Identifying Underdiagnosed ADHD in Females

Holly Erickson

Executive Summary

Attention deficit hyperactivity disorder (ADHD / ADD) is a life-long neurodevelopmental disorder. ADHD may lead to serious consequences including school dropout, depression, problems with relationships, substance abuse, delinquency, and job failure. The CDC reports that the estimated number of children in the United States diagnosed with ADHD is 6.1 million (9.4%). Boys are more likely to be diagnosed with ADHD than girls (12.9% compared to 5.6%).

Underdiagnosis in female children is a serious problem for those who suffer silently, perhaps finding out in adulthood they have a condition that has been holding them back. According to the research done by Dr. Quinn and Sharon Wigal, Ph.D., associate clinical professor of pediatrics at the University of California at Irvine, four out of 10 teachers report more difficulty in recognizing ADHD symptoms in girls than in boys. (Connolly, 2019)

Females with ADHD are often misdiagnosed. The most common diagnosis of a woman before she receives her ADHD diagnosis is depression. One hypothesis is that this is due to our difficulty separating out physical hyperactivity (prevalent in boys) to identify cognitive hyperactivity (prevalent in girls).

I created a classification model that could potentially help identify these undiagnosed females using the CDC National Children's Health Survey (NCHS) 2011 - 2012 dataset. The dataset included 366 variables. I explore several approaches for determining which subset of features to use for the model to identify females who would be more likely to remain undiagnosed if they are suffering from the disorder.

One of the key predictors of ADHD is behavioral and conduct problems, however, there are many people who "mask" their problems and have a 0 for that variable. This is important for the CDC to study because the subset of children who mask are the ones who go undiagnosed until adulthood. The final model includes a Random Forest Binary Classifier, and predicted K2Q31A with 0.955 accuracy on all females and 0.935 area ROC. The False Positives may be identified as having the highest likelihood to have undiagnosed ADHD.

Technical Report

Introduction and Background

The CDC National Children's Health Survey (NCHS) 2011-2012 dataset contains data provided by parents and caregivers of children. The dataset records responses for over 45,000 female and over 45,000 male youth (under the age of 18).

The target for my model is question K2Q31A: "Has a doctor or other health care provider ever told you the child has ADHD / ADD?" Males were diagnosed with ADD or ADHD at over twice the rate of females (12% of males vs 5.5% of females had been diagnosed). Females are more likely to mask their symptoms and go undiagnosed or misdiagnosed. (Bowman, 2020).

The Dataset

CDC's National Center for Health Statistics (NCHS), State and Local Area Integrated Telephone Survey program conducted this survey. In their landline and cell samples, a total of 847,881 households in the 50 states and DC were screened for age-eligible children (0 - 17 years old at the time of call).

The dataset is the results of the completed 95,677 detailed child-level interviews. The number of children with completed interviews per state ranged from 1,811 (South Dakota) to 2,200 (Texas) in the combined sample. Each record contains all interview data for the child and the household in which the child resides, including the child's health and health care, family functioning, parental health, neighborhood and community characteristics, health insurance coverage, and demographics.

The Respondent answering the survey can be either a parent or a guardian with knowledge of the health and health care of the sampled child in the household. For the completed NSCH interviews, 68.6% of the respondents were mothers (biological, step, foster, or adoptive), 24.2% were fathers (biological, step, foster, or adoptive), and 7.2% were other relatives or guardians.

The Goal of the Survey

The NSCH is the only national and state-level survey on the health and well-being of children, their families, and their communities. The data collected are critical for providing states with unique information for the Title V Maternal and Child Health Services Block Grant program, state-level planning and program development, federal policy and program development, and general scientific research. These data are also used by families, child health advocates, policymakers, and others.

