Exploring Children's Depression: The Impacts of Social and Environmental Factors*

Children Facing Daily Bullying, Social Difficulties, Family Mental Illness and Violence Are Most at Risk for Depression

Boxuan Yi

December 1, 2024

Abstract

This paper uses data from the US National Survey of Children's Health to examine factors contributing to depression in children aged 6 to 17, focusing on family, social interactions, and environmental influences. A Bayesian logistic model shows that being bullied, difficulty making friends, living with mental illness, and experiencing violence increase the risk of depression, while family poverty does not have a significant impact. Children often lack the awareness or resources to seek help like adults do, so understanding the signs and causes of children's depression is important for early intervention and support.

Table of contents

*Code and data supporting this analysis is available at: https://github.com/Elaineyi1/Children_Health

1 Introduction

Every child experiences emotional ups and downs, but prolonged periods of sadness or disinterest may indicate depression—a mood disorder marked by a persistent sense of sadness or emptiness affecting daily activities. While depression itself is not typically considered life-threatening, it can lead to suicidal thoughts or self-harm attempts (Leslie Rachel Miller). Early detection and intervention are vital to ensuring children receive appropriate care and resources to manage their mental health. Given that children may lack sufficient understanding of depression and may not know how to seek help, identifying the signs of depression and understanding its underlying causes is important for providing appropriate support. For this analysis, I use data from the 2023 National Survey of Children's Health (NSCH), conducted by the United States Census Bureau. This survey collects demographic, family, and health-related information about children, defined as individuals aged 0–17 years.

In this paper, the estimand being explored is whether a child has depression. I use a Bayesian logistic regression model, as the outcome is binary, to analyze how the following predictors affect children's depression: being bullied, difficulty making or keeping friends, living with someone who is mentally ill or severely depressed, whether the child is a victim of violence or has witnessed violence in the neighbourhood, and family poverty. The reason for choosing these predictors is that it is known that some children are more likely to develop anxiety or depression when they experience trauma, violence, being bullied, or when their own parents have anxiety or depression (Centers for Disease Control and Prevention 2024). The model results indicate that children facing daily bullying, having a lot of difficulty making and keeping friends, living with someone who is mentally ill, suicidal, or severely depressed, and being a victim of violence or witnessing violence in their neighbourhood are more likely to have depression. Interestingly, poverty does not affect depression significantly. The model is also tested and shows an accuracy of 92.33%.

The data utilized for analysis is introduced in Section 2. Following that, Section 3 presents the model created to examine whether the child has depression. I will then explain the results derived from the model in Section 4. Lastly, Section 5 discusses the results in a broader context and addresses the weaknesses of this paper, as well as future directions for research. This paper uses the programming language `R` (R Core Team 2022). The analysis, the model and all the visualizations use the following packages: `dplyr` (Wickham et al. 2023), `tidyr` (Wickham, Vaughan, and Girlich 2024), `arrow` (Richardson et al. 2024), `modelsummary` (Arel-Bundock 2022), `here` (Müller 2020), `rstanarm` (Goodrich et al. 2024), `janitor` (Firke 2023), `haven` (Wickham, Miller, and Smith 2023), `ggplot2` (Wickham 2016), tibble (Müller and Wickham 2023), `testthat` (Wickham 2011).

2 Data

2.1 Overview

The dataset I use is from the 2023 National Survey of Children's Health (NSCH) conducted by the United States Census Bureau (US Census Bureau 2024a). A screener instrument identifies households with children

and conducts the screener survey, after which one child is randomly selected from each household as the subject of a more detailed follow-up topical questionnaire (US Census Bureau 2024b). The information collected includes the well-being of children, such as access to and quality of health care, family interactions, parental health, school and after-school experiences, and neighbourhood characteristics [reginfoDocuments]. The dataset used in this analysis is from the follow-up survey, which also includes questions from the screener survey. Data were collected from June 2023 to January 2024.

Children in this paper are defined as individuals aged 0–17 years. During the surveys, they are separated into three groups: T1, T2, and T3, representing the age ranges 0–5, 6–11, and 12–17, respectively (inclusive). The follow-up questionnaires have three versions, with slight differences based on the child's age group. For example, children aged 0 to 5 were not asked questions about difficulty making friends, while children older than 5 were not asked whether they could focus on a task for at least a few minutes. Since it is unlikely for children aged 5 and under to develop depression due to external factors, this analysis focuses only on the age groups 6–11 and 12–17.

The key variables from the dataset that are used in the paper are:

- Depression: Whether the child has depression.
- Depression Current: Whether the child currently has depression.
- Depression Level: The level of depression if the child currently has depression.
- Age: The age of the child.
- Lived with Mental Illness: Whether the child lives with someone who was mentally ill, suicidal, or severely depressed.
- Hopeful: How often the family remains hopeful during difficult times when facing problems.
- Violence: Whether the child was a victim of violence or witnessed violence in their neighbourhood.
- Bullied: The frequency with which the child is bullied, picked on, or excluded by other children, excluding siblings or dating partners.
- Friends: The level of difficulty the child has in making or keeping friends compared to other children their age.
- Screen time: The amount of time the child spends on most weekdays in front of a TV, computer, cellphone, or other electronic device, watching programs, playing games, accessing the internet, or using social media.
- Poverty: The family poverty ratio is calculated as the ratio of total family income to the family poverty threshold and reported as a rounded percentage from 50 to 400. Values beyond this range are reported as 50 or 400, respectively. A larger ratio indicates a wealthier family.

After eliminating rows with empty values and excluding children who do not fall within the appropriate age range, the dataset contains 24,456 observations. I randomly split the dataset into a training set containing 80% of the rows and a testing set containing 20%. The training set, with 19,564 rows, is used to create a model to explore whether a child has depression based on the following predictors: being bullied, difficulty making or keeping friends, living with someone with a mental illness, whether the child is a victim of violence or has witnessed violence in the neighbourhood, and family poverty. The testing set, with 4,892 rows, is used to evaluate the model's accuracy. The data cleaning process can be found in Section 6.3.

