Report of MA678 Midterm Project

Kosuke Sasaki

2021/12/11

Abstract

Autism spectrum disorder (ASD) has increased substantially over the decades, and some environmental risk factors has been identified. However, the association between parental socioeconomic status (SES) and child's ASD is unknown. In this project, large birth cohort data set was analyzed to examine the association, and the odds of children with high parental SES developing ASD was found to be significantly higher than that with low parental SES.

Introduction

The prevalence of ASD has increased substantially over the past decades, from 6.7 in 2000 to 18.5 in 2016 per 1000 children, and ASD accounted for 7.7 million disability-adjusted life years (DALYs) in 2010, which was the leading mental cause of disability in children under five years old in terms of years lived with disability (YLDs).[1,2] Although up to 40–50% of the variance in ASD liability is determined by environmental factors and several pieces of evidence are presented for possible risk factors such as advanced parental age and maternal diseases, the association between SES and ASD is still to be clear and is inconsistent in previous studies.[2]

To address this research question, I used and analyzed data set from large Japanese birth cohort, Japan Environment and Children's Study (JECS). JECS is a nationwide epidemiological study involving 100,000 mother-child pairs living throughout Japan, and the data was collected in 15 regions throughout the country.[3] I used the data of 40438 children in JECS who responded to the 3-year-old questionnaire, which included a question about ASD diagnosis, and fit the multiple logistic regression model to see if SES is related to child's ASD.

Method

Data wrangling

I applied for the data provision to National Institute for Environmental Study in Japan, and the data set was extracted from JECS data server, which is not public for now. Two raw data sets were extracted from the JECS server which included 104062 observations and different variables; one includes id, district, child's sex, Income, parental age and parental ASD history, and the other includes id and child's ASD.

After I merged the two data sets based on id as an identifier, I recategorize income values to align with the objective of this project. For the income category, there were originally nine levels of classification, from Category 1 which corresponds to an annual income of less than 2 million yen, to Category 9 which corresponds to more than 20 million yen. In this project, I will look for the relationship between SES and ASD, and particularly see if low-SES or high-SES could be related with high odds of ASD. In this perspective, I need to recategorize income into 3 levels of classification, low, middle and high, as a proxy of SES, and I defined the categories below the bottom quintile of income as low income, the categories above the top quintile of income as high income and the others as middle income, which also follows the income category of Cabinet Office in Japan.[4]

Secondly, I assumed missing completely at random and removed observations with missing value for ASD, sex, Income, paternal age, and parental ASD and observations with sex value of 3 categorized as "unknown", which reduced original 104062 observations to 40438 observations. Finally, I converted variable type for "ASD", "sex","fASD","mASD","unit_no" from numeric to factor, assigned appropriate label or name to variable name and values, and assigned levels in ascending order to income category. Each variable name and explanation used in this project is shown below.

column names	explanation
NO	participant's id
district	location where the participant's data is collected (one of 15 districts)
ASD	whether the child develops ASD
sex	child's sex (male or female)
Income	family's income (low, middle, or high)
fathersage	paternal age at the time of delivery
momsage	maternal age at the time of delivery
fASD	whether the father develops ASD
mASD	whether the mother develops ASD

Exploratory Data Analysis

Contingency table for child's sex and ASD is shown below. ASD seems to be more frequent for male compared to female in this cohort, which is already proved and accepted in the world.[5]

sex	Child with ASD	Child w/o ASD	Total
Male	20501 (99.4%)	131 (0.6%)	20632
Female	19762 (99.8%)	44 (0.2%)	19806
Sum	40263 (99.6%)	175 (0.4%)	40438

Contingency tables for parental ASD and child's ASD are shown below. The prevalence of ASD in parents of both sexes is less than 0.03%, which is extremely low compared to the prevalence of 0.4% in children. This might be because adults developing ASD have difficulty in marriage or having children. Based on this table, parental ASD and child's ASD seem to be related since one out of 20 parents developing ASD has a child with ASD, but there is an uncertainty due to the very small sample size of parental ASD.

	Child with ASD	Child w/o ASD	Total
Mother w/o ASD	40254 (99.6%)	174 (0.4%)	40428
Mother with ASD	9 (90.0%)	1 (10.0%)	10
Sum	40263 (99.6%)	175 (0.4%)	40438

	Child with ASD	Child w/o ASD	Total
Father w/o ASD	40254 (99.6%)	175 (0.4%)	40429
Father with ASD	9 (100.0%)	0 (0.0%)	9
Sum	40263 (99.6%)	175 (0.4%)	40438

Then, I will look at the mosaic graph for child's ASD and Income grouped by district as below.

