land uses. While rural encompasses all population, housing, and territory not included within an urban area (Bureau, 2021). The numerator of this index is the total number of residents who live in a rural area of a county. The denominator is the total residents who live in both urban and rural areas. Percent of rural population ranges from 0 to 100, and zero indicates all the residents live in urban, when the value over 50 means more people live in rural, 100 means complete rural. For this database, only CA San Francisco county has all residents live in the urban area. There are thirty-eight counties located in a complete rural area. All one-hundred were re-ordered from 99.72 to 100 by each county's population to handle the one hundred values. The data is from the 2010 Census of Population and Housing issued by July 2012.

3.4 Descriptive Statistics

All the response and predictors variables were treated as continuous variables. The means (*M*) and standard deviation (*SD*) for all interested continuous variables were displayed in **Table 1**. As shown in **Table 1**, of 184 counties, 15 counties did not have *potential missing percentage of ASD* due to missing actual observed ASD cases number. Forty-six counties lacked residential segregation between black and white, the missing value is reported for counties with a black population less than 100 in the time frame (*2020 County Health Rankings Key Findings Report*, 2020). The missing value of residential segregation of black and white in Idaho was the highest at 81.82% (36/44). Based on the report from U.S. Census Bureau, the percent of white in Idaho is around 93.00%, while black is only around 0.90%, which caused some data suppression. In contrast, the percent of black in CA is 6.50%, and MS is 37.80%. There are 49 counties lack child mortality; a missing value is informed for counties with fewer than 10 children deaths in the time frame (*2020 County Health Rankings Key Findings Report*, 2020).

In **Table 1**, the mean (*M*) value of the percentage of missing ASD is 28.45%, and CA (*M* =1.96) had the most closely observed ASD cases with expected ASD value. MS's (*M* = 45.01) and ID's (*M* = 33.71) actual observed autism cases were less than the CDC's estimation. Also, primary care physicians (*M* = 70.76) and mental health providers (*M* = 357.89) in CA are greater than the average mean value of the total, and both MS (M_{primary}=40.42, M_{Mental}=101.19) and ID (M_{primary}=56.33, M_{Mental}=121.56) have a close mean of primary care physicians and a lower mental health provider. MS (*M* = 32692.68) has a higher value of children in poverty compared to CA (*M* = 17996.49) and ID (*M* = 17504.55). For uninsured children, CA (*M* = 3321.86) has better insurance coverage than MS (*M* = 5335.13) and ID (*M* = 5712.23). Residential segregation between black and white residents in MS (*M* = 31.93) is lower than both CA (*M* = 50.38) and ID (*M* = 56.68). High school graduation rates in all three states (M_{CA}=82508.95, M_{MS}= 80798.82, M_{ID}=83712.68) are similar to each other. Child mortality in MS (*M* = 89.78) is the highest in all three states, ID (*M* = 41.47) is in the middle, and CA (*M* = 41.47) is the lowest. For median household income, CA (*M* = 67098.66) has the highest, MS (*M* = 39932.72) has the lowest value, and ID (*M* = 52215.75) has a similar value to the total average (*M* = 51433.14). In CA (M = 28.69), more residents live in urban areas, and the population in MS (*M* = 70.59) and ID (*M* = 61.41) is more likely located in rural areas.

Table 1.Characteristics of continuous variables from 2020 County Health Rankings Annual Report (N=184)

Variable Name	Total		CA (n=58)		MS (n=82)		ID (n=44)	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Missing of ASD, %	169	28.45 (31.05)	55	1.92 (28.23)	76	45.01 (22.89)	38	33.71 (22.72)
Child mortality	135	68.06 (30.54)	49	41.47 (11.37)	66	89.78 (26.25)	20	61.54 (19.67)
Primary care physicians	184	53.79 (30.92)	58	70.76 (30.71)	82	40.42 (25.28)	44	56.33 (29.62)
Mental health providers	184	186.98 (182.94)	58	357.89 (180.25)	82	101.19 (134.96)	44	121.56 (85.09)
Uninsured children	183	4798.71 (1568.77)	57	3321.86 (787.39)	82	5335.13 (917.99)	44	5712.23 (1930.76)
Children in poverty	183	24463.39 (10813.29)	57	17996.49 (6869.19)	82	32692.68 (9762.09)	44	17504.55 (4450.08)
Residential segregation – Black/White	138	39.92 (16.37)	49	50.38 (10.99)	81	31.93 (14.05)	8	56.68 (18.18)
High school graduation	183	82032.09 (953.37)	57	82508.95 (10724.78)	82	80798.82 (13649.90)	44	83712.68 (14040.89)
Median household income	184	51433.14 (17822.38)	58	67098.66 (20753.91)	82	39932.72 (8685.11)	44	52215.75 (7409.04)
Rural population, %	184	55.19 (33.19)	58	28.69 (28.89)	82	70.59 (25.90)	44	61.41 (29.75)

3.5 Correlation

Spearman correlations between county level predictor variables and ASD missing outcome

The Spearman correlations between county-level predictor variables and likely percentage missing of ASD in IDEA data were displayed in **Table 2**. Spearman correlation is applied here to measure the strength and direction of association between ordinal variables and evaluate whether the two ordinal variables vary together with another variable. The total value of each variable was found a statistically significant correlation with the outcome at an alpha level of 0.05. According to Cohen, the effect size is small when r_s varies around 0.1, medium if r_s varies around 0.3, and large if r_s varies more than 0.5. Child mortality (r_s = 0.57) and median household income (r_s= -0.52) have a large association with a percentage missing of ASD. High school graduation (r_s = -0.19) has a small strength of association with the outcome. The rest of the variables- primary care physicians (r_s = -0.33), mental health providers (r_s = -0.47), children in poverty (r_s = 0.44), uninsured children (r_s =0.48), residential segregation between black and white residents (r_s = -0.36), percentage of the rural population (r_s =0.44) all have a medium strength of association. Primary care physicians, mental health providers, residential segregation between black and white residents, high school graduation, median household income have a negative association with the likely percentage missing of ASD. When the value of those predictors increases, the likely percentage missing of ASD decreases. However, children in poverty, uninsured children, child mortality, and the percentage of the rural population have a positive association. As the value of those predictors increase, the likely percentage missing of ASD is also increasing.

The correlations vary considerably between predictors and the outcome in CA, ID, and MS. Some states have a negative association between predictors and the outcome, but others positively associate the same predictors. For instance, child mortality has a small negative association with the outcome in CA, but ID and MS present a positive association. CA has a small positive association for primary care physicians and mental health providers, while ID and MS have a small negative association. Both CA and MS have a small negative association for uninsured children, but ID has a medium positive association. For children in poverty, MS (r_s = 0.43) has a significant positive medium association with the outcome at the county level, but CA and ID have a small none significant association. For residential segregation between black and white, ID has a small negative association, but CA and MS barely have any association. For high school graduation, only MS (r_s = -0.44) has a significant negative medium association with the outcome, CA and ID have a small association. CA (r_s = 0.31) has a significant positive medium association with outcome, but MS (r_s = -0.44) has a significant negative medium association with outcome, and ID has no association. CA (r_s = -0.33) has a significant negative medium association with the outcome for the percentage of the rural population, but ID and MS have a non-significant small positive association. The different direction and strength association between county-level predictor variables and likely percentage missing of ASD discovered the difference and complexity among states in health outcomes, access to clinical care, education, income, and family and social support.

