Autism and ADHD prevalence in Chile through Bayesian prevalence analysis, machine learning and clinical record data linkage

Student 5526, Newnham College

2023-07-14

Declaration: 'This dissertation is submitted for the degree of Master of Philosophy.'

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

The dissertation does not exceed the word limit for the respective Degree Committee.

Word count: xxx TODO (including footnotes, but excluding tables, appendices, and bibliography)

Acknowledgements

TODO

Contents

1 Abstract

TODO

2 Introduction

TODO - copy in intro text

3 Aims

This project aims to better understand the prevalence of autism and ADHD in children in Chile. To that end, its objectives are to:

1. Find a lower bound on the prevalence of autism and ADHD across sex, health service, SES, ethnicity and rurality in Chile from school data using a frequentist method.
2. Identify common characteristics of Chilean children with autism from clinical data using machine learning clustering.
3. Link Chilean school data and clinical data and thus obtain an accurate autism prevalence estimate in one health service, then project prevalence bounds across all health services using Bayesian prevalence prediction.

4 Methods

4.1 Deviation from research protocol

This investigation differs very substantially from the research protocol provided at Appendix X. The protocol intended to investigate autism prevalence in the Cambridgeshire region using clinical data from the Cambridgeshire and Peterborough NHS Foundation Trust and school data from the UK Department for Education's National Pupil Database. Unfortunately this clinical data was found to be of insufficient size and quality to conduct the proposed investigation and while this school data was of high quality, it had been well analysed by Roman-Urrestarazu already (8)(10). Therefore school and clinical data from Chile was used instead as they were of high quality and no research on them could be found. The available data from Chile did not have fields on age at autism diagnosis and it was therefore not possible to progress the research protocol's first aim of using machine learning to analyse autism diagnosis age. The protocol's second aim of school autism diagnoses with clinical diagnoses was adapted to be supplementation of the Chilean school data diagnoses with clinical diagnoses, see below. Additional aims and analysis that were appropriate to the Chile data were developed and are detailed in Section Aims and the rest of this section.

4.2 Data management

The statistical software package R has been used to clean, link, analyse and visualise data. R analysis scripts and version control are managed in GitHub and are available at https://github.com/delatee/Autism-diagnosis-age-ML. Raw data and analysis outputs have not been uploaded to GitHub. All data is stored on local devices and will be deleted at conclusion of the project.

4.3 Data collection

4.3.1 School data

This research uses data from the Chilean Government's Ministerio de Educación (Ministry of Education) that was provided under a freedom of information request and Chile's Law 21545 (28). This dataset contains anonymised records of 3.6 million students in all Chilean schools in 2021. It was collected and curated by the Chilean Unidad de Estadísticas (Statistics Unit), Centro de Estudios (Study Centre) and Ministerio de

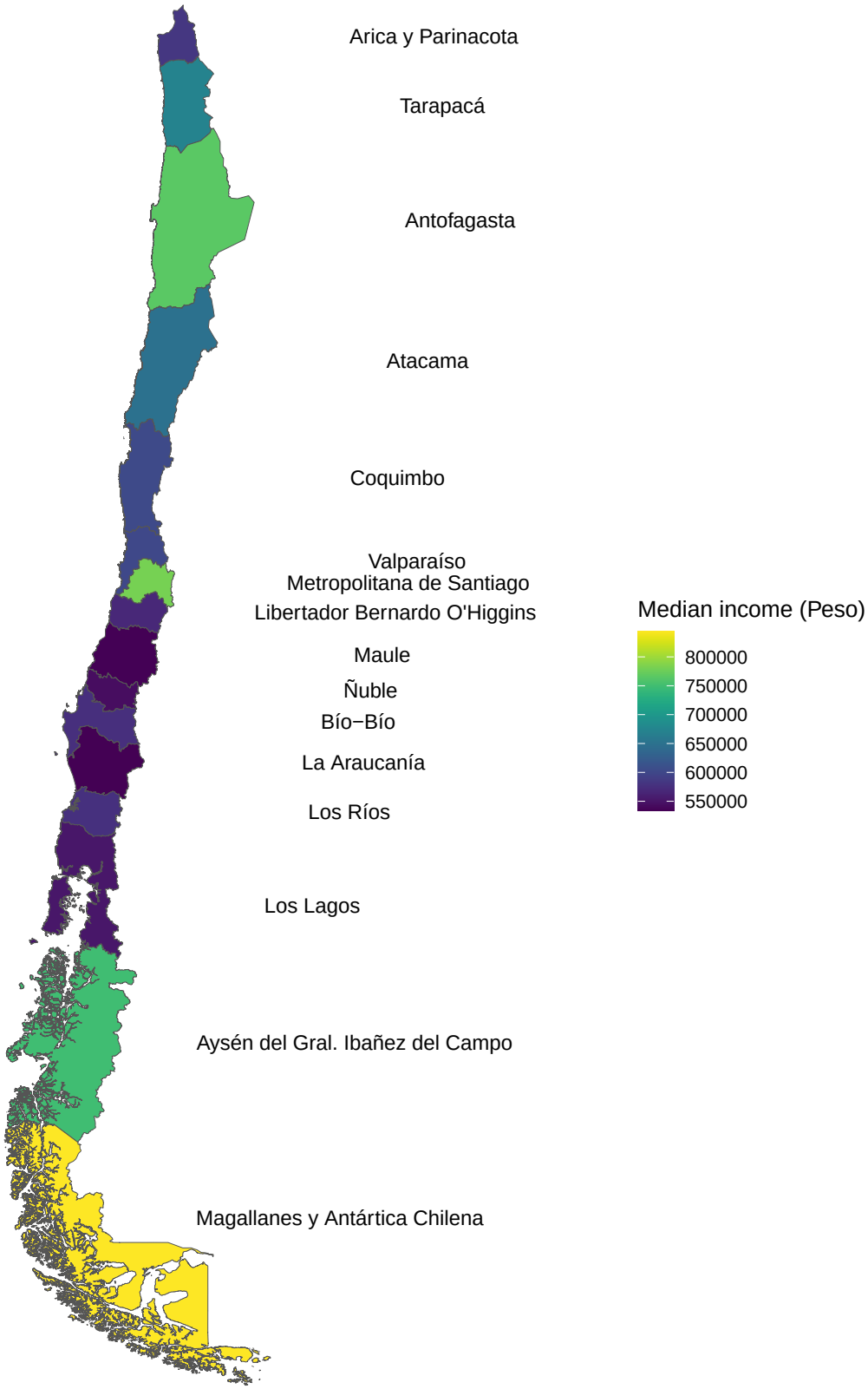

Figure 1: Net income from main job in Chile in 2021 by region, from the INE's Supplementary Income Survey.

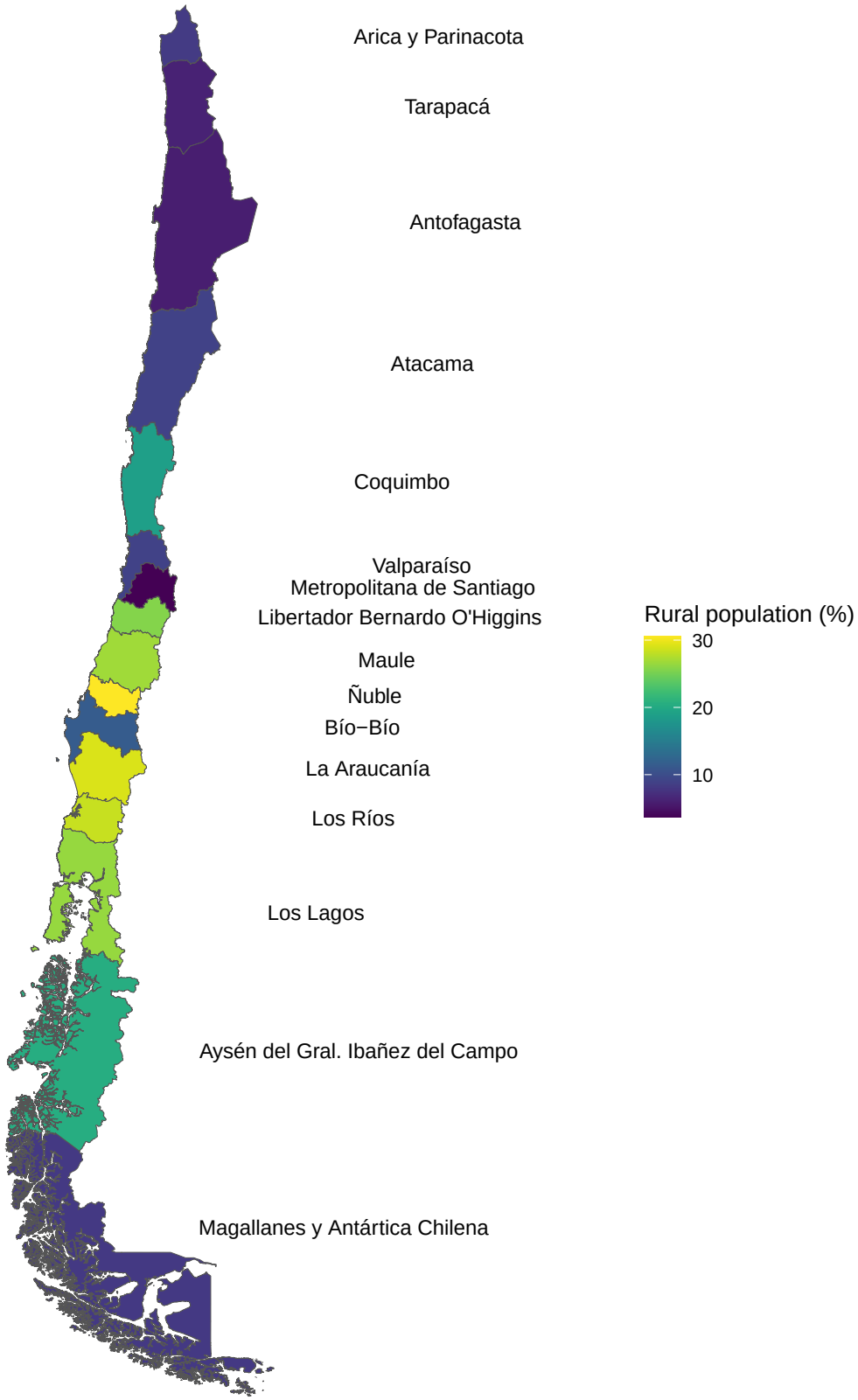

Figure 2: Percentage of population living in rural areas in Chile in 2017 from 2017 census.