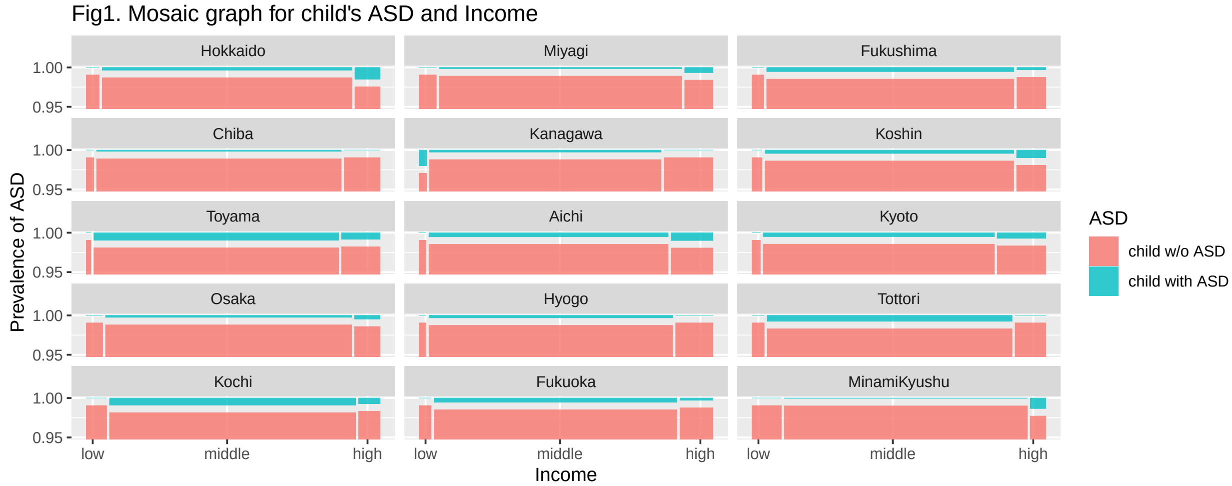


Figure 1: Mosaic graph for child's ASD and Income

Figure 1 shows the prevalence of ASD by income, for each of the 15 regions. Here, note that the y-axis starts at 0.95 because the prevalence of ASD is very low. Although there is variation across regions, the prevalence of ASD in the low-income category seems to be generally lower than in other income categories.

Finally, I look at the relationship between ASD and paternal age as below.

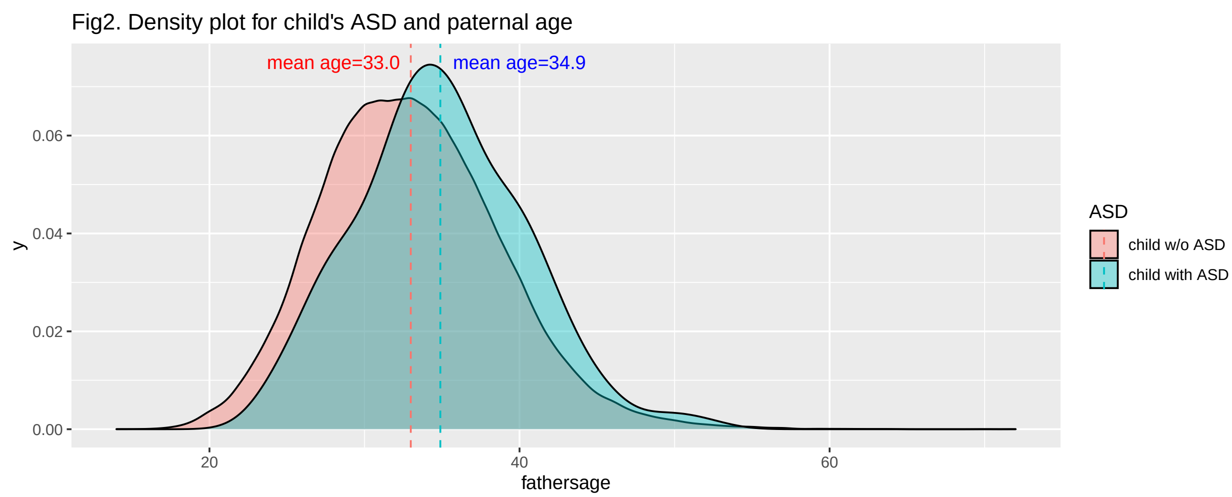


Figure 2: Density plot for child's ASD and paternal age

Figure 2 shows density plot for the paternal age by child's ASD. Based on the graph, paternal age with a ASD child seems to be higher than that with a non-ASD child.

Model Fitting

I decided to use the multilevel logistic regression model to fit the data, selecting ASD as an outcome and Income, sex, parental ASD, and paternal age as predictors. All the predictors excluding income are thought of as risk factors for ASD.[1] As the maternal age is highly correlated with paternal age (Appendix), which will make the model unstable if included, and is still to be investigated as a risk factor for ASD, I will not include maternal age in this model.[6]

Based on the EDA, prevalence of ASD by income seems to be different in different district. In addition, the accuracy and the rate of appropriate ASD diagnosis could differ in different districts which have different facilities and diagnostic skills, which can affect the base ASD prevalence in each district, I assume random effect for the intercept as well as income coefficient. On the other hand, the effect of child's sex, parental ASD, and paternal age on development of ASD is through biological pathway, which cannot be affected by diagnostic skills or social effect. Thus, their effects can be thought of as fixed one. Based on this way of thinking, I fit the model to the data as below.

```
fit <- glmer(ASD ~ Income + sex +fASD + mASD+ fathersage + (1 + Income|district),
             data = eco2, family = binomial)
```