Table 2. Spearman correlations between county level predictor variables and likely percentage missing of ASD in IDEA data

	Total	CA	ID	MS
Child mortality	**0.57*** **	-0.11	0.38	0.08
Primary care physicians	**-0.33*** **	0.16	-0.21	-0.06
Mental health providers	**-0.47*** **	0.22	-0.25	-0.16
Uninsured children	**0.48*** **	-0.19	0.30	-0.02
Children in poverty	**0.44*** **	-0.15	-0.05	0.43***
Residential segregation – Black/White	**-0.36*** **	0.00	-0.19	0.09
High school graduation	**-0.19* **	-0.10	0.10	-0.44**
Median household income	**-0.52*** **	0.31*	0.03	-0.44***
Rural population, %	**0.44*** **	-0.33*	0.30	0.16

Note. * Indicates statistical significance at an alpha level of .05; 0.001 '***' 0.01 '**' 0.05 '*'

Spearman correlations between the independent variables with each other

Combined States. The Spearman correlations between the independent variables with each other were displayed in **Table 3**. The correlation matrix shows that most of the predictors highly correlate with each other except high school graduation. For primary care physicians, the correlation of *r* suggests a large positive association with mental health providers (r_s =0.67), residential segregation between black and white (r_s =0.54), and median household income (r_s =0.51); whereas a medium negative association with children in poverty (r_s = -0.47), uninsured children (r_s = -0.43); and a large negative association with child mortality (r_s = -0.56) and percentage of the rural population (r_s = -0.56) in one county. For mental health providers, both residential segregation between black and white (r_s =0.61) and median household income (r_s =0.55) have a large positive association. Children in poverty (r_s = -0.44), uninsured child (r_s = -0.63), child mortality (r_s = -0.57) and percentage of the rural population (r_s = -0.67) have a medium to large negative association. Children in poverty has a large negative association with median household income (r_s = -0.90) but has a large positive association with child mortality (r_s =0.76), and a medium positive association with uninsured children (r_s =0.30) and percentage of the rural population (r_s =0.44), also a negative medium association with residential segregation (r_s = -0.48) and high school graduation (r_s = -0.33). Uninsured children have a large positive association with child mortality (r_s =0.64) and percentage of the rural population (r_s =0.72); medium negative association with residential segregation (r_s = -0.5) and median household income (r_s = -0.45). Residential segregation between black and white has the same strength but different direction with median household income (r_s =0.53), child mortality (r_s = -0.53), and percentage of the rural population (r_s = -0.53). Median household income has a large negative association with child mortality (r_s = -0.86) and percentage of the rural population (r_s = -0.58). Child mortality has a large positive association with percentage of the rural population (r_s =0.73).

Table 4 displayed the Spearman correlations between independent variables with each other in CA. Based on Cohen, the significant positive association between independent variables include primary care physicians with mental health providers (r_s =0.64), primary care physicians with median household income (r_s =0.65), children in poverty with child mortality (r_s =0.73), and percentage of the rural population with child mortality (r_s = 0.58). The significant large negative association between independent variables include primary care physicians with children in poverty (r_s = -0.62), primary care physicians with child mortality (r_s = -0.66), median house income with children in poverty (r_s = -0.85), median household income with child mortality (r_s = -0.81), and median household income with percentage of the rural population (r_s = -0.69). The significant medium positive association between independent variables include children in poverty with percentage of the rural population (r_s = 0.37), uninsured children with percentage of the rural population (r_s = 0.46), and residential segregation between black and white with percentage of the rural population (r_s = 0.40). The significant medium negative association between independent variables include: primary care physicians with percentage of the rural population (r_s = -0.43), and mental health providers with child mortality (r_s = -0.41). The significant small positive association between independent variables include mental health providers with median household income (r_s = 0.29), and with residential segregation between black and white (r_s = 0.26). The significant small negative association between independent variables include children in poverty and mental health providers (r_s = -0.29) and median household income with uninsured children (r_s = -0.26).

Table 5 displayed the Spearman correlations between independent variables with each other in ID. Based on Cohen, the significant large positive association between independent variables include primary care physicians with mental health providers (r_s = 0.62), uninsured children with child mortality (r_s = 0.68), percentage of the rural population with uninsured children (r_s = 0.57), and child mortality with percentage of the rural population (r_s = 0.68). The significant large negative association between independent variables include mental health providers with child mortality (r_s = -0.52), percentage of the rural population with mental health providers (r_s = -0.51), children in poverty with median household income (r_s = -0.84). The significant medium positive association between independent variables include children in poverty with percentage of the rural population (r_s = 0.30). The significant medium negative association between independent variables include primary care physicians with child mortality (r_s = -0.48) and mental health providers with uninsured children (r_s = -0.43). There is no significant small positive or negative association between independent variables.

Table 6 displayed the Spearman correlations between independent variables with each other in MS. Based on Cohen, the significant large positive association between independent variables include primary care physicians with mental health providers (r_s =0.53), children in poverty with child mortality (r_s =0.58), and percentage of the rural population with uninsured children (r_s = 0.59). The significant large negative association between independent variables include primary care physicians with percentage of the rural population (r_s = -0.56), mental health providers with percentage of the rural population (r_s = -0.61), median household income with children in poverty (r_s = -0.89), median household income with child mortality (r_s = -0.66). The significant medium positive association between independent variables include residential segregation between black and white with primary care physicians (r_s =0.41), and with mental health (r_s = 0.45), also high school graduation with median household income (r_s = 0.42). The significant medium negative association between independent variables include uninsured children with primary care physicians (r_s = -0.33), with mental health providers (r_s = -0.40), and with children in poverty (r_s = -0.32), residential segregation between black and white with percentage of the rural population (r_s = -0.41), and high school graduation with children in poverty (r_s = -0.45). The significant small positive association between independent variables include median household income with primary care physicians (r_s = 0.08), and with residential segregation between black and white (r_s = 0.27), uninsured children with high school graduation (r_s = 0.06), child mortality with percentage of the rural population (r_s = 0.27). The significant small negative association between independent variables include primary care physicians with child mortality (r_s = -0.21) and median household income with percentage of the rural population (r_s = -0.17).

Table 3. Spearman correlations between the independent variables with each other in 2020 County Health Rankings annual report

	Primary care physicians	Mental health providers	Children in poverty	Uninsured children	Residential segregation Black/White	High school graduation	Median household income	Child mortality	Rural population, %
Primary care physicians	1								
Mental health providers	0.67***	1							
Children in poverty	-0.47***	-0.44***	1						
Uninsured children	-0.43***	-0.63***	0.30**	1					
Residential segregation-Black/White	0.54***	0.61***	-0.48**	-0.5***	1				
High school graduation	-0.01	-0.33	-0.33*	-0.04	0.04	1			
Median household income	0.51***	0.55***	-0.90***	-0.45***	0.53***	0.28	1		
Child mortality	-0.56***	-0.57***	0.76***	0.64***	-0.53***	-0.26	-0.86***	1	
Rural population, %	-0.56***	-0.67***	0.44***	0.72***	-0.53***	-0.03	-0.58***	0.73***	1

Note. * Indicates statistical significance at an alpha level of .05; 0.001 '***' 0.01 '**' 0.05 '*'

Table 4. Spearman correlations between independent variables with each other in 2020 County Health Rankings annual report in CA

	Primary care physicians	Mental health providers	Children in poverty	Uninsured children	Residential segregation Black/White	High school graduation	Median household income	Child mortality	Rural population, %
Primary care physicians	1								
Mental health providers	0.64***	1							
Children in poverty	-0.62***	-0.29*	1						
Uninsured children	-0.16	0.01	0.09	1					
Residential segregation-Black/White	0.16	0.26*	-0.03	0.25	1				
High school graduation	-0.01	-0.30	-0.32	-0.06	-0.02	1			
Median household income	0.65***	0.29**	-0.85***	-0.26*	-0.13	0.28	1		
Child mortality	-0.66***	-0.41**	0.73***	0.09	0.02	-0.28	-0.81***	1	
Rural population, %	-0.43***	-0.21	0.37*	0.46***	0.40**	-0.05	-0.69***	0.58***	1

Note. * Indicates statistical significance at an alpha level of .05; 0.001 '***' 0.01 '**' 0.05 '*'

Table 5. Spearman correlations between the predictors with each other in 2020 County Health Rankings annual report in ID