Table 1: Chilean health services by region

Region	Health services
Antofagasta	Antofagasta
Arica y Parinacota	Arica
Atacama	Atacama
Aysén del Gral. Ibañez del Campo	Aisén
Bío-Bío	Arauco
Bío-Bío	Biobío
Bío-Bío	Concepción
Bío-Bío	Talcahuano
Coquimbo	Coquimbo
La Araucanía	Araucanía Norte
La Araucanía	Araucanía Sur
Libertador Bernardo O'Higgins	Libertador B.O'Higgins
Los Lagos	Chiloé
Los Lagos	Osorno
Los Lagos	Reloncaví
Los Ríos	Valdivia
Magallanes y Antártica Chilena	Magallanes
Maule	Maule
Metropolitana de Santiago	Metropolitano Central
Metropolitana de Santiago	Metropolitano Norte
Metropolitana de Santiago	Metropolitano Occidente
Metropolitana de Santiago	Metropolitano Oriente
Metropolitana de Santiago	Metropolitano Sur
Metropolitana de Santiago	Metropolitano Sur Oriente
Tarapacá	Iquique
Valparaíso	Aconcagua
Valparaíso	Valparaíso San Antonio
Valparaíso	Viña del Mar Quillota
Ñuble	Ñuble

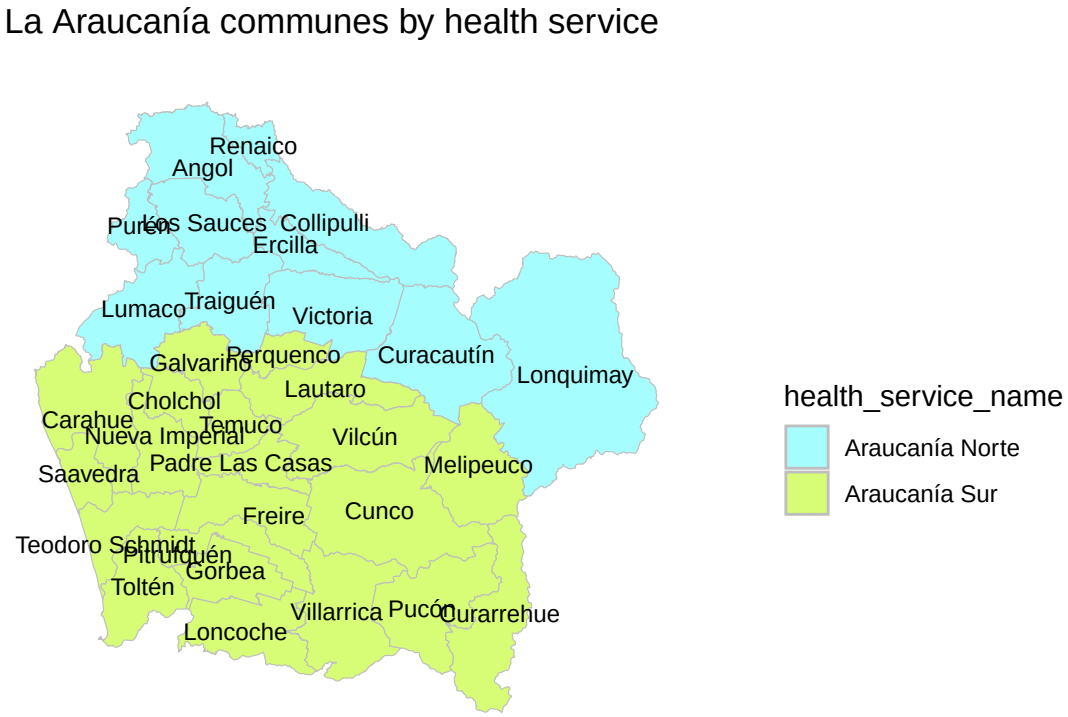

Figure 3: Communes in the Araucanía region, coloured red for the Araucanía Norte (north Araucanía) health services and blue for the Araucanía Sur (south Araucanía) health service.

Educación and is housed in the Sistema de Información General de Estudiantes (General Student Information System, SIGE). It includes data on school type and location, student characteristics, academic performance, special needs and monthly school fee contributions. This dataset will be referred to as the school data, see Supplementary Table 25 for details on its features.

The school data shows whether students have accessed the Subvención de Educación Especial Diferencial (Differential Special Education Grant, SEED), which is provided to students with severe physical or mental disabilities, including autism and ADHD, that require education adjustments such as specialist schools or small class sizes (29). Specialists in the SEED network of providers conduct an external assessment of students that apply for SEED funding to determine whether they have a special educational need (SEN). The school data includes four groups of students who have autism but are recorded in this dataset as not having autism: students with autism who attend private schools which are not eligible for SEED; students with autism who choose not to apply for SEED; students with autism who apply for SEED but are found to not be sufficiently severely affected to receive SEED; and students with autism who receive SEED but for a different disability as only one can be recorded, perhaps one that more severely affects them. The number of students in each of these groups is unknown and these groups are analogous for students with ADHD. Thus all estimates of autism and ADHD prevalence from this dataset alone are likely to be underestimates of the true population prevalence.

4.3.2 Clinical data

This research also uses data from Chile's Servicio de Salud Araucanía Sur (South Araucanía health service, SSAS), obtained under a freedom of information request. These data are collated from secondary care clinical records, particularly from mental health community care services. They comprise clinical records for public sector specialist health visits of patients aged 6-18 with a primary diagnosis of autism between February 2014 and December 2021 for all communes in the SSES catchment. These data include wage deducted health insurance contributions from Chile's Fondo Nacional de Salud (National Health Fund, FONASA), from which a colleague previously inferred socio-economic status of patients' families. The majority of records are for patients resident in the SSAS catchment area and they do not include any records for privately provided healthcare. These data will be referred to as the clinical data.

A subset of the clinical data, those for appointments provided by Villarrica Hospital in the Villarrica commune of Chile's Araucanía region, were manually validated prior to this investigation. Validation involved a neurologist and a psychiatrist in the SEED network of providers checking the clinical records against central health records to confirm individuals did have autism, adding secondary and tertiary diagnoses where relevant, and adding demographic information including ethnicity, rurality of residence, disability status and experience of foster care status where available. We can be very confident that patients in this subset do have autism. Most patients in this subset were resident in Loncoche, Pucón and Villarrica communes in the SSAS catchment area. These data will be referred to as the validated clinical data, see Figure 4.

4.3.3 Additional data

Chile's 2017 census data, held by the Instituto Nacional de Estadisticas (National Statistics Institute, INE) (30), was used to create a standard population of Chile's age and sex distribution from projections of population size by age and sex in 2021. It was also used to obtain projections of population of youth aged 0-14 in 2021 by region.

Data from the INE's Encuesta Suplementaria de Ingresos (Supplementary Income Survey, ESI) of 2021 (31) was used to obtain the nominal median income by region in 2021.

For mapping, shapefiles of administrative areas were obtained from The Humanitarian Data Exchange of the United Nations Office for the Coordination of Humanitarian Affairs (32). Additional region and commune naming information was taken from the R package 'chilemapas'.

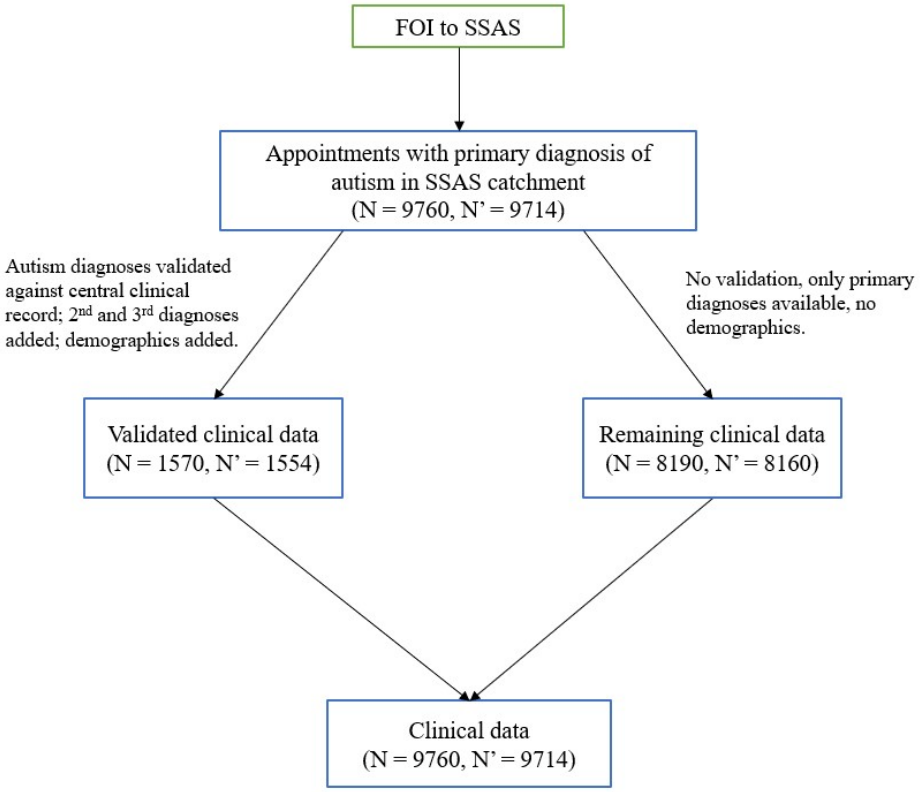

Figure 4: Clinical data pre-processing pathway. N = number of appointment records, N' = number of appointment records for patients resident in SSAS communes.