Result

Fixed effects and fitted model are shown below.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-10.17	1.74	-5.86	4.63e-09 ***
Incomemiddle	3.31	1.71	1.94	0.05238 .
Incomehigh	3.51	1.72	2.04	0.04118 *
sexfemale	-1.06	0.18	-6.05	1.46e-09 ***
father with ASD	-28.13	6.29e+06	0.000	1.00000
mother with ASD	3.47	1.13	3.06	0.00221 **
fathersage	0.05	1.28e-02	3.91	9.36e-05 ***

$$logit(\pi) = -10.17 + 3.31 * (MiddleIncome) + 3.51 * (HighIncome) - 1.06 * (female) - 28.13 * (paternalASD) + 3.47 * (maternalASD) + 0.05 * (paternalage)$$

The effect estimates of high income as well as sex, maternal ASD and paternal age were significant at the significance level of 0.05. For example, the interpretation for high income and paternal age are as follows; the odds of children with high family income developing ASD is exp(3.51)=33.4 times as high as that with low income, holding other predictors and district constant; the odds of children whose father is older by one year developing ASD is exp(0.05)=1.05 times as high as that of children whose father is younger holding other predictors constant.

For some districts as an example, effect estimates are shown below (all the results are in Appendix). Though in Hokkaido and Fukushima there are some variation for intercept and income effect, those are close to each other and also close to fixed effect estimates. However, in Kanagawa the intercept is far larger than the others, and the sign of income effect is even reversed.

	(Intercept)	Incomemiddle	Incomehigh	sexfemale	fASDfather with ASD	mASDmother with ASD	fathersage
Hokkaido	-10.899393	3.9716257	4.3557706	-1.065926	-28.12842	3.470207	0.0501224
Fukushima	-9.991406	3.4310217	3.2905623	-1.065926	-28.12842	3.470207	0.0501224
Kanagawa	-6.505435	-0.7286586	-0.6657998	-1.065926	-28.12842	3.470207	0.0501224

Discussion

All the covariates except for paternal ASD were significant. Those results are reasonable because various previous studies shown that male child, parental ASD and paternal older age are risk factors for child's ASD, and those effects are not likely to be affected by region or social environment. In this project the paternal ASD only showed negative effect, which is opposed to what we expected. This can be because the sample size of paternal ASD was only 10 and there was no child's ASD case with paternal ASD, which lead to large standard deviation of the effect estimate. The other results consistent with previous studies might support the validity of this project.

With regard to income, high income was significantly associated with a high prevalence of ASD in this study. This is a trend seen in countries without universal health insurance, such as the United States, and this trend can be partially due to diagnostic bias where the higher the income, the more likely ASD will be diagnosed as the patient with high income will be able to visit a highly specialized medical institution.[7] However, Japan has adopted universal health insurance system, so the possibility of diagnostic bias might be low. On the other hand, etiologically, it is thought that the stress on the mother due to low income can cause epigenetic changes in the fetal brain and increase the risk of autism, and in fact some countries with universal health insurance has shown negative or no association between SES and ASD.[6] In addition, parental age can be a confounding factor for the association between high income and ASD, but the model accounted for parental age so that this confounding is unlikely to be the cause of the positive association between income and ASD. Considering these facts, the positive association is surprising and not what we expected. There might be another diagnostic bias where those parents with higher income are more sensitive to their child's health and tend to take children to a doctor, but this reasoning needs further research.

For the mixed effect, we had negative income effect in Kanagawa. When compared to the effects in other regions, the value seems to be an outlier, but as mentioned earlier, it is consistent with the results for countries with universal health care. This needs further comparison between Kanagawa and other regions with context and other data set.

Finally, this study needs sensitivity analysis. We assumed the missing values are completely at random. However, those who have ASD children might have more difficulty in keeping up with this study and might drop easily, or more firmly stick to this study because this study can be a channel to the specialist and specific treatment for ASD. Most of the dropped observations in this study are due to lack of fathers' answers, which were systematically not collected because of limited data collection period, but we still need sensitivity analysis to show the validity of our results.

Citation

1. Center for Disease Control and Prevention: Autism Spectrum Disorder (ASD) https://www.cdc.gov/ncbddd/autism/data.html

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

3. National Institute for Environmental Study in Japan: Japan Environment and Children's Study https://www.env.go.jp/chemi/ceh/en/index.html

4. Cabinet Office in Japan: Annual Report on the Japanese Economy and Public Finance. https://www5.cao.go.jp/j-j/wp/wp-je09/pdf/09p03022.pdf

5. Center for Disease Control and Prevention: What is Autism Spectrum Disorder? https://www.cdc.gov/ncbddd/autism/facts.html

6. Kelly, B., Williams, S., Collins, S., Mushtaq, F., Mon-Williams, M., Wright, B., Mason, D., & Wright, J. (2019). The association between socioeconomic status and autism diagnosis in the United Kingdom for children aged 5–8 years of age: Findings from the Born in Bradford cohort. Autism, 23(1), 131–140. https://doi.org/10.1177/1362361317733182

7. Khaiman C, Onnuam K, Photchanakaew S, Chonchaiya W, Suphapeetiporn K. Risk factors for autism spectrum disorder in the Thai population. Eur J Pediatr. 2015 Oct;174(10):1365-72. doi: 10.1007/s00431-015-2544-2. Epub 2015 Apr 23. PMID: 26226890.