	Primary care physicians	Mental health providers	Children in poverty	Uninsured children	Residential segregation Black/White	High school graduation	Median household income	Child mortality	Rural population, %
Primary care physicians	1								
Mental health providers	0.62***	1							
Children in poverty	-0.26	-0.30	1						
Uninsured children	-0.24	-0.43*	0.36	1					
Residential segregation-Black/White	-0.31	0.19	-0.13	0.05	1				
High school graduation	-0.21	-0.20	0.13	0.07	-0.17	1			
Median household income	0.05	0.11	-0.84***	-0.17	0.48	-0.19	1		
Child mortality	-0.48*	-0.52*	0.50	0.68**	0.04	-0.15	-0.35	1	
Rural population, %	-0.23	-0.51***	0.30*	0.57***	-0.12	0.20	-0.13	0.68***	1

Note. * Indicates statistical significance at an alpha level of .05; 0.001 '***' 0.01 '**' 0.05 '*'

Table 6. Spearman correlations between independent variables with each other in 2020 County Health Rankings annual report in MS

	Primary care physicians	Mental health providers	Children in poverty	Uninsured children	Residential segregation Black/White	High school graduation	Median household income	Child mortality	Rural population, %
Primary care physicians	1								
Mental health providers	0.53***	1							
Children in poverty	-0.02	-0.05	1						
Uninsured children	-0.33*	-0.40**	-0.32*	1					
Residential segregation-Black/White	0.41***	0.45**	-0.17	-0.17	1				
High school graduation	-0.07	-0.06	-0.45*	0.06*	-0.04	1			
Median household income	0.08**	0.11	-0.89***	0.25	0.27*	0.42*	1		
Child mortality	-0.21*	-0.09	0.58***	-0.08	-0.33**	-0.17	-0.66***	1	
Rural population, %	-0.56***	-0.61***	0.00	0.59***	-0.41***	0.01	-0.17**	0.27**	1

Note. * Indicates statistical significance at an alpha level of .05; 0.001 '***' 0.01 '**' 0.05 '*'

3.6 Statistical Analysis

To understand if there is an association between county-level predictor variables and likely percentage missing of ASD in IDEA data, a series of linear regression was conducted using the 2020 County Health Rankings annual report and ASD IDEA data. The study starts with a simple linear regression by modeling the *likely percentage missing of ASD* in IDEA data given one predictor of each socio-demographic covariates (**Figure 1**). Model 1 was used to study the association between the *percentage missing of ASD* and each independent variable from the entire three states. The intercept is the average *percentage missing of ASD* for each continuous variable with a value of zero. The slope gives the expected change of outcome for each unit change of predictor. Then multiple linear regressions were run to predict the percentage missing of ASD from each socio-demographic covariates and state. Model 2 forced the slope of the socio-demographic variable to be the same for different states. Model 3 included interactions between states and the socio-demographic variable, which allows the slopes to vary across different states. The coefficient of the interaction term in model 3 represents the difference in the slope of the socio-demographic variable for one state as compared to CA, the reference state. All data cleaning and analyses for this study were done in RStudio with R version 4.0.3.

Figure 1. Detail on the Function

Model 1- Bivariate Regression Models $Y = \beta_0 + \beta_1 X + \epsilon$ with $\epsilon \sim N(0,\sigma^2)$, X = independent variable.
Model 2- Regression Models with Continuous and State Predictors $Y = \beta_0 + \beta_1 X + \beta_2 S_{ID} + \beta_3 S_{MS} + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$, $S_{ID} = 1\ for\ ID\ and\ 0\ for\ others, S_{MS} = 1\ for\ MS\ and\ 0\ for\ other\ states.$
Model 3- Regression Models with Interactions Between Continuous and State Predictors $Y = \beta_0 + \beta_1 X + \beta_2 S_{ID} + \beta_3 S_{MS} + \beta_4 X S_{ID} + \beta_5 X S_{MS} + \epsilon$ with$\epsilon \sim N(0, \sigma^2)$
Model 4- Random-intercept models $$Y_{ij} = \beta_{i0} + \beta_1 X_{ij} + \epsilon_{ij}$$ where i=1,…,3 index states j=1,… n_i index counties for state i random intercept $\beta_{i0} \sim N(\beta_0, \sigma_s^2)$, random error $\epsilon_{ij} \sim N(0, \sigma^2)$
Model 5- Random-slope models $$Y_{ij} = \beta_{i0} + \beta_{i1} X_{ij} + \epsilon_{ij}$$ where i=1,…,3 index states j=1,… n_i index counties for state i random effects $\begin{pmatrix} \beta_{i0} \\ \beta_{i1} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right)$ random error $\epsilon_{ij} \sim N(0, \sigma^2)$

Chapter 4

RESULTS

4.1 Associations between likely percentage missing of ASD and relevant county level socio-demographic covariates using simple and multiple linear regression

Statistical analysis of this study was done using simple linear regressions and multiple linear regression (**Table 7**). Each socio-demographic predictor and *likely percentage missing of ASD* were modeled separately.

Bivariate Regression Models

Primary care physicians. In **Table 7**, model 1 found a significant association between primary care physicians and ASD missingness metric (p = .002). The parameter estimate for primary care physicians was -0.23. For every one more primary care physician, the likely percentage missing of ASD was estimated to decrease by 0.23. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.05.

Mental health providers. In **Table 7**, model 1 found a significant association between mental health providers and ASD missingness metric ($p < .001$). The parameter estimate for mental health providers was -0.05. For every one more mental health provider, the likely percentage missing of ASD was expected to decrease by 0.05. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.11.

Children in poverty. In **Table 7**, model 1 found a significant association between children in poverty and ASD missingness metric ($p < .001$). The parameter estimate for children in poverty providers was 0.001. For every one more child in poverty, the likely percentage missing of ASD was expected to increase by 0.001. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.16.

Uninsured children. In **Table 7**, model 1 found a significant association between uninsured children and ASD missingness metric ($p < .001$). The parameter estimate for uninsured children providers was 0.009. For every one more uninsured children, the likely percentage missing of ASD was expected to increase by 0.009. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.17.

Residential segregation between black and white. In **Table 7**, model 1 found a significant association between residential segregation between black and white and ASD missingness metric ($p < .001$). The parameter estimate for residential segregation between black and white providers was -0.66. For every one more residential segregation between black and white, the likely percentage missing of ASD was estimated to decrease by 0.66. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.12.

High school graduation. In **Table 7**, model 1 found an insignificant association between high school graduation and ASD missingness metric ($p = 0.11$). The parameter estimate for high school graduation providers was -0.0005. For every one-unit increase in high school graduation, the likely percentage missing of ASD was expected to decrease by 0.0005. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.01.

Median household income. In **Table 7**, model 1 found a significant association between median household income and ASD missingness metric ($p < .001$). The parameter estimate for median household income providers was -0.0006. For every one-unit increase in median household income, the likely percentage missing of ASD was expected to decrease by 0.0006. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.12.

Child mortality. In **Table 7**, model 1 found a significant association between child mortality and ASD missingness metric ($p < .001$). The parameter estimate for child mortality providers was

0.45. For every one more in child mortality, the likely percentage missing of ASD was estimated to increase by 0.44. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.21.

Percentage of the rural population. In **Table 7**, model 1 found a significant association between percentage of the rural population and ASD missingness metric ($p < .001$). The parameter estimate for percentage of the rural population providers was 0.31. For every one-unit increase in percentage of the rural population, the likely percentage missing of ASD was expected to increase by 0.31. The R^2 value showing the proportion of dependent variance explained by model 1 was 0.11.

Regression Models with Continuous and State Predictors

Primary care physicians. In **Table 7**, model 2 found an insignificant association between primary care physician metric ($p = 0.538$) and a significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.38.

Mental health providers. In **Table 7**, model 2 found an insignificant association between mental health provider metric ($p = 0.280$) and a significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.38.

Children in poverty. In **Table 7**, model 2 found an insignificant association between children in poverty metric ($p = 0.088$) and a significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.39.

Uninsured children. In **Table 7**, model 2 found an insignificant association between uninsured children metric ($p = 0.635$) and a significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.38.

Residential segregation between black and white. In **Table 7**, model 2 found an insignificant association between residential segregation between black and white metric ($p = 0.779$) and significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.41.