2.2 Methodology and Measurement

The survey selected approximately 385,000 addresses from the Census Master Address File to identify households with children under 18 years old. State-level samples were designed to produce roughly equal numbers of completed questionnaires across states and the District of Columbia (US Census Bureau 2024b).

A screener questionnaire was sent to the sampled households to confirm occupancy and the presence of eligible children aged 0–17. It included demographic questions and a series of health questions used to determine whether each eligible child could be classified as having a special health care need. For households with eligible children, one child was subsampled as the subject of the topical questionnaire, which is the follow-up

survey. This follow-up asked about access to and quality of health care, family interactions, parental health, school and after-school experiences, and neighbourhood characteristics. Subsampling prioritized children with special health care needs (oversampled at 80%) and children aged 0–5 (oversampled at 60%) to address their underrepresentation in surveys and administrative records. This design supported robust representation of children and enabled reliable multi-year trend analyses (US Census Bureau 2024b).

61% of respondents were mothers, and 29% were fathers (biological, step, foster, or adoptive). Six percent were other relatives or caregivers. The weighted Screener Completion Rate was 41.9%, and the weighted Topical Completion Rate was 27.1% (US Census Bureau 2024b).

Selected households were mailed instructions to complete the 2023 NSCH, with the option to respond in English or Spanish. Non-responding households received follow-up mailings as reminders. To complete the survey online, households received invitation letters with instructions for accessing the web instrument. Respondents first completed the screener, which determined eligibility and directed them to the topical questionnaires. The subsampling occurred immediately after the screener. For paper completion, households received a self-administered paper-and-pencil screener, which was returned by mail. If eligible children were identified, a topical questionnaire was sent to the same household for completion and return. Thirty percent of the sample most likely to respond by paper received the paper questionnaire in the initial mailing. Paper questionnaires were also included in the third and fourth follow-up mailings for addresses requiring at least two follow-ups. The 2023 NSCH implemented unconditional incentives for 90% of the sample, offering a $5 cash incentive in the initial mailing to boost response rates. A control group of 10% of the sample received no cash incentive to assess the impact of the incentive on response rates (US Census Bureau 2024b).

The survey questions related to the variables used in this paper, which translate real-world phenomena into data, will be attached in Section 6.2. More details about surveys and samplings can be found in Section 6.1.

2.3 Data Visualizations and Analysis

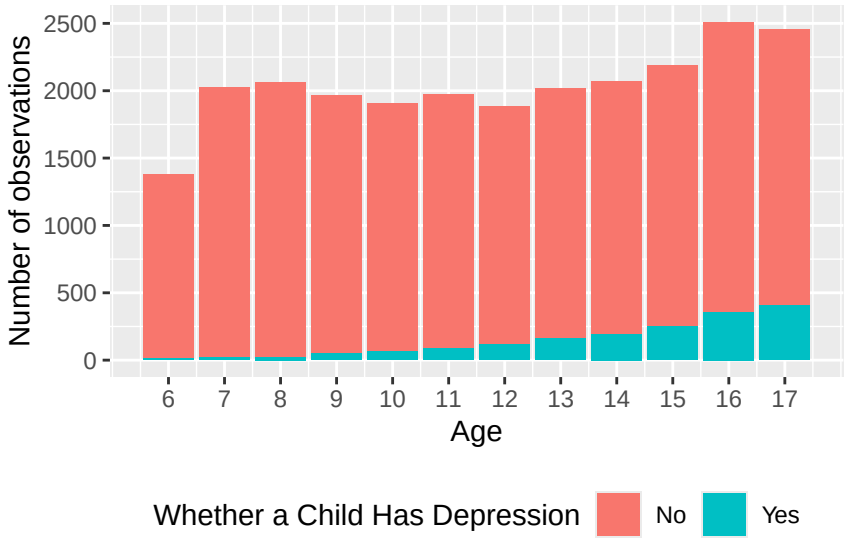


Figure 1: Distribution of Child's Age and Whether the Child Has Depression

Figure 1 illustrates the distribution of children with and without depression across different ages. Both the number of children and the proportion experiencing depression increase with age. The proportion of children with depression is very small under age 10, whereas it is significantly higher among teenagers.

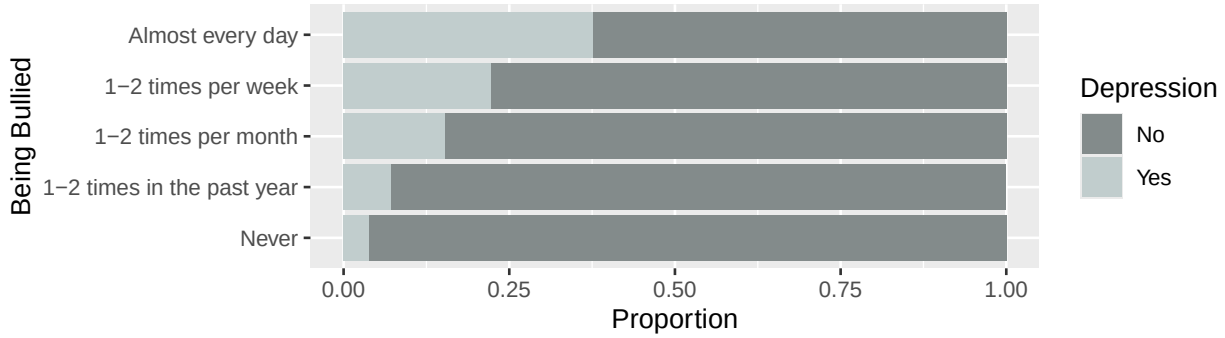
The proportion of different depression severity levels among children currently experiencing depression is shown in Figure 2. The severity of depression increases with age. Approximately 60% of children younger

than 12 have mild depression, while only half of teenagers have mild depression. For children aged 13 and older, the proportion experiencing severe depression is much higher than that of children aged 10 and younger.