Appendix

EDA continued

Fig3−1. Mosaic graph for child's sex and ASD

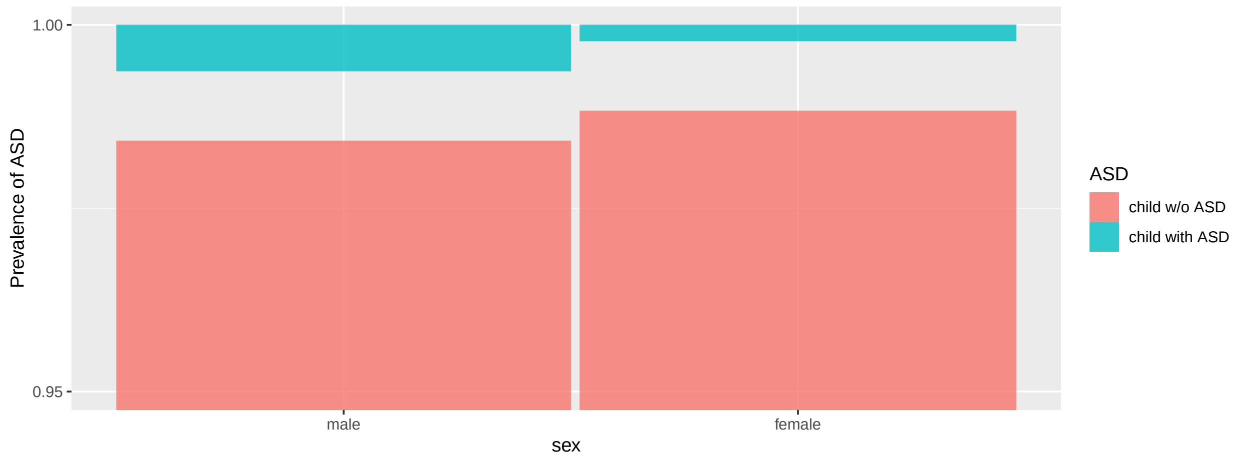


Fig3−2. Mosaic graph for child's sex and ASD by district

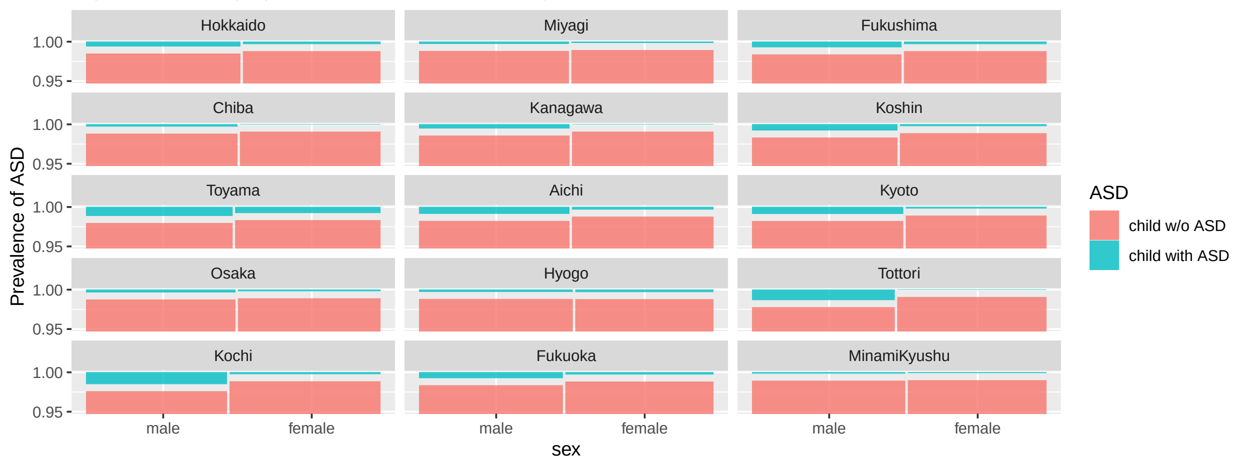


Fig4−1. Mosaic graph for child's ASD and father's ASD