High school graduation. In **Table 7**, model 2 found a significant association between high school graduation metric ($p = 0.01$) and a significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.40.

Median household income. In **Table 7**, model 2 found an insignificant association between median household income metric ($p = 0.328$) and a significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.38.

Child mortality. In **Table 7**, model 2 found an insignificant association between child mortality metric ($p = 0.491$) and a significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.36.

Percentage of the rural population. In **Table 7**, model 2 found an insignificant association between percentage of the rural population metric ($p = 0.526$) and a significant effect of state wherein ID and MS had higher levels of ASD missingness compared to CA (both $p < .001$). The R^2 value showing the proportion of dependent variance explained by model 2 was 0.38.

Regression Models with Interactions Between Continuous and State Predictors

Primary care physicians. In **Table 7**, model 3 found that the parameter estimate for primary care physicians was 0.24 in CA ($p = 0.03$). For every one-unit increase in primary care physicians in CA, the likely percentage missing of ASD increased by 0.24. The interaction terms indicate that compared to that for CA, the slope of primary care physicians significantly decreased by 0.31 for MS and non-significantly decreased by .33 for ID. When the number of primary care physicians was zero, ID and MS have greater rates of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.40.

Mental health providers. In **Table 7**, model 3 found that the parameter estimate for mental health providers was 0.04 in CA ($p = 0.02$). For every one-unit increase in mental health providers in CA, the likely percentage missing of ASD increased by 0.04. The interaction terms indicate that compared to that for CA, the slope of mental health providers non-significantly decreased by 0.05 for MS and non-significantly decreased by 0.08 for ID. When the number of mental health providers was zero, ID and MS have greater rates of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.40.

Children in poverty. In **Table 7**, model 3 found that the parameter estimate for children in poverty was -0.00047 in CA ($p = 0.32$). For every one-unit increase in children in poverty in CA, the likely percentage missing of ASD decreased by 0.00047. The interaction terms indicate that compared to that for CA, the slope of children in poverty significantly increased by 0.00132 for MS and non-significantly increased by 0.00012 for ID. When the number of children in poverty was zero, ID and MS have greater rates of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.41.

Uninsured children. In **Table 7**, model 3 found that the parameter estimate for uninsured children providers was -0.009 in CA (p = 0.04). For every one-unit increase in uninsured children in CA, the likely percentage missing of ASD decreased by 0.009. The interaction terms indicate that compared to that for CA, the slope of uninsured children significantly increased by 0.0109 for MS and significantly increased by 0.0128 for ID. When the number of uninsured children was zero, ID has a lower rate of missingness than CA, and MS has a greater rate of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.40.

Residential segregation between black and white. In **Table 7**, model 3 found that the parameter estimates for residential segregation between black and white was -0.22 in CA (p = 0.47). For every one-unit increase in residential segregation between black and white in CA, the likely percentage missing of ASD decreased by 0.22. The interaction terms indicate that compared to that for CA, the slope of residential segregation between black and white non-significantly increased by 0.28 for MS and non-significantly decreased by 0.0045 for ID. When the number of residential segregation between black and white was zero, MS and ID have greater rates of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.41.

High school graduation. In **Table 7**, model 3 found that the parameter estimate for high school graduation was -0.0005 in CA (p = 0.09). For every one-unit increase in high school graduation in CA, the likely percentage missing of ASD decreased by -0.0005. The interaction terms indicate that compared to that for CA, the slope of high school graduation non-significantly decreased by 0.0011 for MS and non-significantly increased by 0.0005 for ID. When the high school graduation was zero, MS has a greater rate of missingness than CA, but ID has a lower rate of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.41.

Median household income. In **Table 7**, model 3 found that the parameter estimate for median household income was 0.0004 in CA ($p = 0.01$). For every one-unit increase in median household income in CA, the likely percentage missing of ASD increased by 0.0004. The interaction terms indicate that compared to that for CA, the slope of median household income significantly decreased by 0.0014 for MS but non-significantly decreased by 0.0002 for ID. When the median household income was zero, MS and ID have greater rates of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.43.

Child mortality. In **Table 7**, model 3 found that the parameter estimate for child mortality was 0.44 in CA ($p = 0.18$). For every one-unit increase in child mortality in CA, the likely percentage missing of ASD decreased by 0.44. The interaction terms indicate that compared to that for CA, the slope of child mortality non-significantly increased by 0.52 for MS and significantly increased by 0.86 for ID. When the number of child mortality was zero, MS has a greater rate of missingness than CA, but ID has a lower rate of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.38.

Percentage of the rural population. In **Table 7**, model 3 found that the parameter estimate for percentage of the rural population was -0.57 in CA ($p < 0.001$). For every one-unit increase in percentage of the rural population in CA, the likely percentage missing of ASD decreased by 0.57. The interaction terms indicate that compared to that for CA, the slope of percentage of the rural population significantly increased by 0.75 for both MS and ID. When the percentage of the rural population was zero, MS and ID have greater rates of missingness than CA. The R^2 value showing the proportion of dependent variance explained by model 3 was 0.46.

Table 7. Comparison of linear regression models between bivariate regression models (model 1), regression models with continuous and state predictors (model 2) and regression models with interactions between continuous and state predictors (model 3)

	Model 1	Model 2	Model 3
Intercept[a]	40.92 (4.66)***	-1.12 (5.95)	-15.14 (8.49)
Continuous Predictor			
Primary care physicians	-0.23 (0.08)**	0.05 (0.07)	0.24 (0.11) *
State			
ID		32.54 (5.36)***	53.99 (12.08)***
MS		44.42 (4.89)***	63.08 (9.99)***
CA		Reference	Reference
Continuous*State			
Primary care physicians ID			-0.33 (0.18)
Primary care physicians MS			-0.31 (0.15)*
Primary care physicians CA			Reference
AIC	1636.45	1570.29	1568.93
R^2	0.05	0.38	0.40
Intercept[b]	38.87 (3.28)***	-3.16 (5.76)	-12.89 (7.27)
Continuous Predictor			
Mental health providers	-0.05 (0.01)***	0.01 (0.01)	0.04 (0.02)*
State			
ID		35.06 (6.02)***	51.61 (10.10)***
MS		46.66 (5.47)***	58.93 (8.09)***
CA		Reference	Reference
Continuous*State			
Mental health providers ID			-0.08 (0.05)
Mental health providers MS			-0.05 (0.03)
Mental health providers CA			Reference
AIC	1627.31	1569.48	1568.37
R^2	0.11	0.38	0.40
Intercept[c]	0.31 (5.49)	-5.56 (5.48)	10.36 (9.12)
Continuous Predictor			
Children in poverty	0.001 (0.0002)***	0.0004 (0.0002)	-0.0005 (-0.0005)
State			
ID		32.02 (5.19)***	29.37 (17.95)
MS		37.18(5.55)***	7.46 (13.41)
CA		Reference	Reference
Continuous*State			
Children in poverty ID			0.0001 (0.001)
Children in poverty MS			0.001 (0.0006)*
Children in poverty CA			Reference
AIC	1616.83	1567.69	1565.16
R^2	0.16	0.39	0.41