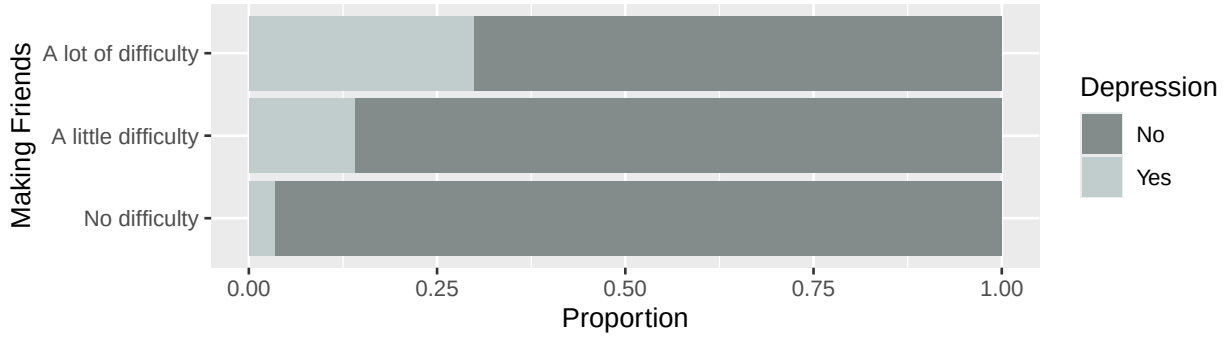
Figure 2: Age and Proportion of Depression Levels Among Depressed Children

Figure 3 (a) shows the relationship between depression and being bullied. The more frequently a child is bullied, the higher the probability of having depression. Approximately 5% of children who were never bullied have depression, compared to around 40% of children who are bullied daily. Figure 3 (b) illustrates the relationship between making friends and depression. Children who struggle more with making or keeping friends have a higher proportion of depression. The depression rate difference between children who have no difficulty and those who face significant difficulty making friends exceeds 25%. Figure 3 (c) shows how the amount of time a child spends in front of screens affects depression. The more time children spend on screens, the higher the likelihood of depression. However, the difference in depression rates for this factor is not as pronounced as for the first two.

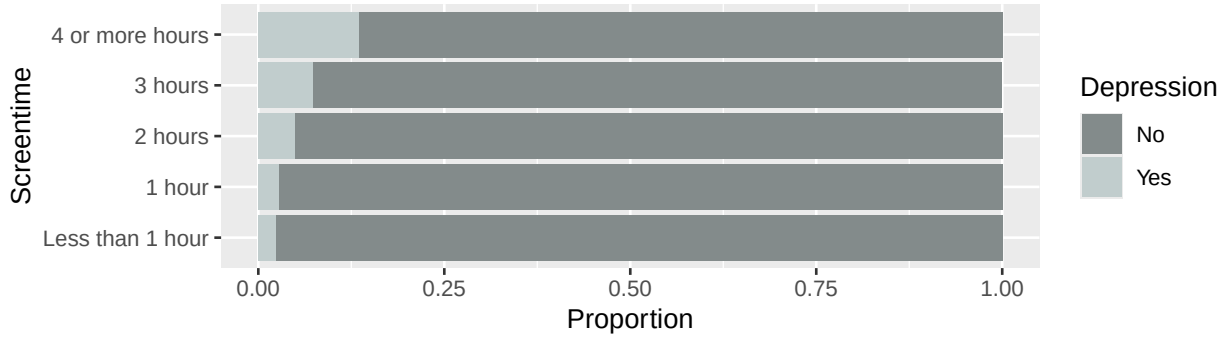
Figure 4 (a) shows the relationship between depression and living with people who have mental illness. Only around 5% of children who do not live with someone who is mentally ill, suicidal, or severely depressed have depression, while 25% of those who live with individuals experiencing mental illness have depression. Children from families that remain hopeful, even during difficult times, have the lowest likelihood of depression, as shown in Figure 4 (b). Conversely, children from families that sometimes remain hopeful have the highest likelihood of depression. The proportions of children with depression in families that never remain hopeful and those that mostly remain hopeful are similar. The difference in depression rates for this factor is less pronounced compared to other factors. In Figure 4 (c), children who have been victims of violence or have witnessed violence in their neighbourhoods have a probability of depression that is approximately 15% higher than those who have never experienced violence.