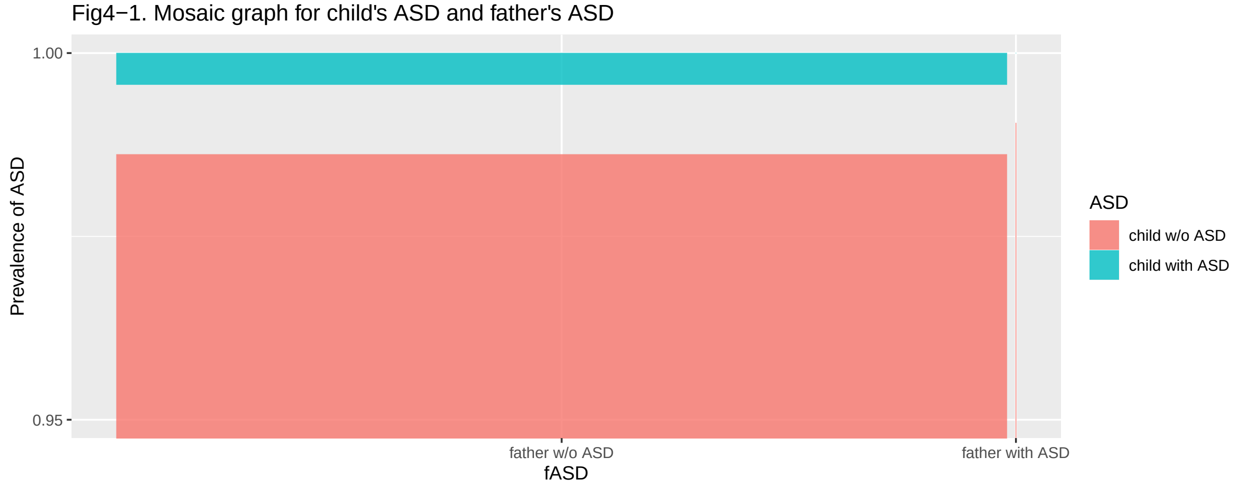


Fig4−2. Mosaic graph for child's ASD and father's ASD by district

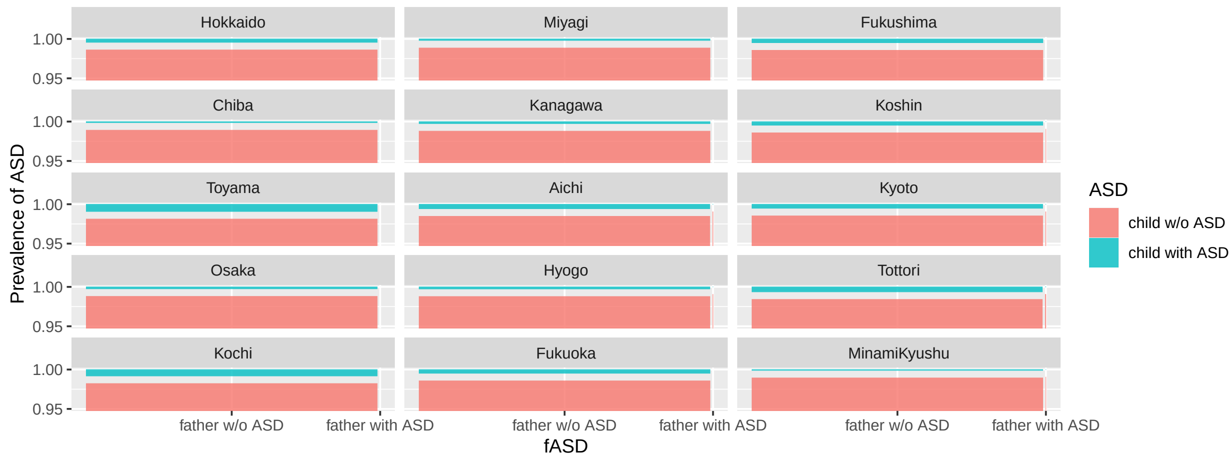


Fig5−1. Mosaic graph for child's ASD and mother's ASD

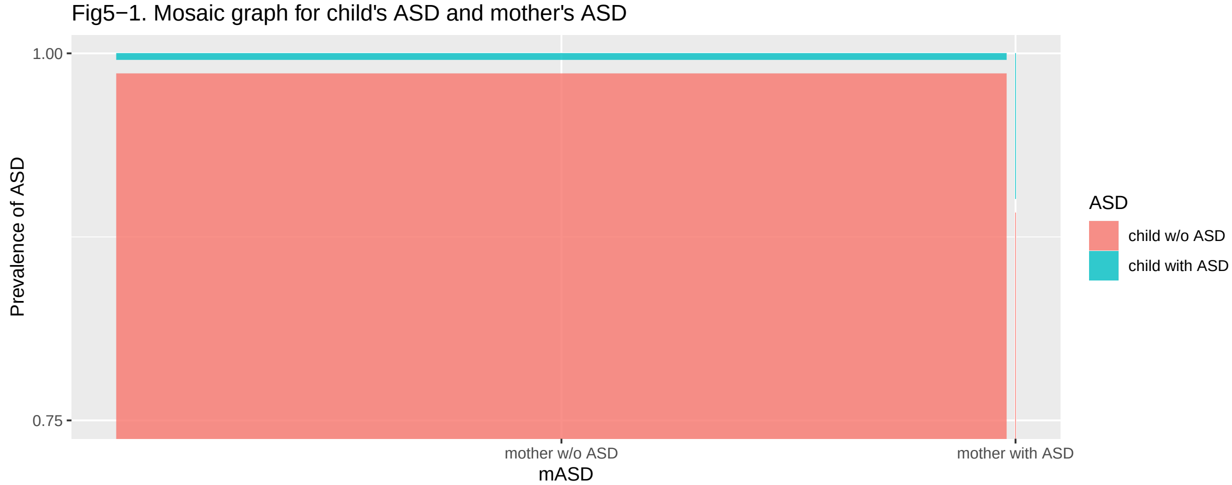