	Model 1	Model 2	Model 3
Intercept[d]	-14.53 (7.59)	-0.98 (6.96)	31.92 (14.52)*
Continuous Predictor			
Uninsured children	0.009 (0.002)***	0.0009(0.002)	-0.009 (0.004)*
State			
ID		29.94 (6.51)***	-18.94 (20.47)
MS		41.37 (5.68)***	3.23 (22.59)
CA		Reference	Reference
Continuous*State			
Uninsured children ID			0.013(0.005)*
Uninsured children MS			0.011 (0.005)*
Uninsured children CA			Reference
AIC	1613.78	1570.45	1567.51
R^2	0.17	0.38	0.40
Intercept[e]	56.06 (6.82)***	6.53 (8.63)	15.57 (16.03)
Continuous Predictor			
Residential segregation – Black/White [e]	-0.66 (0.16)***	-0.04 (0.16)	-0.22 (0.31)
State			
ID		23.79 (9.05)**	25.18 (33.27)
MS		39.94 (5.12)***	27.49 (17.53)
CA		Reference	Reference
Continuous*State			
Residential segregation – Black/White ID			-0.005 (0.58)
Residential segregation – Black/White MS			0.28(0.37)
Residential segregation – Black/White CA			Reference
AIC	1272.48	1224.01	1227.23
R^2	0.12	0.41	0.41
Intercept[f]	70.85 (26.84)**	56.81 (21.37)**	44.10 (25.10)
Continuous Predictor			
High school graduation [f]	-0.0005 (0.0003)	-0.0007(0.0003)*	-0.0005 (0.0003)
State			
ID		34.08 (5.20)***	-6.86 (73.75)
MS		43.39 (4.30)***	133.00 (55.20)*
CA		Reference	Reference
Continuous*State			
High school graduation ID			0.0005 (0.0009)
High school graduation MS			-0.0012 (0.0007)
High school graduation CA			Reference
AIC	1643.34	1563.89	1564.49
R^2	0.01	0.40	0.41

	Model 1	Model 2	Model 3
Intercept[g]	59.36 (6.82)***	-7.28 (9.95)	-24.35 (10.88)*
Continuous Predictor			
Median household income [g]	-0.0006 (0.0001)***	0.0001 (0.0001)	0.0004 (0.0002)*
State			
ID		33.87 (5.63)***	46.76 (28.45)
MS		46.79 (5.78)***	109.00 (17.10)***
CA			Reference
Continuous*State			
Median household income ID			-0.0002 (0.0005)
Median household income MS			-0.0014 (0.0004)***
Median household income CA			Reference
AIC	1624.03	1569.69	1558.76
R^2	0.12	0.38	0.43
Intercept[h]	-2.47 (5.61)	2.32 (5.41)	23.17 (13.75)
Continuous Predictor			
Child mortality [h]	0.45 (0.08)***	0.07 (0.10)	-0.44 (0.33)
State			
ID		25.18 (6.83)***	-16.79 (22.59)
MS		34.93 (6.71)***	13.64 (17.38)
CA		Reference	Reference
Continuous*State			
Child mortality ID			0.86 (0.43)*
Child mortality MS			0.52 (0.35)
Child mortality CA			Reference
AIC	1242.56	1219.67	1219.55
R^2	0.21	0.36	0.38
Intercept[I]	12.14 (4.29)**	3.09 (3.82)	16.39 (4.42)***
Continuous Predictor			
Rural population, % [I]	0.31 (0.07)***	-0.05 (0.07)	-0.57 (0.12)***
State			
ID		33.29 (5.73)***	6.82 (9.49)
MS		45.08 (5.39)***	16.11 (8.83)
CA		Reference	Reference
Continuous*State			
Rural population ID			0.75 (0.18)***
Rural population MS			0.75 (0.16)***
Rural population CA			Reference
AIC	1626.74	1570.26	1549.37
R^2	0.11	0.38	0.46

Note. * Indicates statistical significance at an alpha level of .05; 0.001 '***' 0.01 '**' 0.05 '*'
AIC= Akaike An Information Criterion

[a] Model estimates for primary care physicians per 100,000 (2017)
[b] Model estimates for mental health providers per 100,000 (2019)
[c] Model estimates for children in poverty per 100,000 (2018)
[d] Model estimates for uninsured children per 100,000 (2017)
[e] Model estimates for residential segregation rate between Black and White (2014-2018)
[f] Model estimates for high school graduation per 100,000 (2016-2017)
[g] Model estimates for median household income (2018)
[h] Model estimates for child mortality per 100,000 (2015-2018)
[I] Model estimates for ratio of rural to urban population based on Census Population Estimates (2010)

Chapter 5

DISCUSSION

5.1 Discussion of Results

This study provides an initial analysis into creating a potentially missed ASD case metric that might be useful as a sensitive measure to evaluate if the prevalence of ASD eligibility status in school systems is similar to CDC estimates. The study results showed that the potential missed ASD case metric correlates with relevant socio-demographic covariates, which the literature indicates are associated with ASD prevalence. These data can be taken as initial validity data showing that the potentially missed case metric relates, but is not equivalent, to relevant co-variates. Several socio-demographic covariates, such as uninsured children and percentage of the rural population, are most predictive of potentially missed ASD cases.

The current study also found that the potential missed ASD case metrics is likely widely variable across the states as CA, ID, and MS were substantively different from one another. For example, the total observed ASD cases from the CA state education system are, on average, very close to the CDC's expected autism cases. However, in ID, the actual number of autism is smaller than the CDC's estimation; in MS, the total missing ASD value is the highest among the three states. This provides concern that each state will have possible effects that influence the potential missing ASD cases **(Table 1).** In general, all continuous independent variables were found to be associated with the likely percentage missing of ASD in initial bivariate regression models (Model 1), which suggests that applying the states' combined data showed a significant association between the predictor variables with the potential missing of ASD in general. Adding state predictors as a fixed effect in model 2 allows for estimation of each state's impact. However, most of the predictors are not significant after including states as predictors, suggesting that individual states are a major source of variance in *potentially missed ASD cases.*

Thus, a challenge moving forward is to search for which socio-demographic factors or socioeconomic factors might contribute to the *potential missing of ASD* within and between states. A county-level data within each state may help to identify the specific factors. By adding interactions between continuous predictors and states in model 3, the significant value of interaction effect exists when predictors on the likely percentage missing of ASD changes, depending on the states. In many cases, the relationship is zero. In others, it displays variance in predictor-outcome relationships that are quite disparate. Collectively, this suggests a challenge in examining the predictor relation to the likely percentage missing of ASD as the complex interactions lead to the outcome.

There are several significant differences between states with predictors of the relationship to missing ASD in model 3. For the predictors- uninsured children and rural population, the relationship between missing ASD is different in degree but in the same direction. ID and MS all have positive interaction terms for uninsured children. Compared to CA, the slope of uninsured children significantly increases quickly for ID than MS. Similarly, for rural population predictor, ID and MS have positive and close degrees on the effect of the missing ASD set with CA as reference. It suggests that for ID and MS, insurance coverage among children and the percentage of the rural population have a similar trend in the influence of the likely percentage missing of ASD. However, there is some evidence for the relationship of missing ASD and predictors which is significant in some states, but not in others. For instance, primary care physicians in MS have a significant negative interaction with ASD missingness, but in ID it is not significant. Similarly, the slope of children in poverty in MS is significantly higher compared to CA, but not in ID. The slope of median household income in MS is significantly lower in reference to CA, but not in ID. In addition, the slope of child mortality is significant greater compared to CA in ID, but not in

MS. This suggests that health outcome, clinical care coverage, and income are significantly different among the three states and may lead to a varied degree of missing ASD within those states.

Moreover, some predictors have clear opposite relationships on the effect of potentially missing ASD. For predictors high school graduation and residential segregation between black and white, even though the interaction terms are not significant, the direction of those two interactions are opposite. The effect of high school graduation in ID is positive, but MS is negative. In ID, the residential segregation of black and white is negative but positive in MS. In this database, the missing value of residential segregation of black and white in Idaho was the highest because the white population in Idaho is around 93.00%, which may cause the opposite relationships. To date, the different degree and even opposite direction of interaction between predictors and states conducted are that states have a complex effect on the socio-demographic and socioeconomic influence on the relationship between predictors and outcome (MacFarlane & Kanaya, 2009).

Several variables had differential relationships with the *Potential percentage missing of ASD* outcomes across ID, MS, and CA, indicating that researchers should be careful moving forward with analyses. Linear mixed models would be a better choice here since they can account for the correlations between data coming from county-level and state-level and avoid issues with multiple comparisons while using separate regressions. The linear mixed models allow for the estimate of fixed and random effects. The fixed effects can also be called explanatory variables, which are expected to affect the dependent variable. Random effects refer to groups (e.g., "nestings"), such as states or counties, that may influence the relationship between predictor variables and dependent variables (Bates et al., 2014). The random effect explains the total variance: how much variance among states, plus the residual variance, which aims to capture all the influence of states on dependent variables. In this case, the continuous variables are fixed effects, the state is a random effect, and the likely percentage missing of ASD is the dependent variable. The linear mixed-effects models can be determined using the lmer function in the lme4 package for R (Bates et al., 2014).