4.4 Aim 1: Use school data and frequentist method to find a lower bound on autism and ADHD prevalence

Using the school data, the crude prevalence of autism and ADHD was found. Then frequentists methods were used to find the age- and sex-adjusted autism and ADHD prevalence.

4.4.1 School data preparation

The school dataset was restricted to students aged 6-18 as of 30th June 2021 to capture children of school age in Chile. No restriction on sex was necessary as it contained only females and males. Students' commune of residence was mapped to their respective health service catchment areas. Two issues due to boundary changes were corrected. Firstly, the commune now called Tocopilla which falls across the Antofagasta and Tarapacá regions was formerly two communes, Tocopilla in Antofagasta and Pozo Almonte in Tarapacá. The old names have been retained to ensure appropriate region mapping and they are mapped to Antofagasta and Iquique health services respectively. Secondly, the communes recorded as belonging to the former Ñuble sub-region of the Bío-Bío region were mapped to their corresponding communes in the recently formed Ñuble region.

Commune of residence was missing for 4078 students (0.13%). The commune of their school was imputed as most students are likely to go to school near their place of residence.

Students' age as of 30th June 2021 was mapped to age-bands of 6-8, 9-11, 12-14 and 15-18, preserving the divide between primary and secondary school as secondary school starts at age 12 in Chile, and with a larger final band as fewer students are expected to be diagnosed older and the 18 years old group may be small due to students leaving schooling. Students' ethnicity was mapped to being a member of the Mapuche Indigenous group, being a member of another Chilean Indigenous group, or not being a member of an Indigenous group based on their recorded ethnicity which can take at most one value. The 6859 students (0.22%) with ethnicity recorded as 'no registry' were mapped to not being a member of an Indigenous group.

Students' school fee status was mapped to a proxy socio-economic status feature. Students with free schooling were given SES of 1 indicating low status as families with low SES are entitled to schooling rebates. Students paying Chilean Peso $1,000-$100,000 monthly were given SES of 2, indicating medium status, and students paying more than $100,000 monthly were given 3 to indicate high status. For school fee status, 49973 students (1.64%) were coded as 'No information' and 29 students (0.00095%) were missing school fee values and these were re-categorised as 'No information' to reflect their unknown school fee status. No other features of interest for this analysis had missingness.

4.4.2 Crude prevalence

The crude prevalences of autism and of ADHD were calculated as the count of cases divided by the relevant population size. Grouping features of sex, health service of residential commune, monthly school fee as a proxy for SES, ethnicity and school's rurality were used, and calculations were made for aggregate, by sex, and by sex and age group.

4.4.3 Frequentist age and sex adjustment

For each grouping feature, prevalence was standardised by age and sex using direct standardisation following the method presented by Fay and Feuer (33). The 2017 Chile census projections of the distribution of people by age and sex in 2021 was used as the standard population. Gamma confidence limits were calculated at the 95% level using chi-squared distributions. The adjusted prevalence rates were multiplied by their respective population sizes to give adjusted prevalence counts which were rounded to integers, therefore loosing a small amount of precision.

Ethnicity analysis compared individuals from the Mapuche Indigenous group to those from other Indigenous groups or with no Indigenous group and used data from the Araucanía, Aysén, Biobío, Los Lagos, Los Rios, Magallanes and Región Metropolitana de Santiago regions only as these regions are collectively the five

regions with the largest Mapuche populations and the five regions with the highest proportion of Mapuche in their population.

4.5 Aim 2: Use clinical data and machine learning to identify autism diagnosis characteristics

The machine learning technique of multiple component analysis (MCA) was used to explore the feature categories in the validated clinical data that explain its variance.

4.5.1 Clinical data preparation

The full clinical dataset was cleaned to ensure internal consistency in feature values and translated into English where appropriate. Patients were categorised as having autism if any of their diagnosis codes were in ICD F84-F89 for any appointment. Patients were categorised as having an intellectual disability if any of their diagnoses codes was in ICD F70-F79 for any appointment. No restriction on sex was necessary as these data contained only females and males. The clinical data includes economic data in the form of patients' family's health insurance contribution level. Values taken are 'FONASA-A' to 'FONASA-D' which are respectively larger contributions to this public health fund and are assumed to map to increasing socio-economic status, and 'Private health insurance' which indicates contributions to a private health insurance provider rather than the public fund and therefore is assumed to map to the highest socio-economic level. The clinical data was restricted to patients with a diagnosis of autism.

For multiple correspondence analysis, only the validated clinical dataset was used as it contained additional demographic fields not present in the remaining section of the clinical data. In the validated clinical data, patients' age as of 30th June 2021 was mapped to age-bands of 0-2, 3-5, 6-8, 9-11, 12-14 and 15-18. Students below school age were retained for MCA only to more fully explore the variation in the autism patient data. The validated clinical dataset had many values for ethnicity, disability status and foster care status that were recorded as 'no information'. For disability and foster care status, additional features were created with the 'no information' values imputed as 'no disability' and 'no foster care' respectively, as it is likely that patients who did have a disability or had experienced foster care would have this recorded.

The validated clinical data was aggregated to patient level by selecting each patient's most recent commune and rurality, and most common health insurance contribution level, hospital and medical specialty of appointment. For ethnicity, Mapuche was selected if present in any of that patient's appointment records, then Chilean if present, then Foreign if present and else 'No information' was selected. Similarly, disability and foster care status were selected as 'Yes' if recorded for any appointment, then 'No' was selected if present, then 'No information' if no status was recorded for any appointment. Only patients living most recently in a commune in the SSAS catchment area were included.

4.5.2 Multiple correspondence analysis

The machine learning technique of multiple correspondence analysis (MCA) was conducted using R's FactoMineR package with the number of output dimensions equal to the number of input features. MCA was an appropriate technique to use here because the small clinical dataset contains primarily categorical features which MCA is designed to handle.

Data-led analytics was used to identify features in the small clinical dataset that explained the variance in this patient-level data. Initially, all available features with suspected association with autism were included in MCA, specifically: sex, age group, commune of residence, health insurance contribution level as a proxy for SES, ethnicity, residential rurality, disability status and foster care status. The data grouped well into having information about disability and about foster care and not having information about each, therefore the MCA was rerun with the disability and foster care features exchanged for their respective imputed versions. This identified age, commune and ethnicity as important for explaining the variance, so the MCA was rerun with only these three features.

4.5.3 Alternative machine learning approaches

Other machine learning techniques including component factor analysis (CFA) and principle component analysis (PCA) were also considered to analyse the features associated with autism diagnosis in the small clinical data. Both are powerful methods for uncovering the latent structure of data, however CFA typically requires ordered categorical variables and PCA requires continuous variables. As the commune feature, an unordered categorical variable, was thought to be important to the features associated with autism diagnosis and therefore needed to be included in analysis, using CFA or PCA would require one-hot-encoding of commune which would reduce the appropriateness of these tests.

4.6 Aim 3a: Use machine learning to link school and clinical records

The school data and clinical data were linked using manual and probabilistic record linkage.

4.6.1 School data preparation

For data linkage, the school data was further restricted to students with autism that were living in communes in the SSAS catchment in 2021 to maximise comparability with the clinical data. A false empty record was added to the school dataset before linkage that allowed the algorithm to correctly match on SES. This false record was only used during linkage, did not match to any patient records and was removed before comparing matched and unmatched records. This will be referred to as the SSAS school data

4.6.2 Clinical data preparation

For data linkage, the full clinical dataset was used. It was restricted to appointments for individuals resident in communes in the SSAS catchment as the data is believed to be complete for this catchment area only. It was also restricted to patients aged 6-18 as of 30th June 2021 to maximise compatibility to the school data. Appointment year was not restricted in order to retain more data and thus maximise linkage opportunities, and only patients of female and male sex were present. Patients with familial socio-economic level of FONASA-A were interpreted to be of low SES status and given proxy status of 1 (equivalent to students with free school fees); patients with FONASA-B, FONASA-C and FONASA-D were interpreted to have moderate SES status and given 2 (equivalent to students paying $1,000-$100,000 monthly for state schooling); and patients with private health insurance were interpreted to have high SES and given 3 (equivalent to students paying more than $100,000 monthly).

The clinical dataset had no missingness in the features of interest for data linkage.

The clinical data was aggregated to one row per patient per commune in the SSAS catchment, to maximise the opportunity for matching patients that had moved within this health service. The majority of patients lived in only one commune in the SSAS catchment during the study period. Aggregation used the most common SES value for each patient. The resulting dataset will be referred to as the patient data.

4.6.3 Selection of features for matching

The features available for matching that occurred in both the SSAS school data and the patient data were sex, date of birth, commune of residence, and the proxies for socio-economic status – monthly school fees in the SSAS school data and mode health insurance contributions in the patient data.

4.6.4 Manual record linkage

The SSAS school and patient data were blocked on sex and date of birth to improve match quality and reduce runtime. As both datasets are from trusted, large-scale data collections, it is reasonable to assume sex and dates of birth are highly accurate in both datasets, and it would not be reasonable to accept any proposed matches that do not agree on either sex or date of birth.

Two versions of manual record linkage were tried. First, the SSAS school and patient data were merged with perfect matches required on sex, date of birth, commune of residence and proxies for SES. Second, the SSAS

school and patient data were merged with perfect matches required only on sex, date of birth and commune of residence, as the proxies for SES are known to be approximate and therefore requiring perfect matches on SES is not reasonable.

4.6.5 Probabilistic record linkage

The machine learning technique of probabilistic matching was then used to link the SSAS school and patient data. All possible pairs of blocked matches across these two datasets were generated and agreement weights were calculated for each feature using expectation maximisation, then aggregated into a weight for the pair. Records with missing values were retained as this linkage method is robust to missingness. The similarity comparison method was exact matching for commune of residence, diagnosis with autism and socio-economic status; there was no value in using a string comparison method for commune of residence as all commune names were already standardised and two communes having similarly spelled names does not increase the likelihood of a match between those communes. Linkage was implemented using R's RecordLinkage package. This included consideration of the average frequencies of categories in each feature and estimated errors rates were supplied: the default estimated error rate of 0.01 was supplied for the commune of residence and diagnosis with autism features as they are expected to be fairly accurate features, and an estimated error rate of 0.1 was supplied for the socio-economic status feature to reflect that it is a loosely defined proxy.