(a) Whether the Child Was Bullied and Has Depression

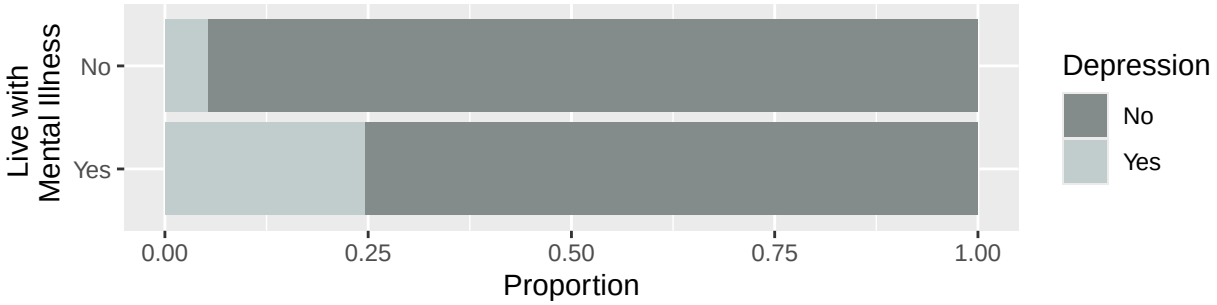


(b) Whether the Child Had Difficulty Making Friends and Has Depression

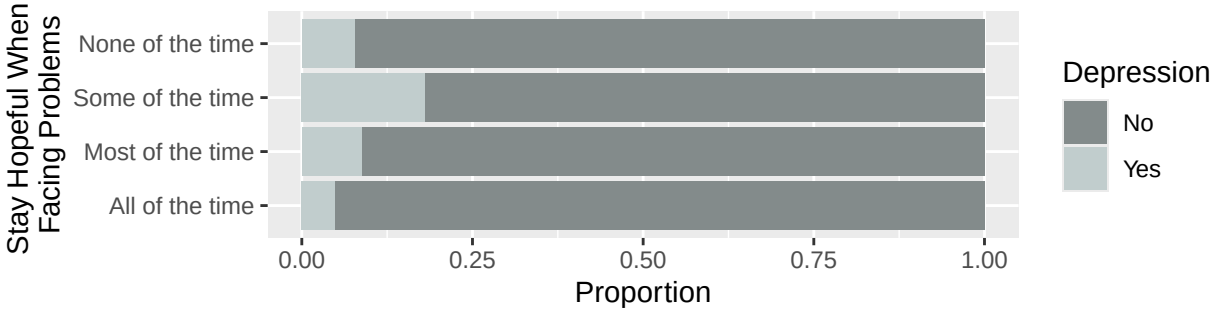


(c) Child's Weekday Screentime and Whether the Child Had Depression

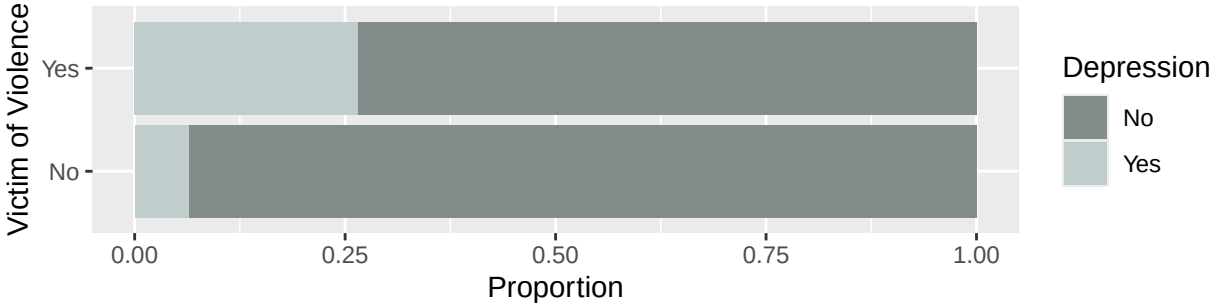
Figure 3: Getting Bullied refers to the frequency with which the child is bullied, picked on, or excluded by other children, excluding siblings or dating partners. Making Friends describes the level of difficulty this child has in making or keeping friends compared to other children their age. Screentime refers to the amount of time the child spends on most weekdays in front of a TV, computer, cellphone, or other electronic device, watching programs, playing games, accessing the internet, or using social media.



(a) Whether the Child Lived with Mental Illness and Has Depression



(b) Whether the Child's Family Stayed Hopeful and Whether the Child Has Depression



(c) Whether the Child Was a Victim of Violence and Whether the Child Has Depression

Figure 4: Lived with Mental Illness refers to living with someone who was mentally ill, suicidal, or severely depressed. Hopefulness describes how often the family remains hopeful when facing problems even during difficult times. Victim of Violence means the child was either a victim of violence or witnessed violence in the neighbourhood.

The family poverty ratio is calculated as the ratio of total family income to the family poverty threshold and reported as a rounded percentage ranging from 50 to 400. Values beyond this range are reported at 50 or 400, respectively. As shown in Figure 5, overall, the wealthier the family, the less likely a child is to experience depression. The highest proportion of depression is seen when the family poverty ratio ranges between 175 and 200, while the lowest proportion is observed when it exceeds 375, representing the wealthiest range. A second local peak in depression occurs at a ratio between 250 and 275, with a local minimum at a ratio between 200 and 225.