This study suggested a random slope model based on the previous analysis of the data. The mixed-effects models can assess the relationship between predictors and *potential missing ASD* from where the predictors were collected. The data used in this study contain county-level data from ID, MS, and CA. A random slope model allows each state line to have a different slope which means that the random slope model allows the predictor to have a different effect for each state. The following equation was used to fit the random slope model. $Y_{ij} = \beta_{i0} + \beta_{i1}X_{ij} + \epsilon_{ij},$ i indicates the states, and j indicates counties for state i, and counties are nested into states, β_{i0} indicates random intercept, β_{i1} is a random slope that changed by states, and ϵ_{ij} indicates the random error **(Figure 1**) (Bates et al., 2014). Concerning the fixed and random effects in mixed-effects models, according to the exploration of model 3- regression models with interactions between continuous and state predictors, uninsured children and percentage of the rural population are likely inference for fixed effects since those two have the same direction and little variation in the relationship to missing ASD in slopes between states. At the same time, including states as random effects because of predictors among county-level observations are nested within states.

School system data provides the records, including students' and teachers' assessments, and provides students' and parents' sociodemographic information. In this study, school system data shows a discrepancy of ASD prevalence among states, emphasizing the need to explore the sociodemographic factors related to ASD, such as local clinical care, health outcome, and social-economic disparities. In addition, detailed and complete school system data would benefit better allocated education and health care resources to children with ASD and address differences between states on *potential missing ASD* rates in the public education system and clinical health systems (Boswell et al., 2014). The prevalence of ASD estimates is usually obtained by either a surveillance system such as ADDM or using existing databases collected by education systems such as IDEA or other longitudinal studies (Nevison et al., 2018). Generally, survey-based prevalence studies or epidemiological and educational data typically are not co-considered (Chiarotti & Venerosi, 2020). The current study compared the ASD prevalence estimates by CDC and the observed ASD data collected by IDEA and calculated the likely percentage missing of ASD between those two ASD estimate systems, which provide a profile and overview of how different the value of ASD estimate systems are. All in all, no matter the estimates of ASD prevalence from CDC's ADDM, or the observed data derived from IDEA, they all show that variation of ASD prevalence does exists in geographical areas (Chiarotti & Venerosi, 2020). The current study evaluates how much difference or missed ASD cases might exist between ADDM's and IDEA's ASD prevalence estimation system. The current research suggests that county-level data could be more accurate and objectively reflect the geographical and sociodemographic factors related to the early identification of autism.

The previously reviewed literature has indicated that racial, ethnic, and socioeconomic disparities are crucial indicators for diagnosing and treating children with ASD (Nevison et al., 2018). Further, children's geographic location also plays a vital role in the prevalence of autism (Boswell et al., 2014). This study collected data from CA, ID, and MS county-level school systems and showed that potentially missed ASD cases may differ across states. Those findings are consistent with the reviewed literature that socioeconomic status and location factors are significantly associated with the potential prevalence of missing ASD (Boswell et al., 2014). Furthermore, the current study also found a notable association between socioeconomic factors and geographic location wherein ASD case disparities are significant, especially in rural areas, are highly associated with local clinical care, health outcome, and social-economic status. In addition, this study contributes to the reviewed literature, indicating that sociodemographic factors- rural/urban area and insurance coverage among children need to be considered when *potential missing ASD* are evaluated. In this study, the significant indicators of uninsured children and the rural populations suggest that exploring medical insurance coverage and the demographic backgrounds of children in special education settings may provide further insight into the prevalence of autism throughout the state.

This study aimed to develop a metric of potentially missed ASD cases and outline the associations with this initial metric and socio-demographic factors. The data used in this study interpreted that missing ASD estimates at the state and county levels are related to, but not redundant with, a number of relevant socio-demographic predictor variables. The study also emphasizes that while school data plays a vital role in identifying children with ASD, more may be done with identified case metrics than is typically conducted in analyses (Boswell et al., 2014). Additionally, the results showed a meaningful relationship between predictors and outcome. For instance, as mentioned, that the *likely percentage missing of ASD* is highly related to the location. Consistent with this, the same predictors in the different states show different directions and strengths in relationship with the outcome variable. Thus, the significance of association between predictors varies among states which highlights that geographic factors play an essential element in the outcome.

In this study, models built by IDEA data can answer that the *potential missingness of ASD* is related to a number of county variables across states. However, the correlation varies from small to large with different directions across the states. Moreover, the gap between actual observed ASD cases between the CDC's estimation states and the variance between states exhibit the need to render ASD estimates more broadly to examine if the number of children served in special education meets the number expected from the CDC's estimation. The school data is helpful for this purpose as IDEA requires regular reporting of ASD students by tracking their health behavior, academic performance, family socioeconomic status, and local social demographic (Boswell et al., 2014). In the current study, we also find some variables are likely to be combined; combined variables will most likely occur in further studies because those variables have large correlations. For instance, primary care physicians and mental health providers, children in poverty and median household income, and uninsured children, the percentage of the rural population, and child mortality could be used as combined variables.

5.2 Limitation and Future Directions

One limitation of this study was that it only includes three states' county-level data. This greatly limits our understanding of variation in ASD 'missingness' across the United States, though it does help us to see that there is likely a great deal of variation. Another limitation of the study relates to the county-level sociodemographic covariates with some missing values and the continuous variables having an extensive range. Possible future studies' directions to investigate the relationship between missed ASD case metrics and the percentage of the rural populations could be to categorize rural to urban populations into urbanized areas, urban clusters, mostly rural, and completely rural by the census bureau identification (Ratcliffe et al., 2016). Applying rural to urban populations as a continuous variable measures the ratio between rural and urban populations in each county, however, making the ratio categorizing could avoid extreme values like a metropolitan area or a completely rural area and better describe an association with outcome. Also, some predictors can be modeled together, such as median household income with children in poverty, and primary care physicians with mental health providers as one covariance by principal components analysis. For the independent variables with missing, multivariate imputation by chained equations can be applied in a future study with a larger sample size.

5.3 Conclusion

This study co-considered the ASD prevalence from both the CDC's estimation and the observed value from IDEA to generate *potential missing ASD cases* matrix under the comparison between two different ASD monitoring systems. The current study suggests that the likely percentage missing of ASD is related to the ratio of rural and urban population and medical insurance coverage. The study also highlights that while school data plays a vital role in identifying children with ASD, more may be done with identified case metrics than is typically conducted in analyses. The study also indicates that mixed-effects models examining county-level ASD information is needed to better determine between socio-demographic and socioeconomic factors in relation to missing ASD cases using IDEA data.

References

2020 County Health Rankings Key Findings Report. (2020). County Health Rankings & Roadmaps. Retrieved July 19, 2021, from https://www.countyhealthrankings.org/reports/2020-county-health-rankings-key-findings-report

American Psychiatric Association. (2013). Autism Spectrum Disorder. https://www.apa.org/topics/autism-spectrum-disorder

Barger, B., Rice, C., Simmons, C. A., & Wolf, R. (2018a). A systematic review of part c early identification studies. *Topics in Early Childhood Special Education*, *38*(1), 4–16. https://doi.org/10.1177/0271121416678664

Barger, B., Rice, C., Wolf, R., & Roach, A. (2018b). Better together: Developmental screening and monitoring best identify children who need early intervention. *Disability and Health Journal*, *11*(3), 420–426. https://doi.org/10.1016/j.dhjo.2018.01.002

Barnard-Brak, L. (2019). Educational versus clinical diagnoses of autism spectrum disorder: Updated and expanded findings. *School Psychology Review*, *48*(2), 185–189. https://doi.org/10.17105/SPR-2018-0009.V48-2

Barton, E. E., Harris, B., Leech, N., Stiff, L., Choi, G., & Joel, T. (2016). An analysis of state autism educational assessment practices and requirements. *Journal of Autism and Developmental Disorders*, *46*(3), 737–748. https://doi.org/10.1007/s10803-015-2589-0

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Becerra, T. A., von Ehrenstein, O. S., Heck, J. E., Olsen, J., Arah, O. A., Jeste, S. S., Rodriguez, M., & Ritz, B. (2014). Autism spectrum disorders and race, ethnicity, and nativity: A population-based study. *Pediatrics, 134*(1), e63–e71. https://doi.org/10.1542/peds.2013-3928

Bilaver, L. A., Cushing, L. S., & Cutler, A. T. (2016). Prevalence and correlates of educational intervention utilization among children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*(2), 561–571. https://doi.org/10.1007/s10803-015-2598-z

Boswell, K., Zablotsky, B., & Smith, C. (2014). Predictors of autism enrollment in public school systems. *Exceptional Children*, *81*(1), 96–106. https://doi.org/10.1177/0014402914532230

Bricker, D., Macy, M., Squires, J., & Marks, K. (2013). *Developmental screening in your community: An integrated approach for connecting children with services*. Paul H. Brookes Publishing.