Pairs were then selected based on weight to create a 1-1, bipartite, matching between SSAS school records and patients. These matches were examined to ensure a patient that lived in multiple communes had matched to only one SSAS school record.

4.6.6 Alternative record linkage methods

Record linkage using Bayesian methods in which matched status is modelled, as developed by Sadinle (34) and refined by Stringham (35), was considered for record linkage here. This technique is typically more complex and has longer runtimes than probabilistic matching but can more easily enforce one-to-one matching and is particularly well suited to matching names. As the datasets in this investigation did not include names, there was limited benefit from using Bayesian linkage methods.

Record linkage using machine learning to classify matched and unmatched pairs, as explored by Pita et al, was also considered as it is generally thought to be more accurate than probabilistic matching alone (36). However this was not pursued because the datasets under investigation have very few common features to match on, meaning machine learning algorithms would have few options to trial, and there are no known true matches with which to assess the accuracy of machine learning models.

4.6.7 Comparison of matched and unmatched records

For the SSAS school and patient datasets, each record was classified as either matched or unmatched based on whether it appeared in the bipartite matching. The discrete Kolmogorov-Smirnov test was used to compare matched and unmatched records within each dataset for each of the features used for matching, excluding date of birth which has too many categories to have meaningful results. Missing values in the socio-economic status feature were omitted before testing as the Kolmogorov-Smirnov test is not robust to missingness.

Permutation tests were then performed for each of the features tested in each dataset by permuting the matched status 2000 times and recomputing the discrete Kolmogorov-Smirnov test for each permutation. The p-values for the Kolmogorov-Smirnov tests on the observed data were then compared to the distributions of p-values for the permuted data to determine the significance of the observed results.

4.7 Aim 3b: Accurately estimate autism prevalence and project prevalence bounds across health services using Bayesian prevalence prediction

An updated estimate of autism prevalence in SSAS was made and projected across health services. Bayesian random effect models were used to predict autism prevalence for several priors.

4.7.1 Updated autism prevalence estimation

The patient data was de-duplicated to a single row per patient with the commune of their matched record chosen if they had multiple communes of residence. The unmatched patients were aggregated to counts per age and sex group. The full school dataset was restricted to students resident in SSAS, then was aggregated to counts per age, sex and autism status group. It was assumed that the patients that did not match to the SSAS school data, which only includes students with SEED for autism, do exist in the larger school data of students resident in SSAS but that do not have a diagnosis of autism in the school data because they do not receive SEED or do not receive it for autism. Thus the count of students with autism for each age and sex group in the restricted school data was increased by the number of unmatched patients in that age and sex group, and the number of students without autism was decreased by the same amount. This effectively reallocated the appropriate number of SSAS students recorded as not having autism to having autism and retained the school data sample sizes.

The crude and age- and sex-adjusted prevalence of autism was calculated from the updated counts as in sections 4.3.2 and 4.3.3. The difference between the adjusted updated prevalence for SSAS (using the reallocated autism diagnosis figures and thus including the additional autism cases found from linkage) and the adjusted school data prevalence for SSAS (using only students with SEED for autsim) was calculated and will be reference to as the adjusted prevalence delta. It represents the prevalence of individuals with an autism diagnosis in the public health system that do not access SEED for autism.

4.7.2 Prevalence projection

Prevalence was projected for each of the health services other than SSAS by adding the adjusted prevalence delta to the adjusted prevalence for each health service that was calculated from the school data only. This projects the prevalence of individuals with autism that do not access SEED across the health services. It assumes that all SSAS patients exist in the school data and are recorded in this dataset as resident in SSAS communes. Estimated confidence intervals were calculated for the projections by finding a band around the projection of equal width to the 95% gamma confidence interval for each health service's school data adjusted prevalence.

4.7.3 Bayesian prevalence analysis

Bayesian prevalence analysis of autism was conducted for the school data by health service. Bayesian methods are appropriate here because they allow calculation of prevalence with incomplete data. Here two different types of data, school records and patient appointments, have been combined and both are known to be incomplete. Bayesian analysis is robust to this messiness and allows plausible prevalence predictions to be made. It also provide information about the likelihood of these predictions given the observed school data.

To conduct the Bayesian analysis, a random-effects model was constructed with the random effect on health service as follows.

Set

$$y_i = \text{adjusted count of autism cases in health service } i$$

$$n_i = \text{number of students in health service } i$$

$$\theta_i = \text{prevalence of autism in health service } i$$

For each health service i, the model formula is

$$y_i|(n_i, \theta_i) \sim \text{Binomial}(n_i, \theta_i)$$

With

$$\theta_i \sim \text{Beta}(a, b)$$

And posterior distribution

$$\theta_i|(y_i, n_i) \sim \text{Beta}(y_i + a, n_i - y_i + b)$$

Fitting a binomial model required integer valued counts of autism cases. As adjusted case counts were used throughout, the adjusted counts had to be rounded to integer values which caused a small amount of precision to be lost. It is anticipated that this may cause the posterior credible intervals to be slightly wider but is not expected to have a large effect on findings.

4.7.4 Prior selection

Four priors for θ_i were used when fitting the Bayesian prevalence model. First, a conjugate beta prior common to all health services was constructed with the global age- and sex-adjusted prevalence of autism in the school dataset and its standard deviation used respectively as the mean and standard deviation of the prior. This prior was suitable because the global adjusted prevalence in the school data provides a lower bound on the plausible prevalence of autism in Chile.

Second, a conjugate beta prior specific to each health service using the health service specific age- and sex-adjusted autism prevalence in the school data as the prior means and their standard deviations as the prior standard deviations. This prior was suitable because it was extending the previous prior to each of the random effect categories and it reflects the students know to be receiving SEED for autism. On its own this prior is expected to give uninformative posteriors because it is effectively duplicating the information in the sample data, but it will be used as a more specific lower bound on the plausible prevalence of autism in each health service.

Third, a conjugate beta prior specific to each health service with the health service specific projections of adjusted updated prevalence from data linkage as the prior means, and prior standard deviations the same as the second prior. This prior was suitable as captures the extra information provided by the linkage and thus represents all students with autism diagnosis, not only those that receive SEED for autism. Additionally, this prior has narrow standard deviations which will model a theoretical upper bound on the prevalence of autism in each health service.

Fourth, a uniform prior specific to each health service with the adjusted autism prevalences from the school data for each health service as its lower bounds and the adjusted projected prevalences from data linkage for each health service as its upper bounds. This prior is suitable because it captures the information from both the school and clinical datasets, without specifying where within these bounds the true prevalences are likely to be.

4.7.5 Markov chain Monte Carlo sampling

Bayesian prevalence modelling used the JAGS (Just Another Gibbs Sampler) R package which uses Markov chain Monte Carlo (MCMC) sampling to produce posterior density distributions when given the above priors and adjusted prevalence observations. A burn-in period of 2000 samples was used to ensure models converge, then 2000 iterations were used to model the posterior densities. Visual inspection of trace plots showed no evidence of a lack of convergence and $\hat{r}$ values were less than 1.1.

5 Results

5.1 School data

The school dataset contained records for 3,056,306 Chilean students aged 6-18 in 2021. Of these, 1,487,224 (48.66%) were female and the rest were male. A special needs code was recorded for 339,968 (11.12%) students, indicating they received SEED during that school year. Of these students, 14,549 (4.28%) received SEED for autism and 46,224 (13.6%) received SEED for ADHD. Thus the global crude prevalence of autism in the school data was 0.48% (0.47-0.48%) and the global crude prevalence of ADHD was 1.51% (1.50-1.53%).

The crude prevalence of autism and ADHD vary with age, as shown in Tables 2 and 3, Figures 5 and 6, and Supplementary Figures 49 and 50. Autism prevalence is highest in 6-8 year olds and decreases with age while ADHD prevalence peaks around age 11 then decreases. Both conditions show a small increase in prevalence for age 18.

Table 2: Count and prevalence of autism cases by age band in Chile school data with normal confidence intervals.

Age band	Autism cases	Prevalence % (95% CI)
6-8	5162	0.69 (0.67, 0.71)
9-11	4212	0.55 (0.53, 0.57)
12-14	3038	0.41 (0.39, 0.42)
15-18	2137	0.27 (0.26, 0.28)

Table 3: Count and prevalence of ADHD cases by age band in Chile school data with normal confidence intervals.

Age band	ADHD cases	Prevalence % (95% CI)
6-8	5936	0.79 (0.77, 0.81)
9-11	15549	2.03 (1.99, 2.06)
12-14	14099	1.88 (1.85, 1.91)
15-18	10640	1.35 (1.32, 1.37)

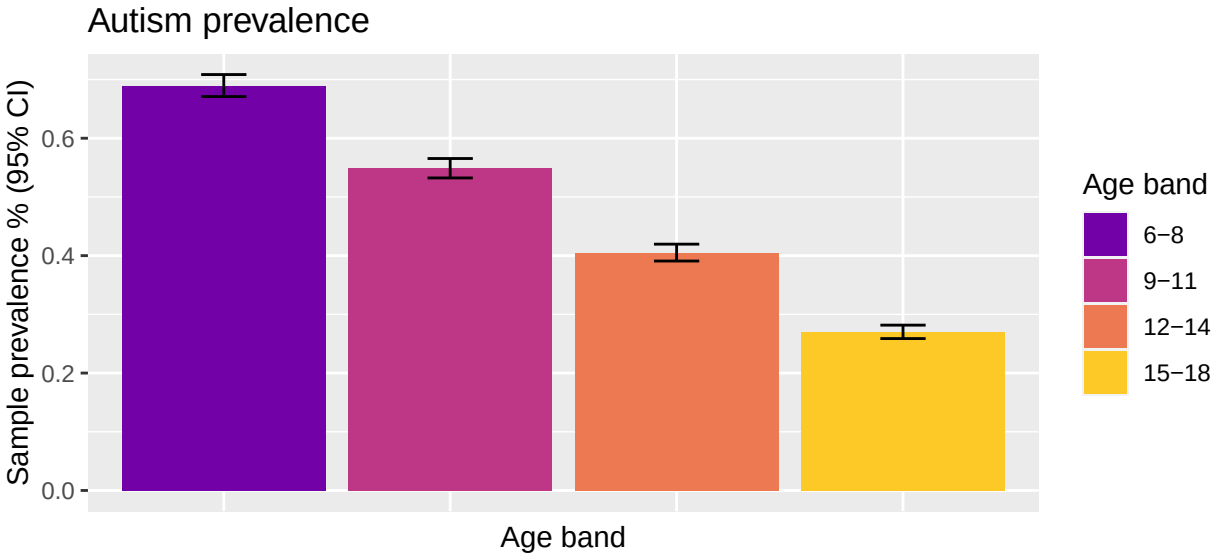

Figure 5: Sample prevalence of autism in school data by age band. Bars show 95% normal confidence intervals.