Fig5−2. Mosaic graph for child's ASD and mother's ASD by district

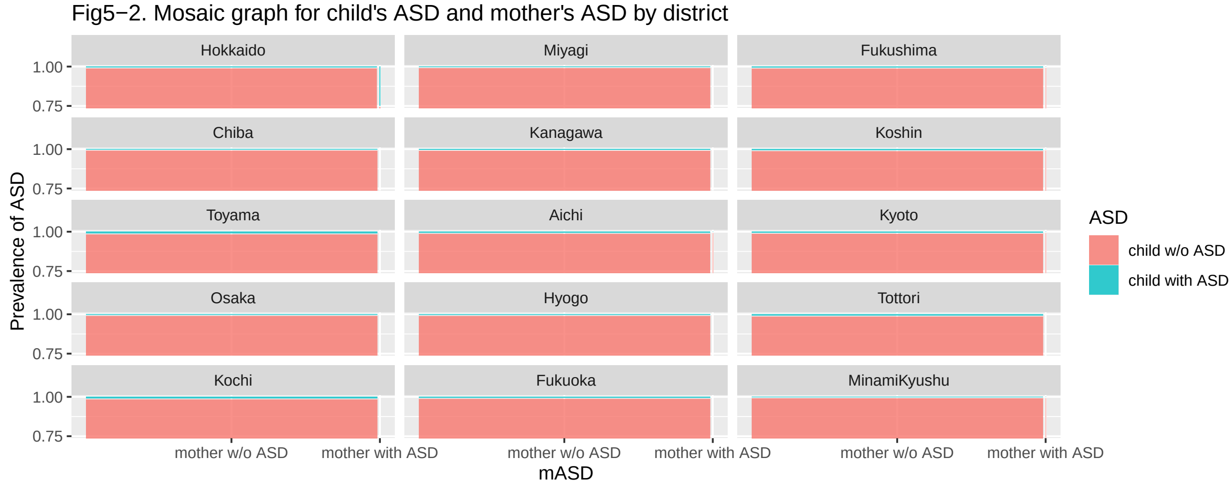


Fig6. Density plot for child's ASD and maternal age

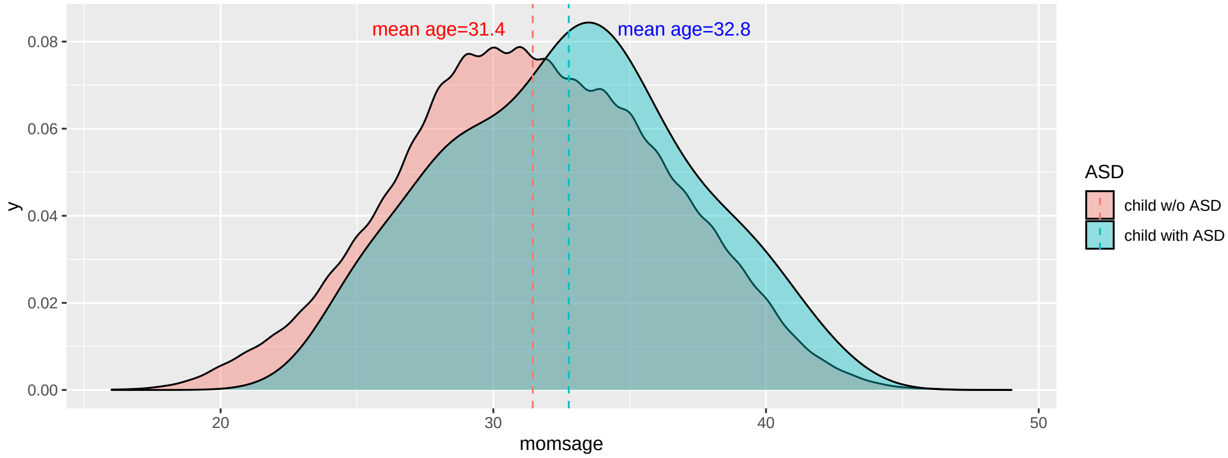
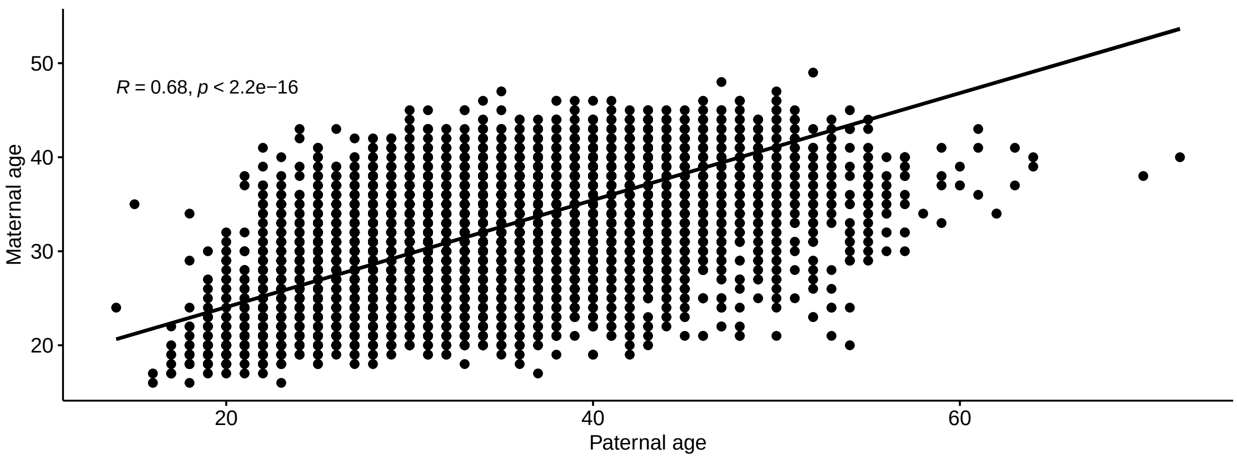
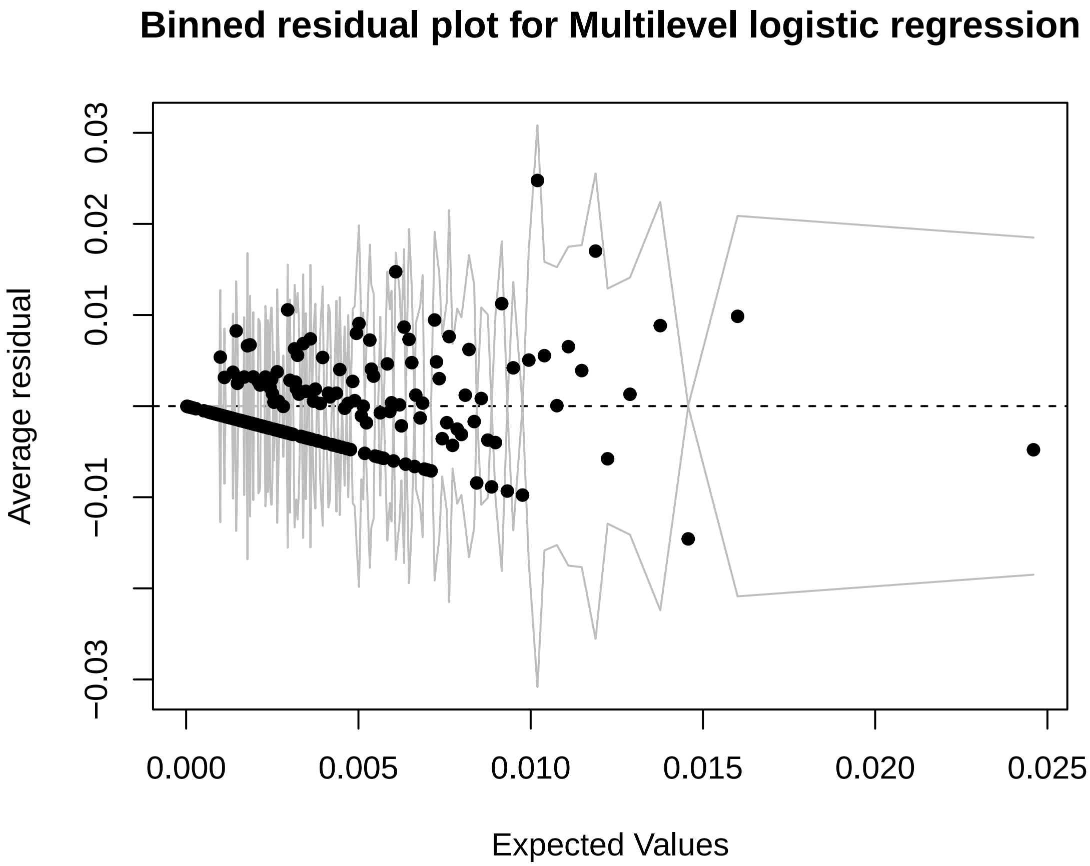