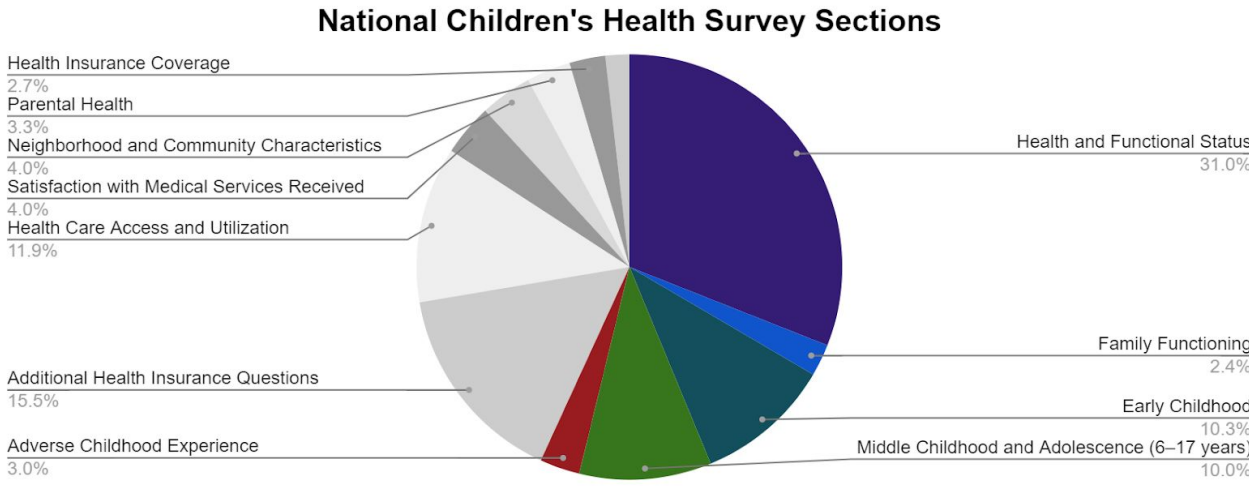
The goals of the survey include assessing the physical and emotional health of children, as well as factors that may relate to child well-being, including medical homes, family interactions, parental health, school and after-school experiences, and neighborhood characteristics. For currently uninsured children, collect detailed information about the reasons why they are

uninsured, including past enrollment and application information for public health insurance programs. Questions include age at first diagnosis for the following conditions: Intellectual disability; cerebral palsy; ADD/ADHD; behavioral or conduct problems; autism or Autism Spectrum Disorder. (CDC, 2013)

The Goal of this Project

My goal was to learn more about ADHD and identify underdiagnosed females who may be struggling for reasons that they don't understand prior to being diagnosed.

The figure below illustrates the breakdown of sections in the survey, where the size of pie slice is equal to the percent of survey questions it makes up. The most important features for predicting ADHD come primarily from the colored pie slices. However, the survey questions for Adverse Childhood Experiences (ACE) are more important for predicting ADHD in those who also have conduct and behavioral problems. Is it possible for me to create a model that predicts ADHD in females who do not have behavioral and conduct problems?



The ADHD Brain

Brain-imaging studies have shown that ADHD is a neurological dysfunction. Neurotransmitters work together in a pattern of stimulation or inhibition, the effects spreading downward, like a cascade, from stimulus input to complex patterns of response leading to feelings of well-being. There is a breakdown in the ADHD brain of the cascade of neurotransmitters in which one reaction triggers another.

ADHD brains have in common at least one defective gene, the DRD2 gene. This is the gene that makes it difficult for neurons to respond to dopamine, the neurotransmitter that is involved in feelings of pleasure, and self-regulation. Several other genes may be damaged, and the severity of ADHD is related to the number of damaged genes present. This results in the

patterns breaking, and the person not processing that feeling of stimulation or inhibition properly. (Chadd, 2020).

ADHD brains have fewer dopamine receptors, causing the human to experience fewer reward feelings from an identical input to a neurotypical brain. The ADHD brain has to flood itself with dopamine to feel the effects. Therefore, their brain becomes a dopamine-seeking missile. It predisposes individuals to high risk for multiple addictive, impulsive, and compulsive behaviors.

The substantia nigra, a part of the brain that plays a role in reward, addiction, and movement (due to its high levels of dopaminergic neurons) is located in the midbrain. In Parkinson's disease, which is characterized by a deficit of dopamine, the death of the substantia nigra is evident. ADHD is similar, and taking certain medications for ADHD may significantly increase the risk of Parkinson's, however, this may be due to correlation, not causation. Perhaps those taking ADHD medication have more severe ADHD which leads to a higher likelihood of Parkinson's. Researchers from the University of Utah found the risk in those taking medicine to be 6- to 8-fold higher to develop early-onset Parkinson's disease or related basal ganglia and cerebellum disease than peers who do not have ADHD. We do know that simply having ADHD adds risk; if they do not take medicine, they are twice as likely to develop early-onset Parkinson's than peers.