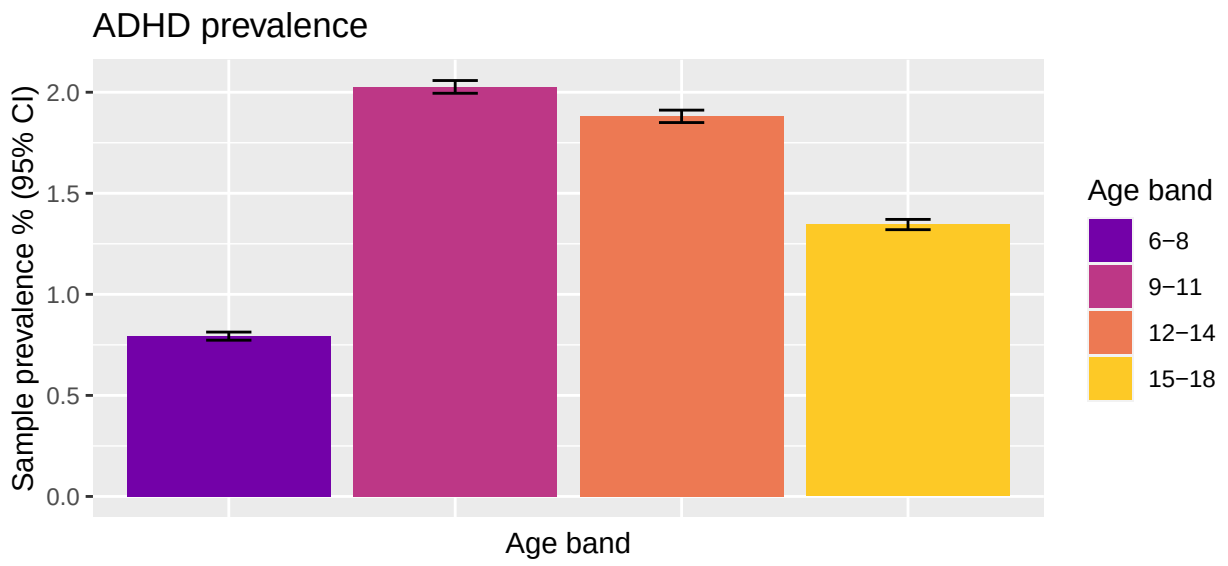

Figure 6: Sample prevalence of ADHD in school data by age band. Bars show 95% normal confidence intervals.

Table 4: Count and prevalence of autism cases by age band in Chile school data for females and males with normal confidence intervals.

	Female		Male	
Age band	Autism cases	Prevalence % (95% CI)	Autism cases	Prevalence % (95% CI)
6-8	774	0.21 (0.20, 0.23)	4388	1.15 (1.11, 1.18)
9-11	523	0.14 (0.13, 0.15)	3689	0.94 (0.91, 0.97)
12-14	391	0.11 (0.10, 0.12)	2647	0.69 (0.66, 0.72)
15-18	290	0.08 (0.07, 0.08)	1847	0.45 (0.43, 0.47)

In the school data, autism prevalence is 0.13% (0.13-0.14%) for females and 0.80% (0.79-0.82%) for males, with male to female ratio of 6.02, see Table 4 and Figure 7. ADHD prevalence is 1.01% (1.00-1.03%) for females and 1.98% (1.96-2.01%) for males, with male to female ratio of 1.96 see Table 5 and Figure 8. Autism and ADHD prevalences are higher in males than females for all ages, see also Supplementary Figures 51 and 52.

Autism varies by health service, as shown in Table 6, Figure 9 and Supplementary Figure 53. Autism prevalence is highest in Ñuble at 1.32% (1.24 - 1.40%) and Antofagasta at 0.84% (0.79- 0.89%), and is lowest in Metropolitano Norte at 0.3% (0.28 - 0.33%) and Araucanía Norte at 0.30% (0.24- 0.36%). Autism peaks in the 6-8 age band across all services except Chiloé and Magallanes where it peaks in the 9-11 band.

ADHD prevalence also varies across health services, as shown in Table 7, Figure 10 and Supplementary Figure 54. ADHD prevalence is highest in Magallanes at 1.42% (1.36 - 1.47%) and Talcahuano at 3.07% (2.93- 3.22%), and is lowest in Atacama at 0.49% (0.44 - 0.55%) and Antofagasta at 1.00% (0.94- 1.06%).

Table 5: Count and prevalence of ADHD cases by age band in Chile school data for females and males with normal confidence intervals.

	Female		Male	
Age band	ADHD cases	Prevalence % (95% CI)	ADHD cases	Prevalence % (95% CI)
6-8	774	0.21 (0.20, 0.23)	3944	1.03 (1.00, 1.06)
9-11	523	0.14 (0.13, 0.15)	10322	2.62 (2.57, 2.67)
12-14	391	0.11 (0.10, 0.12)	9714	2.53 (2.48, 2.58)
15-18	290	0.08 (0.07, 0.08)	7165	1.75 (1.71, 1.79)

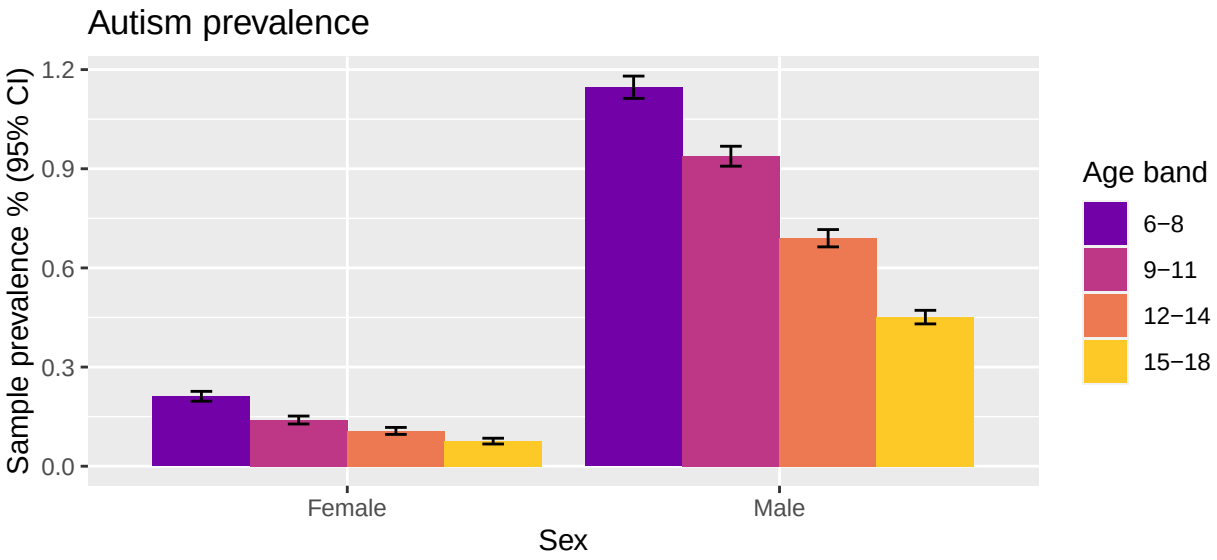

Figure 7: Sample prevalence of autism in school data by age band and sex. Bars show 95% normal confidence intervals.

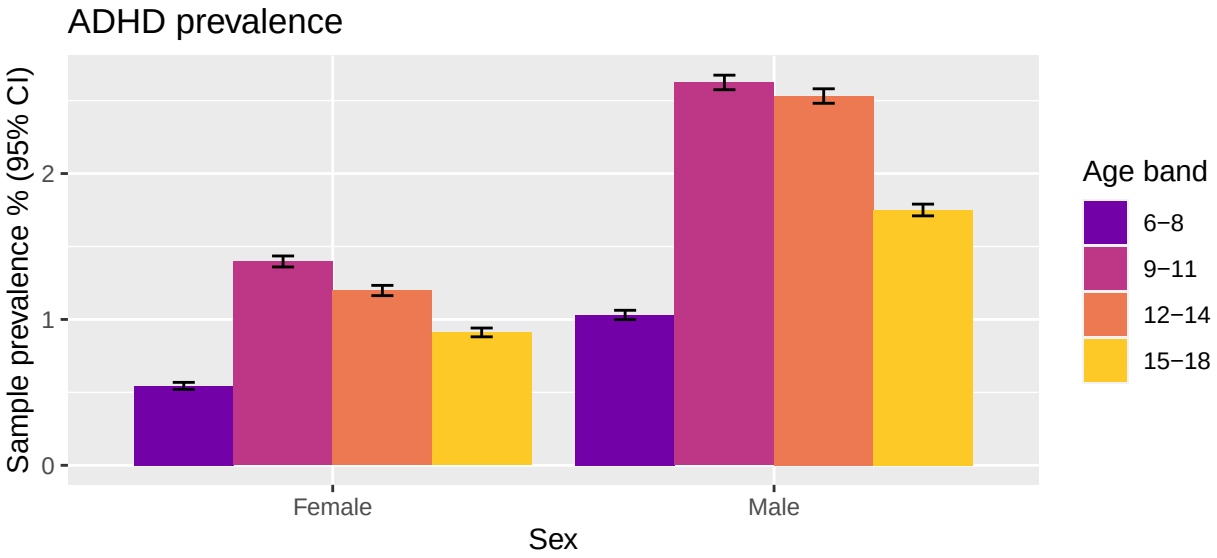

Figure 8: Sample prevalence of ADHD in school data by age band and sex. Bars show 95% normal confidence intervals.