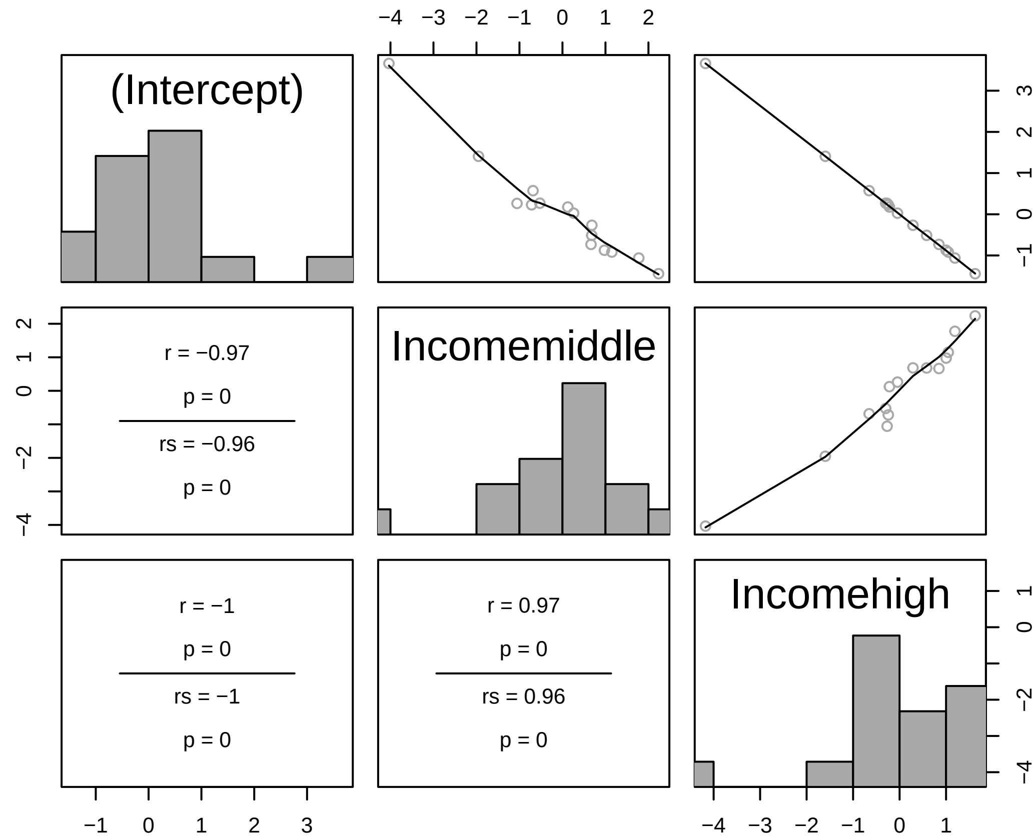
Fig7. Scatter plot for paternal age and maternal age