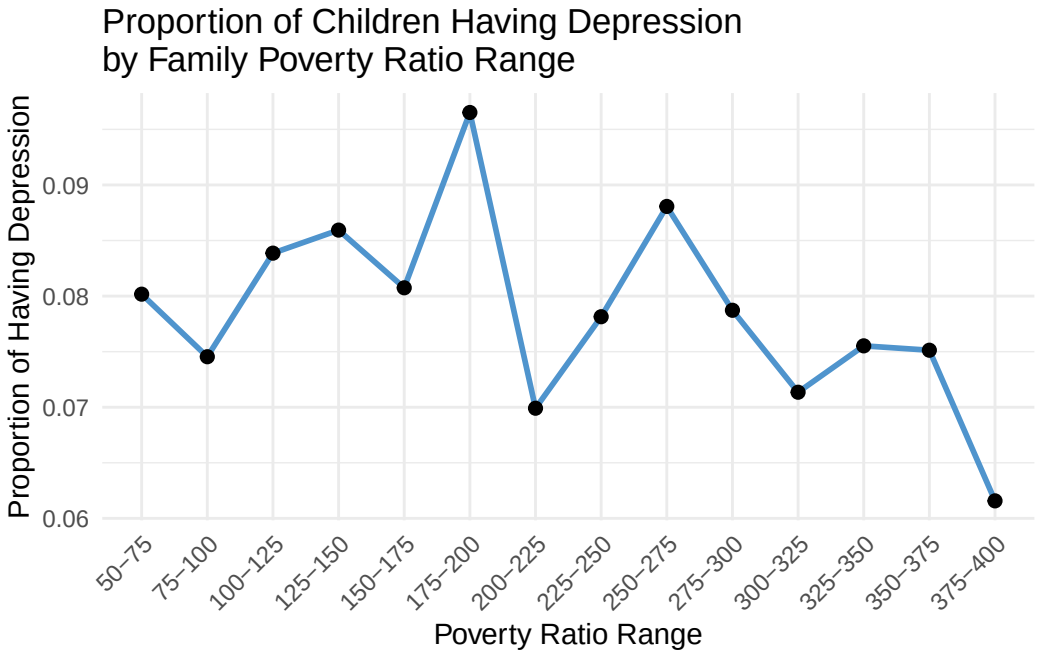


Figure 5: The family poverty ratio is calculated as the ratio of total family income to the family poverty threshold and reported as a rounded percentage from 50 to 400. Values beyond this range are reported as 50 or 400, respectively. A larger ratio indicates a wealthier family.

Therefore, the predictors I have chosen include being bullied, difficulty making or keeping friends, living with someone with a mental illness, being a victim of violence or witnessing violence in the neighbourhood, and family poverty. Screen time and hopefulness will be excluded from the model due to their less significant impacts.

3 Model

3.1 Model Set-up

Logistical regression can model binary outcomes. For this analysis, I use the `stan_glm` function from the `rstanarm` package (Goodrich et al. 2024) in the programming language `R` (R Core Team 2022) to create a Bayesian logistic regression model. The goal of the model is to explore whether a child has depression using five predictors, expressed as follows:

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) \\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times \text{Bullied}_i + \beta_2 \times \text{Friends}_i + \beta_3 \times \text{Live With Mental Illness}_i + \beta_4 \times \text{Violence}_i + \beta_5 \times \text{Poverty}_i \\
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5) \\
\beta_3 &\sim \text{Normal}(0, 2.5) \\
\beta_4 &\sim \text{Normal}(0, 2.5) \\
\beta_5 &\sim \text{Normal}(0, 2.5)
\end{aligned}
$$

where:

- y_i is the dependent variable, representing whether a child has depression, with the outcome being binary (yes or no).
- β_0 is the intercept term, representing the value when all the predictor variables are zero. It has a prior distribution that is normal with a mean of 0 and a standard deviation of 2.5.
- β_1, β_2, β_3, β_4, β_5 represent the effects of being bullied, difficulty making or keeping friends, living with someone with mental illness, whether the child is a victim of violence or witnessed violence in the neighbourhood, and family poverty respectively. Each coefficient has a prior distribution of Normal(0, 2.5).

As shown in Section 2, differences in these five predictors are associated with depression, which is why they were chosen as predictors. Being bullied can lead to emotional distress, low self-esteem, and social isolation, while difficulty making or keeping friends may impact a child's sense of belonging and emotional security, both of which can contribute to depression and other mental health challenges. Living with someone with a mental illness can create a stressful environment that affects the child's emotional well-being, as they may experience increased anxiety and uncertainty. Whether a child is a victim of violence or witnesses violence in their neighbourhood can contribute to trauma, fear, and heightened stress. Family poverty can influence mental health by limiting access to resources or increasing stress levels within the household.

3.2 Model Justification and Limitations

I expect to observe that both being bullied and having difficulty making or keeping friends are associated with a higher risk of depression, as shown in Figure 3. Additionally, both living with someone with a mental illness and being a victim of violence are likely to correlate with an increased risk of depression, as reflected in Figure 4. According to Figure 5, higher levels of family poverty are also expected to be linked to an increased likelihood of depression.

The topical dataset includes a total of 457 variables, containing a wide range of demographic information and health-related questions, such as whether the child has autism or has been treated unfairly due to race or health conditions.

It is known that some children are more likely to develop anxiety or depression when they experience trauma or stress; violence, abuse, or neglect; being bullied or rejected by other children; or when their own parents have anxiety or depression (Centers for Disease Control and Prevention 2024). Therefore, to explore childhood depression, I selected two variables related to family mental health (when their own parents have anxiety or depression), three related to social life (being bullied or rejected by other children), one related to past experiences (violence and trauma), and one pertaining to poverty, aiming to represent various aspects of children's living conditions and potential external factors using a small number of variables. It is possible that some omitted variables not included in the analysis might have provided stronger relationships, making the current model less accurate than it could be. However, many of them are correlated, which can lead to redundancy in the model. To avoid overfitting and over-complication, I chose five out of seven factors as predictors (exclude screen time and hopefulness) that best represent different aspects of children's living conditions, focusing on those with the most significant potential impact on depression. The accuracy of the model on the testing data will be shown in Section 4.

Since the outcome is binary, logistic regression is the most appropriate model to use. However, if the survey were to provide a numeric measure of the severity of children's depression, alternative models, such as linear regression, could be considered.

Table 1: Estimating whether a child has depression based on whether the child is bullied, has difficulty making or keeping friends, lives with someone with mental illness, whether the child is a victim of violence or has witnessed violence in the neighbourhood, and family poverty.