Adults with ADHD have been shown to have damaged dopamine neurons in the basal ganglia. Basal ganglia appear to serve as a gating mechanism for physical movements, inhibiting potential movements until they are fully appropriate for the circumstances in which they are to be executed. With this gating mechanism damaged, ADHD adults often report feeling physically unable to sit still and focus on the required task.

People with ADHD often are hypersensitive in one of the sensory domains: sound, touch, or smell, as well as to emotional stimuli. Any sense of rejection causes an emotional reaction that's much more extreme than usual. 98-99% of adolescents and adults with ADHD experience Rejection Sensitivity Dysphoria (RSD), and for 30%, RSD is the most impairing aspect of their ADHD. (Dodson, 2020)

People with ADHD often do not have a sense of time itself (time blindness). They commonly are lost in the moment and not present to the passage of time. They tend to be poor at knowing how long a task is going to take. Many struggle to survive in a world that isn't suited to the time blind

Methods

Specifics to Note from Dataset

1. K2Q31A - Target Column - Has a doctor or other health care provider ever told you that [S.C.] had attention deficit disorder or attention deficit hyperactive disorder, that is, ADD or ADHD? 0 = No, 1 = Yes

2. K2Q34A - Has a doctor or other health care provider ever told you that [S.C.] had behavioral or conduct problems, such as oppositional defiant disorder or conduct disorder? 0 = No, 1 = Yes

In predicting K2Q31A in underdiagnosed females, I looked at important features for all females verses for all children with 0 for K2Q34A. The data for all females shows that those females who are diagnosed with ADHD as children are those with stronger symptoms, not those who mask. These features include stronger emphasis on extrinsic variables such as ACEs as well as higher levels of emotional dysregulation.

By examining the subset of children with 0 for K2Q34A, I was able to look at feature importance for those females who don't have such extreme symptoms.

The dataset has some complexities to be mindful of in order to prevent data leakage. Some of the questions are follow-up questions that are only to be answered if an answer of "Yes" was given for a different question. I removed the follow-up questions for K2Q31A from my model and set these aside for later analysis (any question beginning in K2Q34).

Preprocessing

- In training the model, I only kept rows with an answer of 1 or 0 in the target, as they are the known outcomes.
- Removed columns with all NaN's and redundant columns with a high correlation. This keeps only those columns with a correlation of less than 0.9.
- Exploratory data analysis, including a correlation matrix to identify 63 desired features with a higher likelihood to produce a good model.
- Divide the desired features into two groups - intrinsic and extrinsic features. Note that this was done subjectively, and more discussion could be had on if they are divided properly. It gave me a place to start though.
- Impute missing values using the strategy "most frequent".

Models

I used a combination of Random Forest and Support Vector Machines to create initial models that I could fine-tune later. Random Forest is an ensemble technique that can be used to easily identify which features are most important in making predictions. The condition is based on impurity, which in case of classification problems is Gini impurity/information gain (entropy). When training a tree we can compute how much each feature contributes to decreasing the weighted impurity. In the case of Random Forest, we are talking about averaging the decrease in impurity over trees.Each tree gets a vote on label, and the majority of votes becomes the prediction.

Support vectors are margins that separate the distribution of features for the target variable. They occur in the multi-dimensional space, one dimension for each feature.The Support Vector Machine tries to place the support vectors in such a way that the margins between features are as wide as possible.

Masking vs Conduct and Behavioral Problems

It is possible to create a model with 92% accuracy in predicting that a child has been diagnosed with ADHD, simply by predicting 1 for all rows with a 1 for variable K2Q34A. This variable is asking whether a responder has been told that the child has behavioral or conduct problems,

such as oppositional defiant disorder or conduct disorder. Children who mask would be more likely to have a 0 for this variable, so I removed the subset of data for children who have a 1 for K2Q34A.

I compared models that included the entire dataset with models that only included children who had not been diagnosed with behavioral and conduct problems (K2Q34A). The highest accuracy models left out those rows, performing better despite there being a smaller dataset (85,510 vs 82,287 rows).