Model Validation



The pairwise plot of by-district random effects



From the binned residual plots we cannot see any specific shape or deviation from the mean of residual. A lot of residual dots lined up on a line on the lower left in this residual plot just represents there are a lot of 0 values in this data set, which means most of the children do not develop ASD.

The The pairwise plot of by-district random effects suggests each random effect for intercept, middle income and high income follows approximately normal, but might have an outlier;the possible outlier is the district with the highest random intercept, lowest random middle income, and lowest random high income value (which represents top-left-most point in the top-middle plot and top-right plot, or bottom-left-most point in the middle-right plot).

Full Results

Random effects of model

```
## Table of random effect

## $district
##              (Intercept) Incomemiddle  Incomehigh
## Hokkaido      -0.7322002    0.6645463  0.84683606
## Miyagi         0.2293870   -0.7174940 -0.24225986
## Fukushima      0.1757870    0.1239423 -0.21837228
## Chiba          1.4071268   -1.9523450 -1.59750308
## Kanagawa       3.6617582   -4.0357380 -4.17473442
## Koshin        -0.8734857    0.9756929  1.00149809
## Toyama        -1.4431420    2.2352135  1.62545707
## Aichi         -0.9183124    1.1487244  1.04728204
## Kyoto         -0.5121233    0.6809643  0.58196405
## Osaka          0.2692574   -0.5256250 -0.29894707
## Hyogo          0.5727624   -0.6845600 -0.65509828
## Tottori       -0.2670429    0.6839939  0.28776285
## Kochi         -1.0625620    1.7754999  1.19041752
## Fukuoka        0.0265976    0.2599082 -0.04493543
## MinamiKyushu   0.2634459   -1.0565638 -0.26763946
##
## with conditional variances for "district"
```

Fixed effects of model

```
## Table of fixed effect

##         (Intercept)        Incomemiddle          Incomehigh           sexfemale
##        -10.16719285          3.30707939          3.50893457         -1.06592622
## fASDfather with ASD mASDmother with ASD          fathersage
##        -28.12841776          3.47020653          0.05012235
```

Coefficients of model

```
## Table of mixed effect

## $district
##              (Intercept) Incomemiddle Incomehigh sexfemale fASDfather with ASD
## Hokkaido      -10.899393    3.9716257  4.3557706 -1.065926           -28.12842
## Miyagi         -9.937806    2.5895853  3.2666747 -1.065926           -28.12842
## Fukushima      -9.991406    3.4310217  3.2905623 -1.065926           -28.12842
## Chiba          -8.760066    1.3547344  1.9114315 -1.065926           -28.12842
## Kanagawa       -6.505435   -0.7286586 -0.6657998 -1.065926           -28.12842
## Koshin        -11.040679    4.2827722  4.5104327 -1.065926           -28.12842
## Toyama        -11.610335    5.5422929  5.1343916 -1.065926           -28.12842
## Aichi         -11.085505    4.4558037  4.5562166 -1.065926           -28.12842
## Kyoto         -10.679316    3.9880437  4.0908986 -1.065926           -28.12842
## Osaka          -9.897935    2.7814544  3.2099875 -1.065926           -28.12842
## Hyogo          -9.594430    2.6225193  2.8538363 -1.065926           -28.12842
## Tottori       -10.434236    3.9910733  3.7966974 -1.065926           -28.12842
## Kochi         -11.229755    5.0825793  4.6993521 -1.065926           -28.12842
## Fukuoka       -10.140595    3.5669876  3.4639991 -1.065926           -28.12842
## MinamiKyushu   -9.903747    2.2505156  3.2412951 -1.065926           -28.12842
##              mASDmother with ASD fathersage
## Hokkaido                3.470207 0.05012235
```

```
## Miyagi                  3.470207 0.05012235
## Fukushima               3.470207 0.05012235
## Chiba                   3.470207 0.05012235
## Kanagawa                3.470207 0.05012235
## Koshin                  3.470207 0.05012235
## Toyama                  3.470207 0.05012235
## Aichi                   3.470207 0.05012235
## Kyoto                   3.470207 0.05012235
## Osaka                   3.470207 0.05012235
## Hyogo                   3.470207 0.05012235
## Tottori                 3.470207 0.05012235
## Kochi                   3.470207 0.05012235
## Fukuoka                 3.470207 0.05012235
## MinamiKyushu            3.470207 0.05012235
##
## attr(,"class")
## [1] "coef.mer"
```