	Depression
(Intercept)	−2.414
bullied1-2 times in the past year	0.286
bullied1-2 times per month	0.662
bullied1-2 times per week	0.743
bulliedAlmost every day	1.339
friendsA little difficulty	1.136
friendsA lot of difficulty	1.838
live_with_mentalNo	−1.242
violenceYes	0.823
poverty	0.000
Num.Obs.	19 564
R2	0.140
Log.Lik.	−4094.593
ELPD	−4105.0
ELPD s.e.	83.6
LOOIC	8210.1
LOOIC s.e.	167.2
WAIC	8210.1
RMSE	0.24

4 Results

In Table 1, the coefficient for being bullied increases with frequency, indicating that the more often a child is bullied, the higher their likelihood of experiencing depression. The coefficient for being bullied almost every day is 1.339, while the coefficient for never being bullied is 0. For difficulty making or keeping friends, a child with a lot of difficulty has a coefficient of 1.838, while a child with a little difficulty has a coefficient of 1.136, and no difficulty has a coefficient of 0. This illustrates that the more difficulty a child has in making or keeping friends, the more likely they are to develop depression. The coefficient for not living with someone with a mental illness is -1.242, while living with someone with a mental illness has a coefficient of 0, suggesting that living with a person with a mental illness is associated with a higher likelihood of depression. For being a victim or witness of violence, the coefficient is 0.823, while not being a victim has a coefficient of 0, showing that experiencing or witnessing violence increases the risk of depression in children. The poverty coefficient is 0, suggesting that family poverty may not have a significant impact on children's depression in this model.

Overall, children who are bullied almost every day, have significant difficulty making friends, live with someone with a mental illness, and have experienced violence are the most likely to suffer from depression. Conversely, children who are not bullied, do not have difficulty making friends, do not live with a mentally ill person, and have not experienced violence are the least likely to have depression.

The model was trained using an 80% training set and tested on a 20% testing set, achieving a 92.33% accuracy. This high level of accuracy demonstrates the model's effectiveness in estimating children's depression.

5 Discussions

5.1 Parental Depression and Its Effects on a Child

As shown in both Section 2 and Section 4, the impacts of family are important for children's depression. Additionally, Figure 6 shows the relationship between parents' mental health and children's depression. The average mental health of parents ranges from 1 to 5, where a lower average indicates healthier parents. Figure 6 shows that the more mentally healthy the parents are, the less likely children are to become depressed. The disparity might be even greater, as sometimes people may perceive themselves as mentally healthier than they actually are.

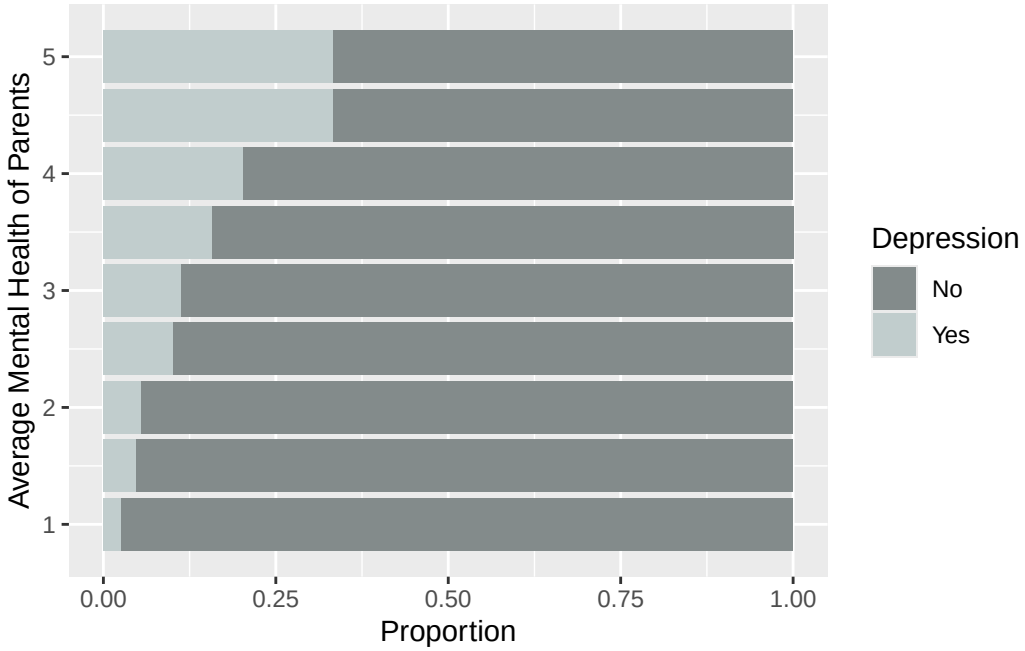


Figure 6: The distribution of the average mental health of parents and the proportion of children experiencing depression. The mental health of parents ranges from 1 to 5, where a lower average indicates healthier and more stable parents.

Parental depression can significantly impact a child's experiences. Depressed parents often interact with their children in ways that hinder development, such as using less emotional expression and making less eye contact. This can extend to everyday activities, causing a sense of isolation and limiting the child's social networks. Research shows that parental depression, including during pregnancy, is linked to long-term difficulties. It impairs parental bonding and nurturing and may lead to neglect in child safety practices, such as using car seats or getting vaccinations. School-age children of depressed parents may struggle academically, exhibit behavioural issues, and have poorer overall health (Yale Medicine).

Some symptoms of parental depression include a lack of responsiveness, where depressed parents may not respond appropriately to their child's cues, such as crying or eye contact. Depressed parents may also display inappropriate parenting behaviours, ranging from neglectful and disengaged to overly intrusive, with both extremes failing to respond sensitively to the child's needs. Additionally, children of depressed parents may experience school-related issues, such as consistent tardiness, as their parents may lack the energy or organizational skills necessary to manage routines effectively (Yale Medicine).

5.2 The Impact of Social Skills on Adolescent Mental Health

Figure 3 shows that children who are bullied more frequently and those who struggle to make or maintain friendships are at a higher risk of depression. Additionally, the amount of time spent on screens correlates with depression, with increased screen time raising the likelihood of depression, although the effect is less pronounced than that of bullying or social difficulties.