Table 6: Count and prevalence of autism cases by health service and age band in Chile school data for females and males with normal confidence intervals.

		Female		Male	
Health service	Age band	Autism cases	Prevalence % (95% CI)	Autism cases	Prevalence % (95% CI)
Aconcagua	6-8	14	0.25 (0.12, 0.38)	59	0.99 (0.74, 1.24)
Aisén	6-8	9	0.40 (0.14, 0.66)	52	2.29 (1.67, 2.90)
Antofagasta	6-8	54	0.39 (0.29, 0.50)	302	2.09 (1.85, 2.32)
Araucanía Norte	6-8	3	0.07 (0.00, 0.15)	32	0.69 (0.45, 0.93)
Araucanía Sur	6-8	26	0.17 (0.10, 0.23)	139	0.86 (0.72, 1.00)
Arauco	6-8	11	0.29 (0.12, 0.46)	67	1.79 (1.36, 2.21)
Arica	6-8	18	0.34 (0.18, 0.49)	94	1.63 (1.31, 1.96)
Atacama	6-8	18	0.26 (0.14, 0.39)	53	0.75 (0.55, 0.95)
Biobío	6-8	12	0.14 (0.06, 0.23)	92	1.06 (0.85, 1.28)
Chiloé	6-8	2	0.06 (0.00, 0.14)	38	1.07 (0.73, 1.41)
Concepción	6-8	44	0.34 (0.24, 0.43)	252	1.87 (1.64, 2.10)
Coquimbo	6-8	36	0.21 (0.14, 0.28)	207	1.14 (0.98, 1.29)
Iquique	6-8	15	0.18 (0.09, 0.27)	100	1.11 (0.89, 1.33)
Magallanes	6-8	6	0.19 (0.04, 0.35)	42	1.30 (0.91, 1.70)
Maule	6-8	41	0.19 (0.13, 0.24)	193	0.84 (0.72, 0.96)
Metro. Central	6-8	22	0.16 (0.09, 0.22)	162	1.07 (0.91, 1.23)
Metro. Norte	6-8	26	0.12 (0.07, 0.17)	174	0.77 (0.65, 0.88)
Metro. Occidente	6-8	62	0.16 (0.12, 0.20)	323	0.81 (0.72, 0.90)
Metro. Oriente	6-8	31	0.14 (0.09, 0.19)	154	0.69 (0.58, 0.80)
Metro. Sur	6-8	44	0.18 (0.13, 0.24)	261	1.04 (0.91, 1.16)
Metro. Sur Oriente	6-8	41	0.15 (0.10, 0.20)	277	0.95 (0.84, 1.07)
O'Higgins	6-8	38	0.20 (0.14, 0.26)	215	1.08 (0.94, 1.22)
Osorno	6-8	9	0.20 (0.07, 0.33)	51	1.08 (0.78, 1.37)
Reloncaví	6-8	22	0.24 (0.14, 0.34)	96	1.02 (0.82, 1.23)
Talcahuano	6-8	25	0.40 (0.24, 0.55)	129	1.96 (1.63, 2.30)
Valdivia	6-8	11	0.15 (0.06, 0.23)	71	0.90 (0.69, 1.10)
Valparaíso	6-8	22	0.23 (0.13, 0.33)	156	1.57 (1.33, 1.82)
Viña del Mar	6-8	55	0.27 (0.20, 0.34)	292	1.37 (1.21, 1.52)
Ñuble	6-8	57	0.63 (0.46, 0.79)	305	3.21 (2.85, 3.56)
Aconcagua	9-11	3	0.05 (0.00, 0.11)	59	0.95 (0.71, 1.20)
Aisén	9-11	2	0.08 (0.00, 0.20)	26	1.05 (0.65, 1.45)
Antofagasta	9-11	45	0.30 (0.22, 0.39)	260	1.67 (1.47, 1.87)
Araucanía Norte	9-11	4	0.09 (0.00, 0.18)	26	0.57 (0.35, 0.79)
Araucanía Sur	9-11	16	0.10 (0.05, 0.15)	120	0.71 (0.58, 0.83)
Arauco	9-11	9	0.23 (0.08, 0.38)	70	1.73 (1.33, 2.13)
Arica	9-11	11	0.20 (0.08, 0.31)	57	0.95 (0.71, 1.20)
Atacama	9-11	7	0.09 (0.02, 0.16)	45	0.59 (0.42, 0.76)
Biobío	9-11	16	0.19 (0.09, 0.28)	68	0.76 (0.58, 0.94)
Chiloé	9-11	2	0.06 (0.00, 0.13)	49	1.27 (0.92, 1.63)
Concepción	9-11	30	0.22 (0.14, 0.30)	216	1.49 (1.29, 1.69)
Coquimbo	9-11	18	0.10 (0.05, 0.14)	145	0.78 (0.65, 0.90)
Iquique	9-11	16	0.19 (0.09, 0.28)	80	0.86 (0.67, 1.04)
Magallanes	9-11	6	0.17 (0.03, 0.31)	60	1.73 (1.30, 2.17)
Maule	9-11	16	0.07 (0.04, 0.11)	140	0.60 (0.50, 0.70)
Metro. Central	9-11	17	0.11 (0.06, 0.17)	136	0.86 (0.72, 1.00)
Metro. Norte	9-11	16	0.07 (0.04, 0.11)	147	0.62 (0.52, 0.72)
Metro. Occidente	9-11	28	0.09 (0.05, 0.12)	234	0.67 (0.59, 0.76)
Metro. Oriente	9-11	24	0.11 (0.06, 0.15)	129	0.55 (0.46, 0.65)
Metro. Sur	9-11	38	0.15 (0.10, 0.20)	238	0.90 (0.79, 1.01)
Metro. Sur Oriente	9-11	24	0.08 (0.05, 0.12)	191	0.63 (0.54, 0.72)
O'Higgins	9-11	30	0.15 (0.10, 0.21)	193	0.92 (0.79, 1.05)
Osorno	9-11	6	0.12 (0.02, 0.22)	54	1.05 (0.77, 1.32)
Reloncaví	9-11	9	0.09 (0.03, 0.15)	81	0.80 (0.62, 0.97)
Talcahuano	9-11	25	0.38 (0.23, 0.52)	132	1.87 (1.55, 2.18)
Valdivia	9-11	4	0.05 (0.00, 0.10)	38	0.45 (0.31, 0.59)
Valparaíso	9-11	15	0.16 (0.08, 0.23)	122	1.23 (1.01, 1.44)
Viña del Mar	9-11	47	0.22 (0.16, 0.28)	294	1.32 (1.17, 1.48)
Ñuble	9-11	39	0.41 (0.28, 0.54)	279	2.82 (2.50, 3.15)

Table 6: Count and prevalence of autism cases by health service and age band in Chile school data for females and males with normal confidence intervals. *(continued)*