- 0.946 accuracy - 63 features, no K2Q34A “conduct / behavioral problems” subset
- 0.944 accuracy - uses the full list of potential features, no K2Q34A subset
- 0.939 accuracy - 63 features, includes K2Q34A subset
- 0.935 accuracy - uses the full list of potential features, includes K2Q34A subset

To highlight how the answer for K2Q34A impacts our model, let’s examine the prediction for K6Q06 (concern about behavior) for two identical models: Model A trained on data that includes K2Q34A = 1subset, and Model B that does not.

Model A is splitting the dataset predictions almost entirely by K2Q34A. Model A underpredicted rows with “Not at all” as the response for K6Q06. It missed 1,035 rows of the test set that are diagnosed with ADHD, but the responder showed little to no concern about their behavior.

Model B correctly predicted that no one in this subset diagnosed with ADHD elicits behavior concerns from the responder. It is very interesting that for rows with ADHD Diagnosed Target = 1 in this subset, we find zero rows where responders are concerned about their behavior.

The highest accuracy models didn’t use the full list of potential features, but only used the subset of features with the highest correlation to ADHD (intrinsic). This means we aren’t losing accuracy by speeding up the model by using fewer features. In fact, we’ve made it slightly more accurate.

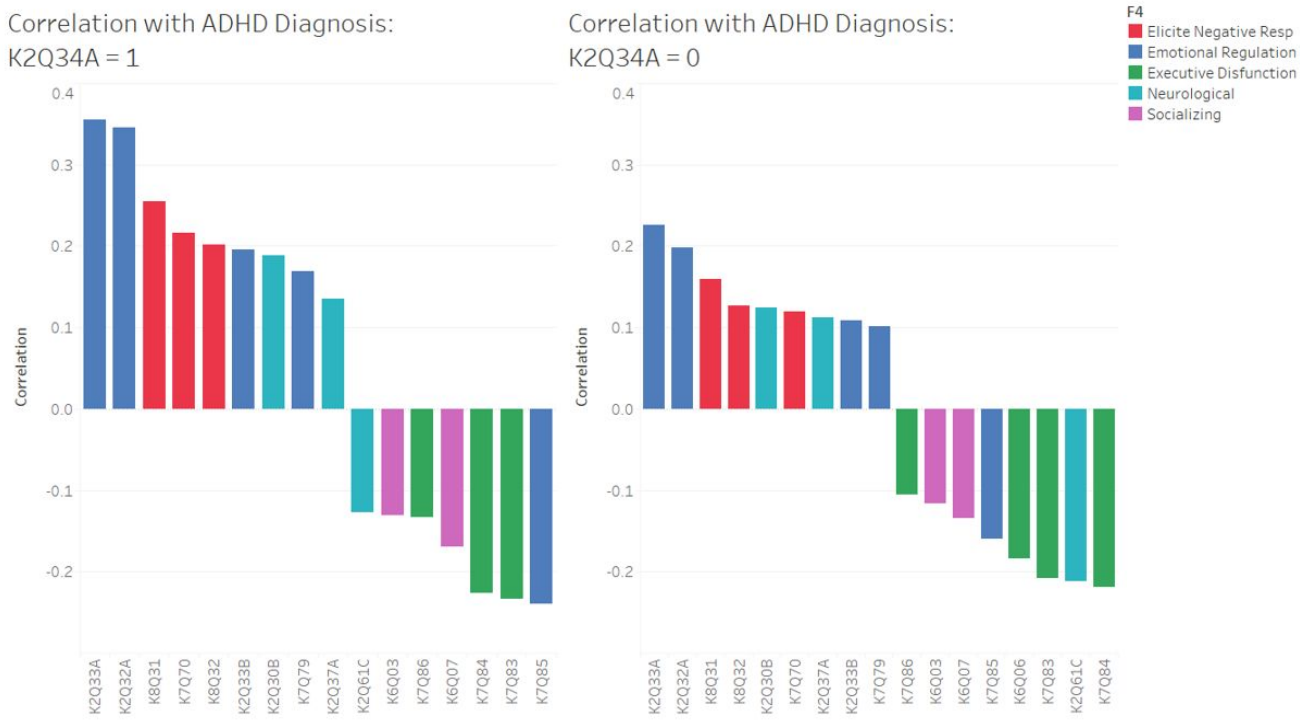
Model Evaluation

Final model numbers on females:

This model was trained on 41314 rows with 56 features identified as important when classifying the subset of children without behavioral and conduct problems. It includes a Random Forest Binary Classifier, and predicted K2Q31A with 0.955 accuracy and 0.935 area ROC. The idea is that you could examine the subset of False Positives (3.39% females) as a higher likelihood of underdiagnosis.