Research has found that during early to middle adolescence, there is a decline in cooperation and assertion—key social skills for positive interactions—while responsibility increases and self-control remains stable. Adolescents who showed greater improvements in responsibility and self-control were less likely to experience a rapid increase in depression, suggesting that these skills can serve as protective factors. Conversely, larger decreases in cooperation and assertion were linked to a faster increase in depression. This implies that as adolescents become less cooperative and assertive, their ability to navigate social situations and manage relationships effectively may diminish, making them more vulnerable to depression (Zhuojun Yao 2021). These findings highlight the importance of fostering social skills as a preventive strategy against depression in adolescents, and help explain why being bullied and experiencing difficulty making friends increase the likelihood of depression.

5.3 Diagnosis and Treatment of Childhood Depression

Diagnosing depression in children can be challenging, as there are no laboratory tests available. Mental health evaluations are essential for an accurate diagnosis. Consulting a child's paediatrician can be an important first step in identifying depression. Common symptoms of depression in children include loss of interest and appetite, persistent sadness, trouble sleeping, and changes in behaviour.

In severe cases, depression can lead children and teens to contemplate suicide. According to the Centers for Disease Control and Prevention (CDC), suicide is the second leading cause of death for children aged 10–14 and the third leading cause of death for children aged 15–18 (Centers for Disease Control and Prevention 2024).

Treatment approaches for depression vary, reflecting the unique nature of the condition among individuals. For mild cases, therapy alone can be as effective as medication. In moderate to severe cases, psychotherapy may be combined with medication. Antidepressants, such as selective serotonin reuptake inhibitors (SSRIs), are often considered first-line treatment options (Leslie Rachel Miller).

5.4 Weaknesses and Limitations

One significant limitation of this study is the exclusion of child-level and household-level weights, which could have enhanced the representativeness of the dataset. These weights are designed to account for variations in the child's race, sex and age. Additionally, the weights help address non-response biases, which can skew results, especially given the potential for respondents to become disengaged due to the length of the follow-up survey. Employing techniques such as post-stratification with data from the American Community Census Survey could have improved the analysis' accuracy. The absence of these adjustments is an important limitation of this study.

Another limitation involves the selection of predictors. This analysis utilized a logistic regression model with five predictors: being bullied, difficulty making or keeping friends, living with someone with a mental illness, being a victim or witness to violence, and family poverty. However, predictors such as being bullied and having difficulty making friends may be highly correlated, potentially leading to redundancy and duplicative effects. Additionally, while a wide range of variables are available in the dataset, some variables might have offered stronger information and contributed to a more precise model.

As indicated in Section 4, the coefficient for poverty is 0, suggesting that it does not appear to affect the likelihood of depression. This finding may be influenced by the high proportion of families in the dataset with a poverty ratio at 400, which reduces variability and could lead to inaccuracies in the analysis.

Lastly, as noted in Section 2, respondents were divided into age groups T1 (ages 0–5), T2 (ages 6–11), and T3 (ages 12–17). However, there are a small number of cases where the recorded age does not align with these age groups. Although these inaccuracies are minor, they could still affect the analysis to some degree.

5.5 Next Steps

Given the challenges and shifts introduced by the COVID-19 pandemic, future studies should investigate how children's mental health has evolved in the post-pandemic era. Comparing patterns before, during, and after the pandemic could help illustrate the effects of social isolation and increased screen time on children and adolescents. Additionally, research could focus on the interplay between parental and child depression, assessing how fluctuations in one may impact the other to better address mental health challenges.

The models developed in this analysis provide a foundation for discussions on resource allocation, including healthcare, clinical interventions, and the importance of familial support in promoting children's mental health. By identifying the complex factors that influence children's well-being, educators and healthcare professionals can make informed decisions. These findings can guide children in seeking help and inform the development of targeted treatment strategies for those in need.

6 Appendix

6.1 Surveys and Sampling

The following information is from 2023 National Survey of Children's Health, United States Census Bureau (US Census Bureau 2024b).

6.1.1 Frame and Sample Selection

The sample frame uses administrative records-based flags to identify four mutually exclusive strata:

- Stratum 1A and 1B: Addresses directly linked to children through administrative records are placed in Stratum 1, which includes about 80% of households with children. Within Stratum 1, if the linked child is 5 years old or younger, the address is assigned to Stratum 1A; if the child is older, the address is assigned in Stratum 1B.
- Stratum 2a: Addresses that are probabilistically linked to children. Approximately 15% of these addresses are households with children.
- Stratum 2b: The remaining addresses. Less than 5% of these addresses are households with children.

To increase sample efficiency, Stratum 1a addresses were sampled at the highest rate, followed by Stratum 1b, while Stratum 2a was sampled at a lower rate, and Stratum 2b was excluded. Sampling rates were optimized by state to maximize households with children without compromising reliability. Nationally, 70.5% of addresses were from Stratum 1, and 29.6% from Stratum 2a. Addresses were sorted by stratum and grouped by block group poverty rate to ensure high-poverty area representation. Nine states (California, Colorado, Nebraska, New York, Ohio, Oregon, Pennsylvania, Tennessee, and Wisconsin) and one metropolitan area (Atlanta, Georgia) funded oversamples to ensure more interviews in those locations.

6.1.2 Data Collection

Data collection for the 2023 National Survey of Children's Health (NSCH) began on June 23, 2023, and ended on January 19, 2024. The survey used a two-phase approach: (1) an initial household screener to determine the presence, demographic characteristics, and special health care needs status of children in the

home, and (2) a detailed topical questionnaire completed by a parent or caregiver of the selected child. Strategies to boost response rates included clear, concise question wording, multiple response options, cash incentives, and other supportive measures.

Respondents had four ways to complete the survey: Web Instrument, Paper Instrument, Telephone Questionnaire Assistance, Email Questionnaire Assistance.