Bureau, U. C. (2021). *Urban and Rural*. The United States Census Bureau. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html

Centers for Disease Control and Prevention. (2020). *Autism and Developmental Disabilities Monitoring (ADDM) Network*. https://www.cdc.gov/ncbddd/autism/addm.html

Centers for Disease Control and Prevention. (2021). *Signs & Symptoms | Autism Spectrum Disorder (ASD) | NCBDDD*. https://www.cdc.gov/ncbddd/autism/signs.html

Chiarotti, F., & Venerosi, A. (2020). Epidemiology of autism spectrum disorders: A review of worldwide prevalence estimates since 2014. *Brain Sciences, 10*(5), 274. https://doi.org/10.3390/brainsci10050274

American Psychiatric Association. (2016). Children diagnosed with autism at earlier age more likely to receive evidence-based treatments. *Psychiatry.org*. Retrieved July 19, 2021, from https://www.psychiatry.org/newsroom/news-releases/children-diagnosed-with-autism-at-earlier-age-more-likely-to-receive-evidence-based-treatments

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, *18*(2), 72–145. https://doi.org/10.1177/1529100617727266

Ennis, R. P., Blanton, K., & Katsiyannis, A. (2017). Child find activities under the individuals with disabilities education act: Recent case law. *TEACHING Exceptional Children*, *49*(5), 301–308. https://doi.org/10.1177/0040059916685063

Fombonne, E. (2018). Editorial: The rising prevalence of autism. *Journal of Child Psychology and Psychiatry*, *59*(7), 717–720. https://doi.org/10.1111/jcpp.12941

Galea, S., Tracy, M., Hoggatt, K. J., DiMaggio, C., & Karpati, A. (2011). Estimated deaths attributable to social factors in the united states. *American Journal of Public Health*, *101*(8), 1456–1465. https://doi.org/10.2105/AJPH.2010.300086

Gist, L., & Stein, M. T. (2014). Developmental screening in your community: An integrated approach for connecting children with services. *Journal of Developmental & Behavioral Pediatrics*, *35*(7), 469. https://doi.org/10.1097/DBP.0000000000000088

Guerin, R. (2004). The American board of pediatrics 2003–04 workforce data. *Chapel Hill, NC: American Board of Pediatrics*.

Jones, W., & Klin, A. (2009). Heterogeneity and homogeneity across the autism spectrum: The role of development. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(5), 471–473. https://doi.org/10.1097/CHI.0b013e31819f6c0d

Landa, R. J. (2018). Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders. *International Review of Psychiatry (Abingdon, England)*, *30*(1), 25–39. https://doi.org/10.1080/09540261.2018.1432574

Lipkin, P. H., Macias, M. M., & Council on children with disabilities, section on developmental and behavioral pediatrics. (2020). Promoting optimal development: Identifying infants and young children with developmental disorders through developmental surveillance and screening. *Pediatrics*, *145*(1), e20193449. https://doi.org/10.1542/peds.2019-3449

Ma, J., Pender, M., & Welch, M. (2016). Education Pays 2016: The benefits of higher education for individuals and society. Trends in higher education series. In *College Board*. https://eric.ed.gov/?id=ED572548

MacFarlane, J. R., & Kanaya, T. (2009). What does it mean to be autistic? Inter-state variation in special education criteria for autism services. *Journal of Child and Family Studies*, *18*(6), 662. https://doi.org/10.1007/s10826-009-9268-8

Maenner, M. J. (2020). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 Sites, United States, 2016. *MMWR. Surveillance Summaries*, *69*. https://doi.org/10.15585/mmwr.ss6904a1

Mandell, D. S., & Palmer, R. (2005). Differences among states in the identification of autistic spectrum disorders. *Archives of Pediatrics & Adolescent Medicine*, *159*(3), 266–269. https://doi.org/10.1001/archpedi.159.3.266

McCarty, A. T. (2016). Child poverty in the united states: A tale of devastation and the promise of hope. *Sociology Compass*, *10*(7), 623–639. https://doi.org/10.1111/soc4.12386

Murphey, D. (2017). *Health Insurance Coverage Improves Child Well-Being*. Child Trends. https://www.childtrends.org/publications/health-insurance-coverage-improves-child-well

Nevison, C., Blaxill, M., & Zahorodny, W. (2018). California autism prevalence trends from 1931 to 2014 and comparison to national ASD data from IDEA and ADDM. *Journal of Autism and Developmental Disorders*, *48*(12), 4103–4117. https://doi.org/10.1007/s10803-018-3670-2

Nevison, C., & Zahorodny, W. (2019). Race/ethnicity-resolved time trends in united states asd prevalence estimates from IDEA and ADDM. *Journal of Autism and Developmental Disorders*, *49*(12), 4721–4730. https://doi.org/10.1007/s10803-019-04188-6

National Institute of Mental Health. (2020). *Autism Spectrum Disorder (ASD)*. Retrieved July 19, 2021, from https://www.nimh.nih.gov/health/statistics/autism-spectrum-disorder-asd

Petrocchi, S., Levante, A., & Lecciso, F. (2020). Systematic review of level 1 and level 2 screening tools for autism spectrum disorders in toddlers. *Brain Sciences*, *10*(3), 180. https://doi.org/10.3390/brainsci10030180

Randall, M., Egberts, K., Samtani, A., & Willians, K. (2018). Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *The Cochrane Database of Systematic Reviews*, 7(7). https://doi.org/10.1002/14651858.CD009044.pub2

Ratcliffe, M., Burd, C., Holder, K., & Fields, A. (2016). Defining Rural at the U.S. Census Bureau. https://www.census.gov/library/publications/2016/acs/acsgeo-1.html

Rice, C. E., Rosanoff, M., Dawson, G., Durkin, M. S., Croen, L. A., Singer, A., & Yeargin-Allsopp, M. (2012). Evaluating changes in the prevalence of the autism spectrum disorders (ASDs). *Public Health Reviews*, *34*(2), 17. https://doi.org/10.1007/BF03391685

Shaw, K. A. (2020). Early identification of autism spectrum disorder among children aged 4 years—early autism and developmental disabilities monitoring network, six sites, United States, 2016. *MMWR. Surveillance Summaries*, *69*. https://doi.org/10.15585/mmwr.ss6903a1

Sheldrick, R. C., Maye, M. P., & Carter, A. S. (2017). Age at first identification of autism spectrum disorder: An analysis of two us surveys. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(4), 313–320. https://doi.org/10.1016/j.jaac.2017.01.012

Sullivan, A. L. (2013). School-based autism identification: Prevalence, racial disparities, and systemic correlates. *School Psychology Review*, *42*(3), 298–316. https://doi.org/10.1080/02796015.2013.12087475

Wang, L., Mandell, D. S., Lawer, L., Cidav, Z., & Leslie, D. L. (2013). Healthcare service use and costs for autism spectrum disorder: A comparison between Medicaid and private insurance. *Journal of Autism and Developmental Disorders*, *43*(5), 1057–1064. https://doi.org/10.1007/s10803-012-1649-y

Zajacova, A., & Everett, B. G. (2014). The nonequivalent health of high school equivalents. *Social Science Quarterly*, *95*(1), 221–238. https://doi.org/10.1111/ssqu.12039

Zirkel, P. A. (2011). Autism litigation under the IDEA: A new meaning of “Disproportionality”? *Journal of Special Education Leadership*, *24*(2), 92–103.