Themes in the Most Important Features:

I found five common themes in the questions that make the best predictors in my model. Note that these 5 themes are found in both children who display behavioral and conduct problems and in those who do not. In contrast, ACEs are important only in children who display behavioral and conduct problems. This is true for both male and female children, which is why it was important to select features based off of children without behavioral and conduct problems vs based off of all females.



Once the features had been selected, I noticed the following themes, along with a few examples of questions in each theme:

The child's tendency to elicit negative feelings from the responder:

- During the past month, how often have you felt angry with [him/her]?
- During the past month, how often have you felt [he/she] does things that really bother you a lot?

Emotional regulation:

- He/She] stays calm and in control when faced with a challenge - negative correlation.
- [He/She] is unhappy, sad, or depressed - positive correlation - (often misdiagnosed in females when it is a symptom of ADHD)

Executive Dysfunction:

- [He/She] finishes the tasks [he/she] starts and follows through with what [he/she] says [he'll/she'll] do
- [He/She] does all required homework

Neurological Related:

- Has a doctor or other health care provider ever told you that [S.C.] had speech or other language problems?
- How would you describe [his/her] usual ability to walk?

Socializing:

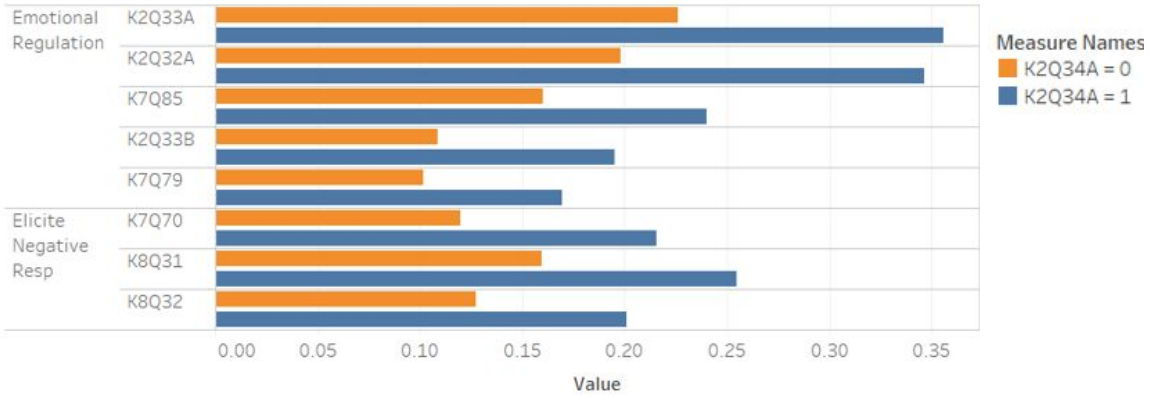
- Are you concerned a lot, a little, or not at all about how [he/she] gets along with others

Variable	K2Q34A = 1	K2Q34A = 0	Description
K2Q33A	0.3554771766	0.2259161567	Has a doctor or other health care provider ever told you that [S.C.] had anxiety problems?
K2Q32A	0.346150616	0.1977519008	Has a doctor or other health care provider ever told you that [S.C.] had depression?
K8Q31	0.254650235	0.159554612	During the past month, how often have you felt [S.C.] is much harder to care for than most children [his/her] age?
K2Q61A_2	0.2401780273	0.1034377208	How old was [S.C.] when you were first told by a doctor or other health care provider that [he/she] had cerebral palsy? (Unit of Measure)
K2Q36A	0.2306746923	0.1577450341	Has a doctor or other health care provider ever told you that [S.C.] had any developmental delay?
K7Q70	0.2158002551	0.1200012887	[He/She] argues too much
K8Q32	0.2012028542	0.1272143591	During the past month, how often have you felt [he/she] does things that really bother you a lot?
K2Q33B	0.1953133537	0.1086303816	Does [S.C.] currently have anxiety problems?
K2Q30B	0.1885354196	0.1246886036	Does [S.C.] currently have a learning disability?
K7Q79	0.1691140364	0.1019005966	[He/She] is unhappy, sad, or depressed