- For Web Instrument, respondents were asked to verify their address. If the address displayed on screen did not match their residence, the survey was ended, and the address was removed from further mailings. If the address matched, a PIN was assigned for the respondent to log back into the survey. Alternatively, the respondent could create a new PIN by answering a security question they had set up during the original PIN creation process. After setting up the PIN, respondents indicated the number of children (ages 0–17) who usually lived at the address. If no children were present, the survey ended, and the address was removed from future mailings. If children were present, the respondent continued with the rest of the screener questionnaire.
- Paper Instrument were printed, trimmed, and stitched through an in-house print on-demand process using a system that allowed personalization to each respondent. If the respondent returned the screener by mail, it was processed to determine the presence of eligible children at the address. Returned forms were processed using iCADE, which captured responses through optical mark recognition, optical character recognition, and keying from image. If the address did not match the respondent's residence or if no children usually resided at the address, the survey was concluded, and the household was removed from further mailings. If children were listed as residing at the address, Census Bureau staff used subsampling methodology to select one child from the household roster as the subject for the topical questionnaire.
- Completion of Telephone Questionnaire Assistance over the phone was supported by trained Automated Tracking and Control staff.
- For Email Questionnaire Assistance, An email address was provided on all invitation letters and the Centurion login page. Emails were managed by call centre staff in Arizona, who logged each inquiry in a tracking spreadsheet and assigned purpose codes similar to those used for Telephone Questionnaire Assistance. Email agents used scripted responses to address common concerns and questions, ensuring consistent and accurate information. When replying, agents removed any potentially personally identifiable information from the response emails.

6.2 Survey Questions Related to the Variables

The survey questions mentioned in this paper, which translate real-world phenomena into data, are as follows (US Census Bureau 2024a):

1. **Has a doctor or other health care provider EVER told you that this child has Depression?**
    - 1 = Yes
    - 2 = No
2. **If yes, does this child CURRENTLY have the condition?**
    - 1 = Yes
    - 2 = No
3. **If yes, is it**
    - 1 = Mild
    - 2 = Moderate
    - 3 = Severe
4. **What year was this child born?**

5. **To the best of your knowledge, has this child EVER experienced any of the following? Was a victim of violence or witnessed violence in their neighbourhood**

    - 1 = Yes
    - 2 = No

6. **To the best of your knowledge, has this child EVER experienced any of the following? Lived with anyone who was mentally ill, suicidal, or severely depressed**

    - 1 = Yes
    - 2 = No

7. **When your family faces problems, how often are you likely to do each of the following? Stay hopeful even in difficult times**

    - 1 = All of the time
    - 2 = Most of the time
    - 3 = Some of the time
    - 4 = None of the time

8. **DURING THE PAST 12 MONTHS, how often was this child bullied, picked on, or excluded by other children? Do not include siblings or dating partners. If the frequency changed throughout the year, report the highest frequency.**

    - 1 = Never (in the past 12 months)
    - 2 = 1-2 times (in the past 12 months)
    - 3 = 1-2 times per month
    - 4 = 1-2 times per week
    - 5 = Almost every day

9. **Compared to other children their age, how much difficulty does this child have making or keeping friends?**

    - 1 = No difficulty
    - 2 = A little difficulty
    - 3 = A lot of difficulty

10. **ON MOST WEEKDAYS, about how much time did this child spend in front of a TV, computer, cellphone or other electronic device watching programs, playing games, accessing the internet or using social media?**

- 1 = Less than 1 hour
- 2 = 1 hour
- 3 = 2 hours
- 4 = 3 hours
- 5 = 4 or more hours

The family poverty ratio is calculated as the ratio of total family income to the family poverty threshold and reported as a rounded percentage from 50 to 400.

6.3 Data Cleaning

Variables of interest were selected from the raw dataset. Cases with `Formtype` equal to "T1"(ages 0–5) were filtered out, as the analysis focused on children aged 6–17. Missing values were removed for most variables, except for `depression_current` and `depression_level`. The `age` variable was computed based on the birth year, and a new variable, `parent_mental_health`, was derived as the average of two mental health indicators. Selected variables were then renamed for clarity and relevance. Data entries outside the target age range (6–17) were filtered out. The cleaned dataset underwent further recoding to convert categorical variables into descriptive factors with meaningful levels. For example, `depression` was recoded into "Yes" or "No," and other variables, such as `screentime` and `bullied`, were recoded into interpretable categories based on the original coding scheme. Finally, the data was split into training and testing sets (80%-20% split) using random sampling to facilitate subsequent model building and evaluation. This cleaning process ensured the dataset was structured, consistent, and ready for analysis.

References

Arel-Bundock, Vincent. 2022. "Modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Centers for Disease Control and Prevention. 2024. "Anxiety and Depression in Children." https://www.cdc.gov/children-mental-health/about/about-anxiety-and-depression-in-children.html.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Leslie Rachel Miller, Jennifer Katzenstein. "Depression in Teens and Children." https://www.hopkinsmedicine.org/health/conditions-and-diseases/depression-in-children.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to Apache Arrow*. https://github.com/apache/arrow/.

US Census Bureau. 2024a. "NSCH Datasets." https://www.census.gov/programs-surveys/nsch/data/datasets.html.

———. 2024b. "Technical Documentation Complete List." https://www.census.gov/programs-surveys/nsch/technical-documentation/complete-technical-documentation.html.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2024. *Tidyr: Tidy Messy Data*. https://tidyr.tidyverse.org.

Yale Medicine. "Parental Depression: How It Affects a Child." https://www.yalemedicine.org/conditions/how-parental-depression-affects-child.

Zhuojun Yao, Robert Enright. 2021. "A Longitudinal Analysis of Social Skills and Adolescent Depression: A Multivariate Latent Growth Approach." https://pmc.ncbi.nlm.nih.gov/articles/PMC8297573/.