APPENDIX A

Table 8. Spearman correlations between independent variables with each other in 2020 County Health Rankings annual report in CA, ID and MS

	PCP			MHP			CIP			UC		
	CA	ID	MS	CA	ID	MS	CA	ID	MS	CA	ID	MS
PCP	1	1	1									
MHP	**0.64*** **	**0.62*** **	**0.53*** **	1	1	1						
CIP	**-0.62*** **	-0.26	-0.02	-0.29*	-0.30	-0.05	1	1	1			
UC	-0.16	-0.24	**-0.33* **	0.01	**-0.43* **	**-0.40** **	0.09	0.36	**-0.32* **	1	1	1
RSBW	0.16	-0.31	**0.41*** **	0.26*	0.19	**0.45** **	-0.03	-0.13	-0.17	0.25	0.05	-0.17
HSG	-0.01	-0.21	-0.07	-0.30	-0.20	-0.06	-0.32	0.13	**-0.45* **	-0.06	0.07	0.06*
MHI	**0.65*** **	0.05	0.08**	0.29**	0.11	0.11	**-0.85*** **	**-0.84*** **	**-0.89*** **	-0.26*	-0.17	0.25
CM	**-0.66*** **	**-0.48* **	-0.21*	**-0.41** **	**-0.52* **	-0.09	**0.73*** **	0.50	**0.58*** **	0.09	**0.68** **	-0.08
RP %	**-0.43*** **	-0.23	**-0.56*** **	-0.21	**-0.51*** **	**-0.61*** **	**0.37* **	**0.30* **	0.00	**0.46*** **	**0.57*** **	**0.59*** **

Spearman correlations between independent variables with each other in 2020 County Health Rankings annual report in CA, ID and MS (continued)

	RSBW			HSG			MHI			CM		
	CA	ID	MS	CA	ID	MS	CA	ID	MS	CA	ID	MS
PCP												
MHP												
CIP												
UC												
RSBW	1	1	1									
HSG	-0.02	-0.17	-0.04	1	1	1						
MHI	-0.13	0.48	0.27*	0.28	-0.19	**0.42***	1	1	1			
CM	0.02	0.04	**-0.33****	-0.28	-0.15	-0.17	**-0.81*****	-0.35	**-0.66*****	1	1	1
RP %	**0.40****	-0.12	**-0.41*****	-0.05	0.20	0.01	**-0.69*****	-0.13	-0.17**	**0.58*****	**0.68*****	**0.27****

Note. * Indicates statistical significance at an alpha level of .05; 0.001 '***' 0.01 '**' 0.05 '*'; Primary care physicians = PCP, Mental Health Providers = MHP, Children in poverty = CIP, Uninsured children = UC, Residential segregation- Black/White = RSBW, High school graduation = HSG, Median household income = MHI, Child mortality = CM, Rural population, %= RP%

APPENDIX B

Table 9. Comparison of mixed effects models between Random-intercept model (Model 4) and Random-slope model (Model 5)

	Model 4 Random-intercept model	**Model 5 Random-slope model**
Fixed effects		
Intercept [a]	25.08 (11.45)	23.61 (29.41)
Primary care physicians	0.03 (0.07)	0.02 (0.18)
Random effects		
State	339.20 (18.42)	254.10 (50.41)
Residual	609.40 (24.69)	584.30 (24.17)
Primary care physicians		0.08 (0.28)
AIC	1581.60	1582.5
R^2 c	0.36	0.69
Fixed effects		
Intercept [b]	24.68 (11.59)	22.94 (29.14)
Mental health providers	0.01 (0.01)	0.005 (0.03)
Random effects		
State	371.50 (19.27)	2515.00 (50.41)
Residual	606.70 (24.63)	583.30 (24.15)
Mental health providers		0.003 (0.05)
AIC	1581.10	1583.00
R^2 c	0.38	0.75
Fixed effects		
Intercept [c]	16.6 (10.86)	17.83 (6.08)
Children in poverty	0.0005 (0.0002)	0.0003 (0.0012)
Random effects		
State	255.20 (15.98)	0.0001 (0.002)
Residual	600.20 (24.50)	577.90 (24.04)
Children in poverty		0.0001 (0.002)
AIC	1578.30	1582.3
R^2 c	0.32	0.84
Fixed effects		
Intercept [d]	20.32 (12.97)	23.24 (8.64)
Uninsured children	0.0014 (0.0018)	-0.0001 (0.0032)
Random effects		
State	278.90 (16.70)	0.0002 (0.02)
Residual	610.20 (24.70)	590.30 (24.29)
Uninsured children		0.0001 (0.005)
AIC	1581.30	1583.00
R^2 c	0.32	0.45
Fixed effects		
Intercept [e]	82.08 (23.94)**	85.60 (31.90)**

	Model 4 Random-intercept model	Model 5 Random-slope model
High school graduation	-0.0007 (0.0003)*	-0.0007 (0.0003)*
Random effects		
State	338.80 (18.41)	1685.00 (41.04)
Residual	586.80 (24.22)	578.10 (24.04)
High school graduation		0.0001 (0.0001)
AIC	1575.30	1580.40
R^2 c	0.38	0.67
Fixed effects		
Intercept [f]	19.73 (10.38)	23.21 (8.26)
Child mortality	0.11 (0.09)	0.02 (0.14)
Random effects		
State	188.40 (13.72)	52.44(7.24)
Residual	572.40 (23.93)	564.91 (23.77)
Child mortality		0.02 (0.15)
AIC	1228.70	1232.1
R^2 c	0.26	0.36
Fixed effects		
Intercept [g]	28.96 (12.23)	30.57 (10.99)
Residential segregation-Black/White	-0.07(0.16)	-0.09 (0.16)
Random effects		
State	269.40 (16.41)	201.31(14.19)
Residual	551.00 (23.47)	546.55 (23.37)
Residential segregation-Black/White		0.007 (0.08)
AIC	1233.30	1236.90
R^2 c	0.33	0.37
Fixed effects		
Intercept [h]	21.35 (13.40)	29.28 (31.56)
Median household income	0.0001 (0.0001)	-0.0001 (0.0003)
Random effects		
State	366.70 (19.15)	2848.00 (23.80)
Residual	607.40 (24.65)	566.20 (23.77)
Median household income		0.0001 (0.0005)
AIC	1581.30	1576.1
R^2 c	0.38	0.62

	Model 4 Random-intercept model	Model 5 Random-slope model
Fixed effects		
Intercept [I]	28.47 (11.49)	24.09 (4.88)*
Rural population, %	-0.03(0.07)	-0.0647(0.2085)
Random effects		
State	344.90 (18.57)	33.57 (5.79)
Residual	609.50 (24.69)	527.42 (22.97)
Rural population, %		0.12(0.34)
AIC	1581.70	1563.2
R^2 c	0.36	0.56

Note. * Indicates statistical significance at an alpha level of .05; 0.001 '***' 0.01 '**' 0.05 '*'
R^2c=conditional R^2
AIC= Akaike An Information Criterion
[a] Model estimates for primary care physicians per 100,000 (2017)
[b] Model estimates for mental health providers per 100,000 (2019)
[c] Model estimates for children in poverty per 100,000 (2018)
[d] Model estimates for uninsured children per 100,000 (2017)
[e] Model estimates for high school graduation per 100,000 (2016-2017)
[f] Model estimates for child mortality per 100,000 (2015-2018)
[g] Model estimates for residential segregation rate between Black and White (2014-2018)
[h] Model estimates for median household income (2018)
[I] Model estimates for ratio of rural to urban population based on Census Population Estimates (2010)