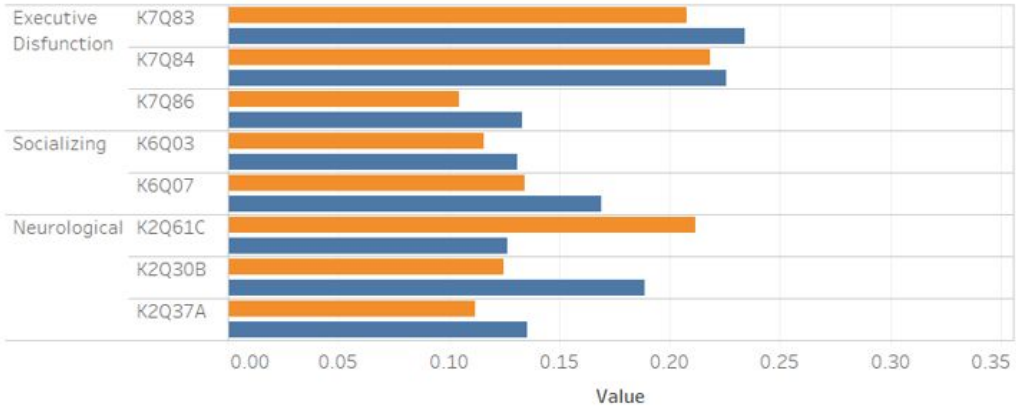
K2Q37A	0.1354453724	0.1120150336	Has a doctor or other health care provider ever told you that [S.C.] had speech or other language problems?
K2Q61C	-0.1266573969	-0.2117332926	How would you describe [his/her] usual ability to walk?
K6Q03	-0.1310884082	-0.1160859328	Are you concerned a lot, a little, or not at all about how [he/she] understands what you say?
K7Q86	-0.1329936667	-0.1049852036	[He/She] shows interest and curiosity in learning new things
K6Q07	-0.169101322	-0.1343858152	Are you concerned a lot, a little, or not at all about how [he/she] gets along with others?
K6Q06	-0.1982191395	-0.1839477925	Are you concerned a lot, a little, or not at all about how [he/she] behaves?
K7Q84	-0.2257916287	-0.2186283734	[He/She] finishes the tasks [he/she] starts and follows through with what [he/she] says [he'll/she'll] do
K7Q83	-0.233880047	-0.2079687067	[He/She] does all required homework
K7Q85	-0.2400000794	-0.1600068875	[He/She] stays calm and in control when faced with a challenge

To summarize, these children are more likely to exhibit unhealthy emotional regulation and executive dysfunction. These children are more likely to elicit negative responses from the responders on their relationship. Further research does support this, as ADHD researchers estimate that by age 12, children with ADHD get 20,000 more negative messages about themselves than other kids their age. They have comorbidity with other neurological health issues.

However, the data shows that children who display behavioral and conduct problems (K2Q34A = 1) have a much stronger tendency towards Emotional Regulation and Eliciting Negative Responses than children who do not.



In contrast, trouble with executive functioning, socializing, and neurological disorders show up more evenly in those with and without behavioral and conduct problems.



Discussion and Conclusions

I interviewed Rebecca Delgado for this project to learn about potential survey questions the CDC could include in future surveys to improve the model. Rebecca holds a Master's degree in Applied Learning and Instruction with a focus on motivation and cognition. She has completed comprehensive training as an ADHD Life Coach through ADDCA (ADD Coach Academy).

She is passionate about awareness of underdiagnosis of ADHD in females and was glad to hear about this project. Her experience and research backed up my findings about females masking and those who don't display behavioral and conduct problems going undiagnosed.

These are her tips on signs to look for in female children, perhaps the CDC would be interested in collecting data on some of these variables:

- Unusual sensitivity to noise or light
- Chronic lateness
- Poor sense of time
- Strong resistance to switching activity
- Noticeable oversharing
- Repetitive body movements such as teeth grinding (bruxism), pacing, distracting pen-tapping, issues caused by restless legs, and excessive fidgeting
- Social anxiety or lack of self-esteem (vs aggression)

When I asked her what people with ADHD need the most help with in her experience, she replied "acceptance and permission." Once someone understands their disorder, they can stop trying to fit a square peg in a round hole and design a life that works for them.

