**Midterm Project Proposal**

Kosuke SASAKI

November 4, 2021

**Personal Statement**

My career goal is to work for the public health field or healthcare company, and this will require knowledge and skills related to data analysis, epidemiology, and medicine. To acquire and further deepen these skills, I will work on large birth cohort data which comes from Japan Environment and Children’s Study (JECS). This data set contains over 100,000 mother-child pair observations and over 100 variables which includes their environmental data, personal health records, and other possible risk factors related to various child’s diseases and developmental disorders.

For the midterm project, I will use this data set and assess i) the association between family socioeconomic status (SES) and autism spectrum disorder (ASD) of the children (or any other developmental disorder) and ii) other possible risk factors of ASD. This will maximize the use of data analytical skills learned in this program and advance these skills to handle large data set and explore any other health-related risk factors.

**Question**

The prevalence of ASD has increased substantially over the past decades, from 6.7 in 2000 to 18.5 in 2016 per 1000 children.[[1]](#endnote-1) Among all the risk factors of ASD, up to 40–50% of the variance in ASD liability is determined by environmental factors, and evidence to date has proved possible risk factors such as advanced parental age, pregnancy-related factors, and maternal diseases.[[2]](#endnote-2) However, the association between family SES and child’s ASD is still to be clear and is inconsistent.

In this project, I would like to answer the following question: is family SES associated with the development of child’s ASD?

**The data source**

JECS is a birth cohort observational study launched in 2011 with participants of 100,000 mother-child pairs residing in 15 regions throughout Japan. It measures the environmental exposures including social environment during pregnancy and through childhood, while examining children's health periodically with questionnaires, biological samples and personal health records until they become 13 years of age.[[3]](#endnote-3)

The data set used in this project will be extracted from JECS, and it will include about 100,000 observations and more than 100 variables, including family’s SES such as income, parental job category and educational status, and the presence of child’s ASD. The data is expected to be available in a week because it requires data issuance procedures.

**Proposed Timeline of work**

* EDA: November 5 - 12
* Data Processing: November 12 - 19
* Modeling and Validation: November 19 - 26
* Write up: November 26 - December 2

1. Autism Spectrum Disorder (ASD): Center for Disease Control and Prevention  
   https://www.cdc.gov/ncbddd/autism/data.html [↑](#endnote-ref-1)
2. Modabbernia, A., Velthorst, E., & Reichenberg, A. (2017). Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. Molecular autism, 8, 13.   
   https://doi.org/10.1186/s13229-017-0121-4 [↑](#endnote-ref-2)
3. Japan Environment and Children’s Study  
   https://www.env.go.jp/chemi/ceh/en/index.html [↑](#endnote-ref-3)