		Female		Male	
Health service	Age band	Autism cases	Prevalence % (95% CI)	Autism cases	Prevalence % (95% CI)
Aconcagua	12-14	6	0.11 (0.02, 0.20)	38	0.67 (0.45, 0.88)
Aisén	12-14	6	0.24 (0.05, 0.44)	25	0.94 (0.57, 1.31)
Antofagasta	12-14	28	0.19 (0.12, 0.27)	170	1.09 (0.93, 1.25)
Araucanía Norte	12-14	6	0.13 (0.03, 0.24)	21	0.44 (0.25, 0.63)
Araucanía Sur	12-14	12	0.07 (0.03, 0.12)	94	0.57 (0.45, 0.68)
Arauco	12-14	5	0.14 (0.02, 0.25)	46	1.17 (0.84, 1.51)
Arica	12-14	10	0.19 (0.07, 0.31)	44	0.80 (0.57, 1.04)
Atacama	12-14	3	0.04 (0.00, 0.09)	28	0.37 (0.23, 0.51)
Biobío	12-14	5	0.06 (0.01, 0.11)	58	0.63 (0.47, 0.79)
Chiloé	12-14	1	0.03 (0.00, 0.08)	32	0.79 (0.52, 1.06)
Concepción	12-14	28	0.22 (0.14, 0.30)	155	1.14 (0.96, 1.32)
Coquimbo	12-14	14	0.08 (0.04, 0.13)	88	0.51 (0.40, 0.62)
Iquique	12-14	8	0.10 (0.03, 0.17)	52	0.58 (0.42, 0.74)
Magallanes	12-14	7	0.20 (0.05, 0.35)	55	1.55 (1.15, 1.96)
Maule	12-14	16	0.07 (0.04, 0.11)	86	0.37 (0.29, 0.45)
Metro. Central	12-14	5	0.03 (0.00, 0.06)	96	0.62 (0.49, 0.74)
Metro. Norte	12-14	15	0.07 (0.03, 0.11)	78	0.35 (0.27, 0.42)
Metro. Occidente	12-14	25	0.08 (0.05, 0.11)	186	0.56 (0.48, 0.64)
Metro. Oriente	12-14	21	0.10 (0.05, 0.14)	122	0.54 (0.45, 0.64)
Metro. Sur	12-14	18	0.07 (0.04, 0.11)	137	0.54 (0.45, 0.63)
Metro. Sur Oriente	12-14	17	0.06 (0.03, 0.09)	172	0.57 (0.49, 0.66)
O'Higgins	12-14	21	0.11 (0.06, 0.15)	116	0.57 (0.47, 0.67)
Osorno	12-14	6	0.12 (0.02, 0.22)	27	0.53 (0.33, 0.72)
Reloncaví	12-14	13	0.13 (0.06, 0.21)	63	0.61 (0.46, 0.76)
Talcahuano	12-14	11	0.17 (0.07, 0.26)	87	1.23 (0.97, 1.49)
Valdivia	12-14	3	0.04 (0.00, 0.08)	43	0.51 (0.36, 0.66)
Valparaíso	12-14	16	0.17 (0.09, 0.26)	107	1.11 (0.90, 1.32)
Viña del Mar	12-14	31	0.15 (0.10, 0.20)	230	1.08 (0.94, 1.22)
Ñuble	12-14	34	0.35 (0.23, 0.46)	191	1.87 (1.61, 2.13)
Aconcagua	15-18	4	0.07 (0.00, 0.14)	22	0.36 (0.21, 0.51)
Aisén	15-18	5	0.19 (0.02, 0.36)	24	0.85 (0.51, 1.19)
Antofagasta	15-18	25	0.17 (0.10, 0.23)	120	0.76 (0.62, 0.89)
Araucanía Norte	15-18	1	0.02 (0.00, 0.06)	17	0.34 (0.18, 0.51)
Araucanía Sur	15-18	7	0.04 (0.01, 0.07)	74	0.42 (0.32, 0.51)
Arauco	15-18	0	0.00 (0.00, 0.00)	21	0.50 (0.29, 0.71)
Arica	15-18	4	0.07 (0.00, 0.14)	35	0.61 (0.41, 0.81)
Atacama	15-18	3	0.04 (0.00, 0.09)	23	0.30 (0.18, 0.42)
Biobío	15-18	6	0.07 (0.01, 0.12)	51	0.52 (0.38, 0.66)
Chiloé	15-18	1	0.02 (0.00, 0.07)	14	0.31 (0.15, 0.47)
Concepción	15-18	23	0.17 (0.10, 0.24)	111	0.76 (0.62, 0.91)
Coquimbo	15-18	10	0.06 (0.02, 0.10)	63	0.35 (0.26, 0.43)
Iquique	15-18	2	0.02 (0.00, 0.06)	44	0.48 (0.34, 0.62)
Magallanes	15-18	14	0.38 (0.18, 0.57)	43	1.06 (0.74, 1.37)
Maule	15-18	14	0.06 (0.03, 0.09)	51	0.21 (0.15, 0.27)
Metro. Central	15-18	11	0.07 (0.03, 0.11)	67	0.40 (0.30, 0.49)
Metro. Norte	15-18	11	0.05 (0.02, 0.08)	62	0.26 (0.19, 0.32)
Metro. Occidente	15-18	14	0.04 (0.02, 0.07)	122	0.35 (0.29, 0.42)
Metro. Oriente	15-18	9	0.04 (0.01, 0.06)	65	0.26 (0.20, 0.32)
Metro. Sur	15-18	9	0.04 (0.01, 0.06)	86	0.33 (0.26, 0.40)
Metro. Sur Oriente	15-18	15	0.05 (0.02, 0.07)	131	0.40 (0.33, 0.47)
O'Higgins	15-18	16	0.08 (0.04, 0.12)	72	0.33 (0.25, 0.40)
Osorno	15-18	2	0.04 (0.00, 0.09)	24	0.43 (0.26, 0.60)
Reloncaví	15-18	2	0.02 (0.00, 0.05)	52	0.47 (0.34, 0.59)
Talcahuano	15-18	8	0.12 (0.04, 0.19)	42	0.56 (0.39, 0.74)
Valdivia	15-18	1	0.01 (0.00, 0.04)	35	0.39 (0.26, 0.52)
Valparaíso	15-18	11	0.11 (0.05, 0.18)	91	0.85 (0.68, 1.02)
Viña del Mar	15-18	35	0.16 (0.11, 0.22)	172	0.73 (0.62, 0.84)

Table 6: Count and prevalence of autism cases by health service and age band in Chile school data for females and males with normal confidence intervals. *(continued)*

		Female		Male	
Health service	Age band	Autism cases	Prevalence % (95% CI)	Autism cases	Prevalence % (95% CI)
Ñuble	15-18	27	0.26 (0.16, 0.36)	113	1.02 (0.83, 1.21)

Table 7: Count and prevalence of ADHD cases by health service and age band in Chile school data for females and males with normal confidence intervals.

		Female		Male	
Health service	Age band	ADHD cases	Prevalence % (95% CI)	ADHD cases	Prevalence % (95% CI)
Aconcagua	6-8	56	0.99 (0.74, 1.25)	71	1.19 (0.91, 1.46)
Aisén	6-8	12	0.54 (0.23, 0.84)	38	1.67 (1.14, 2.20)
Antofagasta	6-8	66	0.48 (0.36, 0.59)	106	0.73 (0.59, 0.87)
Araucanía Norte	6-8	24	0.57 (0.34, 0.79)	52	1.13 (0.82, 1.43)
Araucanía Sur	6-8	99	0.63 (0.51, 0.76)	206	1.27 (1.10, 1.45)
Arauco	6-8	20	0.53 (0.30, 0.76)	49	1.31 (0.94, 1.67)
Arica	6-8	17	0.32 (0.17, 0.47)	46	0.80 (0.57, 1.03)
Atacama	6-8	7	0.10 (0.03, 0.18)	20	0.28 (0.16, 0.41)
Biobío	6-8	79	0.95 (0.74, 1.16)	133	1.54 (1.28, 1.79)
Chiloé	6-8	28	0.85 (0.54, 1.16)	66	1.86 (1.42, 2.31)
Concepción	6-8	101	0.77 (0.62, 0.92)	162	1.20 (1.02, 1.38)
Coquimbo	6-8	105	0.61 (0.50, 0.73)	234	1.29 (1.12, 1.45)
Iquique	6-8	35	0.42 (0.28, 0.56)	96	1.07 (0.85, 1.28)
Magallanes	6-8	17	0.55 (0.29, 0.81)	57	1.77 (1.31, 2.22)
Maule	6-8	69	0.31 (0.24, 0.39)	182	0.79 (0.68, 0.91)
Metro. Central	6-8	87	0.62 (0.49, 0.75)	173	1.14 (0.97, 1.31)
Metro. Norte	6-8	114	0.53 (0.43, 0.62)	218	0.96 (0.84, 1.09)
Metro. Occidente	6-8	167	0.43 (0.37, 0.50)	296	0.74 (0.66, 0.82)
Metro. Oriente	6-8	115	0.53 (0.43, 0.62)	202	0.90 (0.78, 1.03)
Metro. Sur	6-8	131	0.55 (0.45, 0.64)	240	0.95 (0.83, 1.07)
Metro. Sur Oriente	6-8	189	0.69 (0.59, 0.79)	355	1.22 (1.10, 1.35)
O'Higgins	6-8	124	0.65 (0.54, 0.77)	252	1.27 (1.11, 1.42)
Osorno	6-8	16	0.36 (0.18, 0.53)	32	0.68 (0.44, 0.91)
Reloncaví	6-8	41	0.45 (0.31, 0.59)	76	0.81 (0.63, 0.99)
Talcahuano	6-8	62	0.99 (0.74, 1.23)	110	1.67 (1.36, 1.98)
Valdivia	6-8	41	0.54 (0.38, 0.71)	102	1.29 (1.04, 1.54)
Valparaíso	6-8	34	0.36 (0.24, 0.47)	89	0.90 (0.71, 1.08)
Viña del Mar	6-8	79	0.38 (0.30, 0.47)	171	0.80 (0.68, 0.92)
Ñuble	6-8	57	0.63 (0.46, 0.79)	110	1.16 (0.94, 1.37)
Aconcagua	9-11	144	2.45 (2.06, 2.85)	246	3.98 (3.49, 4.47)
Aisén	9-11	38	1.60 (1.10, 2.11)	94	3.79 (3.04, 4.55)
Antofagasta	9-11	143	0.97 (0.81, 1.12)	275	1.76 (1.56, 1.97)
Araucanía Norte	9-11	34	0.78 (0.52, 1.04)	130	2.85 (2.37, 3.33)
Araucanía Sur	9-11	240	1.48 (1.29, 1.66)	487	2.87 (2.62, 3.12)
Arauco	9-11	47	1.21 (0.86, 1.55)	107	2.64 (2.15, 3.13)
Arica	9-11	60	1.07 (0.80, 1.34)	118	1.97 (1.62, 2.32)
Atacama	9-11	28	0.37 (0.24, 0.51)	62	0.81 (0.61, 1.01)
Biobío	9-11	175	2.03 (1.73, 2.32)	329	3.67 (3.28, 4.06)
Chiloé	9-11	83	2.31 (1.82, 2.80)	183	4.76 (4.09, 5.43)
Concepción	9-11	333	2.41 (2.16, 2.67)	633	4.37 (4.03, 4.70)
Coquimbo	9-11	321	1.76 (1.57, 1.96)	545	2.93 (2.68, 3.17)
Iquique	9-11	108	1.25 (1.02, 1.48)	245	2.62 (2.30, 2.95)
Magallanes	9-11	83	2.39 (1.89, 2.90)	167	4.82 (4.11, 5.54)
Maule	9-11	227	1.03 (0.90, 1.17)	567	2.43 (2.24, 2.63)
Metro. Central	9-11	198	1.34 (1.15, 1.52)	430	2.72 (2.47, 2.97)
Metro. Norte	9-11	299	1.36 (1.20, 1.51)	507	2.15 (1.96, 2.33)
Metro. Occidente	9-11	372	1.14 (1.02, 1.25)	713	2.05 (1.90, 2.20)

Table 7: Count and prevalence of ADHD cases by health service and age band in Chile school data for females and males with normal confidence intervals. *(continued)*