Acknowledgments

I would like to thank Rebecca Delgado for her insight and enthusiasm, I really enjoyed our conversation and appreciate your time! I would also like to thank my professor Catie Williams for an excellent course and for helping me along the way. It was very fun to have the opportunity to do a project that is near and dear to my heart. Many thanks also to my classmates for your feedback.

Appendix

Data Source: 2011 - 2012 National Survey of Children's Health:
https://ftp.cdc.gov/pub/Health_Statistics/NCHS/slaits/nsch_2011_2012/

Github with Project Repository:
https://github.com/Holly-E/ADHD_Study

Centers for Disease Control and Prevention, National Center for Health Statistics. (2013, April). "*2011-2012 National Survey of Children's Health Frequently Asked Questions."* Available from: http://www.cdc.gov/nchs/slaits/nsch.htm

Formatted Frequencies of Survey Responses available from:
https://ftp.cdc.gov/pub/Health_Statistics/NCHS/slaits/nsch_2011_2012/04_List_of_variables_and_frequency_counts/

Survey questions used to obtain the variables available from:
https://www.cdc.gov/nchs/data/slaits/2011NSCHQuestionnaire.pdf

References

Birnbaum HG, Kessler RC, Lowe SW, et al. "Costs of attention deficit-hyperactivity disorder (ADHD) in the US: excess costs of persons with ADHD and their family members in 2000." Curr Med Res Opin. 2005;21(2):195-206. doi:10.1185/030079904X20303. Retrieved from https://pubmed.ncbi.nlm.nih.gov/15801990

Bowman, S. (2020). "ADHD in women - underdiagnosis." Retrieved from https://www.rtor.org/2020/01/15/overdiagnosis-or-underdiagnosis-of-adhd/

Castellanos FX, Lee PP, Sharp W, et al. "Developmental Trajectories of Brain Volume Abnormalities in Children and Adolescents With Attention-Deficit/Hyperactivity Disorder." JAMA. 2002;288(14):1740–1748. doi:10.1001/jama.288.14.1740 Retrieved from https://jamanetwork.com/journals/jama/article-abstract/195386

Chadd. "ADHD and Long-Term Outcomes." Retrieved June, 2020 from https://chadd.org/about-adhd/long-term-outcomes/

Chadd. "The Science of ADHD ." Retrieved June, 2020 from https://chadd.org/about-adhd/the-science-of-adhd/

Connolly,M.(2019). *"ADHD in Girls: Why It's Ignored, Why That's Dangerous."* Retrieved from https://www.additudemag.com/adhd-in-girls-women/.

FiveThirtyEight. (Sept 19, 2014) "How Many Adults Take ADHD Drugs?" Retrieved from https://fivethirtyeight.com/features/dear-mona-how-many-adults-take-adhd-drugs/

Guilherme Polanczyk, M.D., Maurício Silva de Lima, M.D., et al. "The Worldwide Prevalence of ADHD: A Systematic Review and Metaregression Analysis." American Journal of Psychiatry 2007 164:6, 942-948. Retrieved from https://ajp.psychiatryonline.org/doi/full/10.1176/ajp.2007.164.6.942

MASON, O. and ROSIER, T. (Feb 18, 2020). "WHAT CAUSES ADHD? Face It — People with ADHD Are Wired Differently." Retrieved from https://www.additudemag.com/current-research-on-adhd-breakdown-of-the-adhd-brain/

Paul, J. (2018, April 11). *"How it really feels to be time-blind with ADHD."* Retrieved from https://adhdhomestead.net/time-blindness-feels/

Philip Shaw, Mary Gilliam, et al. "Cortical Development in Typically Developing Children With Symptoms of Hyperactivity and Impulsivity: Support for a Dimensional View of Attention Deficit Hyperactivity Disorder." American Journal of Psychiatry 2011 168:2, 143-151. Retrieved from https://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.2010.10030385

Philip Shaw, Argyris Stringaris, Joel Nigg, and Ellen Leibenluft. "Emotion Dysregulation in Attention Deficit Hyperactivity Disorder." American Journal of Psychiatry 2014 171:3, 276-293. Retrieved from https://ajp.psychiatryonline.org/doi/pdf/10.1176/appi.ajp.2013.13070966

U.S. EPA Office of Research and Development. (Jan 18, 2020). "Childhood Chemical Exposures and ADHD." Retrieved from https://catalog.data.gov/dataset/childhood-chemical-exposures-adhd