		Female		Male	
Health service	Age band	ADHD cases	Prevalence % (95% CI)	ADHD cases	Prevalence % (95% CI)
Metro. Oriente	9-11	238	1.07 (0.94, 1.21)	478	2.05 (1.87, 2.23)
Metro. Sur	9-11	386	1.54 (1.39, 1.70)	656	2.48 (2.29, 2.67)
Metro. Sur Oriente	9-11	446	1.56 (1.41, 1.70)	813	2.70 (2.52, 2.88)
O'Higgins	9-11	318	1.60 (1.43, 1.78)	650	3.11 (2.88, 3.35)
Osorno	9-11	40	0.81 (0.56, 1.06)	95	1.84 (1.48, 2.21)
Reloncaví	9-11	80	0.83 (0.65, 1.01)	201	1.97 (1.70, 2.24)
Talcahuano	9-11	200	3.02 (2.60, 3.43)	316	4.47 (3.99, 4.95)
Valdivia	9-11	81	0.99 (0.77, 1.20)	179	2.12 (1.81, 2.42)
Valparaíso	9-11	99	1.03 (0.83, 1.23)	200	2.01 (1.74, 2.29)
Viña del Mar	9-11	231	1.09 (0.95, 1.23)	531	2.39 (2.19, 2.59)
Ñuble	9-11	175	1.83 (1.57, 2.10)	365	3.70 (3.32, 4.07)
Aconcagua	12-14	84	1.52 (1.20, 1.84)	187	3.27 (2.81, 3.74)
Aisén	12-14	39	1.58 (1.09, 2.07)	112	4.22 (3.45, 4.98)
Antofagasta	12-14	94	0.65 (0.52, 0.78)	248	1.59 (1.40, 1.79)
Araucanía Norte	12-14	44	0.97 (0.68, 1.25)	118	2.48 (2.04, 2.92)
Araucanía Sur	12-14	153	0.95 (0.80, 1.10)	389	2.34 (2.11, 2.57)
Arauco	12-14	33	0.89 (0.59, 1.19)	118	3.01 (2.48, 3.55)
Arica	12-14	35	0.67 (0.45, 0.89)	136	2.48 (2.07, 2.89)
Atacama	12-14	31	0.43 (0.28, 0.58)	59	0.78 (0.58, 0.98)
Biobío	12-14	140	1.61 (1.35, 1.87)	358	3.90 (3.50, 4.30)
Chiloé	12-14	95	2.49 (1.99, 2.98)	209	5.15 (4.47, 5.83)
Concepción	12-14	374	2.91 (2.62, 3.20)	641	4.72 (4.36, 5.08)
Coquimbo	12-14	286	1.71 (1.52, 1.91)	565	3.27 (3.01, 3.54)
Iquique	12-14	100	1.22 (0.98, 1.46)	209	2.33 (2.02, 2.65)
Magallanes	12-14	81	2.35 (1.84, 2.86)	178	5.02 (4.30, 5.74)
Maule	12-14	197	0.89 (0.77, 1.02)	544	2.36 (2.16, 2.56)
Metro. Central	12-14	171	1.16 (0.99, 1.33)	400	2.57 (2.32, 2.82)
Metro. Norte	12-14	270	1.26 (1.11, 1.41)	518	2.29 (2.10, 2.49)
Metro. Occidente	12-14	278	0.89 (0.78, 0.99)	597	1.80 (1.65, 1.94)
Metro. Oriente	12-14	212	0.97 (0.84, 1.10)	458	2.04 (1.85, 2.22)
Metro. Sur	12-14	283	1.17 (1.04, 1.31)	557	2.18 (2.00, 2.36)
Metro. Sur Oriente	12-14	378	1.33 (1.19, 1.46)	687	2.29 (2.13, 2.46)
O'Higgins	12-14	263	1.35 (1.19, 1.51)	633	3.11 (2.87, 3.35)
Osorno	12-14	30	0.60 (0.39, 0.82)	106	2.07 (1.68, 2.46)
Reloncaví	12-14	64	0.66 (0.50, 0.82)	156	1.52 (1.28, 1.76)
Talcahuano	12-14	173	2.60 (2.22, 2.98)	367	5.19 (4.67, 5.70)
Valdivia	12-14	52	0.64 (0.47, 0.81)	139	1.65 (1.38, 1.92)
Valparaíso	12-14	76	0.82 (0.64, 1.00)	218	2.25 (1.96, 2.55)
Viña del Mar	12-14	183	0.88 (0.75, 1.01)	432	2.02 (1.83, 2.21)
Ñuble	12-14	166	1.70 (1.44, 1.95)	375	3.67 (3.31, 4.04)
Aconcagua	15-18	58	0.99 (0.74, 1.24)	128	2.11 (1.75, 2.47)
Aisén	15-18	35	1.35 (0.91, 1.80)	76	2.70 (2.10, 3.30)
Antofagasta	15-18	72	0.48 (0.37, 0.60)	191	1.21 (1.04, 1.38)
Araucanía Norte	15-18	34	0.74 (0.49, 0.98)	53	1.07 (0.79, 1.36)
Araucanía Sur	15-18	98	0.58 (0.47, 0.70)	208	1.17 (1.01, 1.33)
Arauco	15-18	37	0.92 (0.63, 1.22)	103	2.44 (1.98, 2.91)
Arica	15-18	30	0.55 (0.35, 0.74)	66	1.14 (0.87, 1.42)
Atacama	15-18	22	0.31 (0.18, 0.43)	60	0.78 (0.58, 0.97)
Biobío	15-18	128	1.40 (1.16, 1.64)	278	2.83 (2.50, 3.16)
Chiloé	15-18	84	2.00 (1.58, 2.42)	167	3.67 (3.13, 4.22)
Concepción	15-18	373	2.74 (2.46, 3.01)	605	4.16 (3.84, 4.49)
Coquimbo	15-18	231	1.36 (1.19, 1.54)	508	2.79 (2.55, 3.03)
Iquique	15-18	72	0.85 (0.66, 1.05)	179	1.96 (1.68, 2.25)
Magallanes	15-18	80	2.15 (1.69, 2.62)	198	4.87 (4.21, 5.53)
Maule	15-18	99	0.44 (0.35, 0.52)	291	1.19 (1.05, 1.32)
Metro. Central	15-18	139	0.89 (0.74, 1.03)	272	1.61 (1.42, 1.80)
Metro. Norte	15-18	256	1.15 (1.01, 1.29)	370	1.55 (1.39, 1.70)

Table 7: Count and prevalence of ADHD cases by health service and age band in Chile school data for females and males with normal confidence intervals. *(continued)*

		Female		Male	
Health service	Age band	ADHD cases	Prevalence % (95% CI)	ADHD cases	Prevalence % (95% CI)
Metro. Occidente	15-18	189	0.59 (0.51, 0.67)	418	1.21 (1.09, 1.32)
Metro. Oriente	15-18	187	0.79 (0.68, 0.90)	342	1.37 (1.22, 1.51)
Metro. Sur	15-18	203	0.83 (0.71, 0.94)	406	1.55 (1.40, 1.70)
Metro. Sur Oriente	15-18	292	0.95 (0.84, 1.06)	538	1.65 (1.51, 1.79)
O'Higgins	15-18	178	0.89 (0.76, 1.02)	366	1.67 (1.50, 1.84)
Osorno	15-18	31	0.59 (0.38, 0.80)	71	1.26 (0.97, 1.55)
Reloncaví	15-18	45	0.44 (0.31, 0.56)	152	1.36 (1.15, 1.58)
Talcahuano	15-18	162	2.33 (1.98, 2.69)	289	3.89 (3.45, 4.33)
Valdivia	15-18	32	0.38 (0.25, 0.51)	87	0.96 (0.76, 1.16)
Valparaíso	15-18	59	0.60 (0.44, 0.75)	161	1.50 (1.27, 1.73)
Viña del Mar	15-18	105	0.49 (0.40, 0.58)	292	1.24 (1.10, 1.38)
Ñuble	15-18	144	1.40 (1.17, 1.63)	290	2.62 (2.33, 2.92)

For school fees, which are used here as a proxy for SES, autism prevalence is highest among students that receive free or low fee education, though the sample size for students that pay $1,000-$10,000 monthly is very small, see Table 8, Figure 11 and Supplementary Figure 55. ADHD prevalence is more consistent across school fee levels, except for the $1,000-$10,000 band which has low prevalence and very few cases, see Table 9, Figure 12 and Supplementary Figure 56. Prevalence is higher among older students for higher fee bands. For both autism and ADHD, prevalence is very low among students paying more than $100,000 monthly, suggesting students from wealthier families may not be accessing SEED or may not be eligible for it due to attending private schools.

Autism prevalence and distribution by age is very similar between Mapuche and non-Indigenous students in the Araucanía, Aysén, Biobío, Los Lagos, Los Rios, Magallanes and Región Metropolitana de Santiago regions, as shown in Table 10, Figure 13 and Supplementary Figure 57. ADHD prevalence is also consistent across Mapuche and non-Indigenous students, see Table 11, Figure 14 and Supplementary Figure 58. Among students of other Indigenous groups, autism and ADHD prevalence appear to peak in older age groups, however this result is based on very few cases.

Autism and ADHD are slightly more prevalent for students at rural schools and both follow the same age patterns observed earlier. See Tables 12 and 13, Figures 15 and 16 and Supplementary Figures 59 and 60.

5.2 Frequentist prevalence estimation

After using frequentist methods to adjust for the age and sex distribution of the child population of Chile in 2021, the adjusted prevalence of autism was 0.46% (0.46-0.47%) and the adjusted prevalence of ADHD was 1.50% (1.48-1.51%). Among females, the adjusted prevalence of autism was 0.13% (0.13-0.14%) and among males it was 0.79% (0.77-0.80%). Among females, the adjusted prevalence of ADHD was 1.01% (1.00-1.03%) and among males it was 1.97% (1.94-1.99%). This gives male to female ratio of 6.00 As we know the prevalences of autism and ADHD in this data are likely to be underestimates, the age- and sex-adjusted prevalences can be considered a lower bound on the true prevalence of autism and of ADHD in Chile.

Considering school data by health service, shown in Table 14, the adjusted prevalence of autism is much higher in Ñuble at 1.29% (1.21 - 1.37%) than other health services and is low in rural areas such as Araucanía. It is also low in the Metropolitano health services which serve Santiago, Chile's largest city. Adjusted ADHD prevalence, shown in Table 15, is medial for Metropolitano health services and very low for Atacama, a fairly urbanised region, at 0.49% (0.43 - 0.56%).

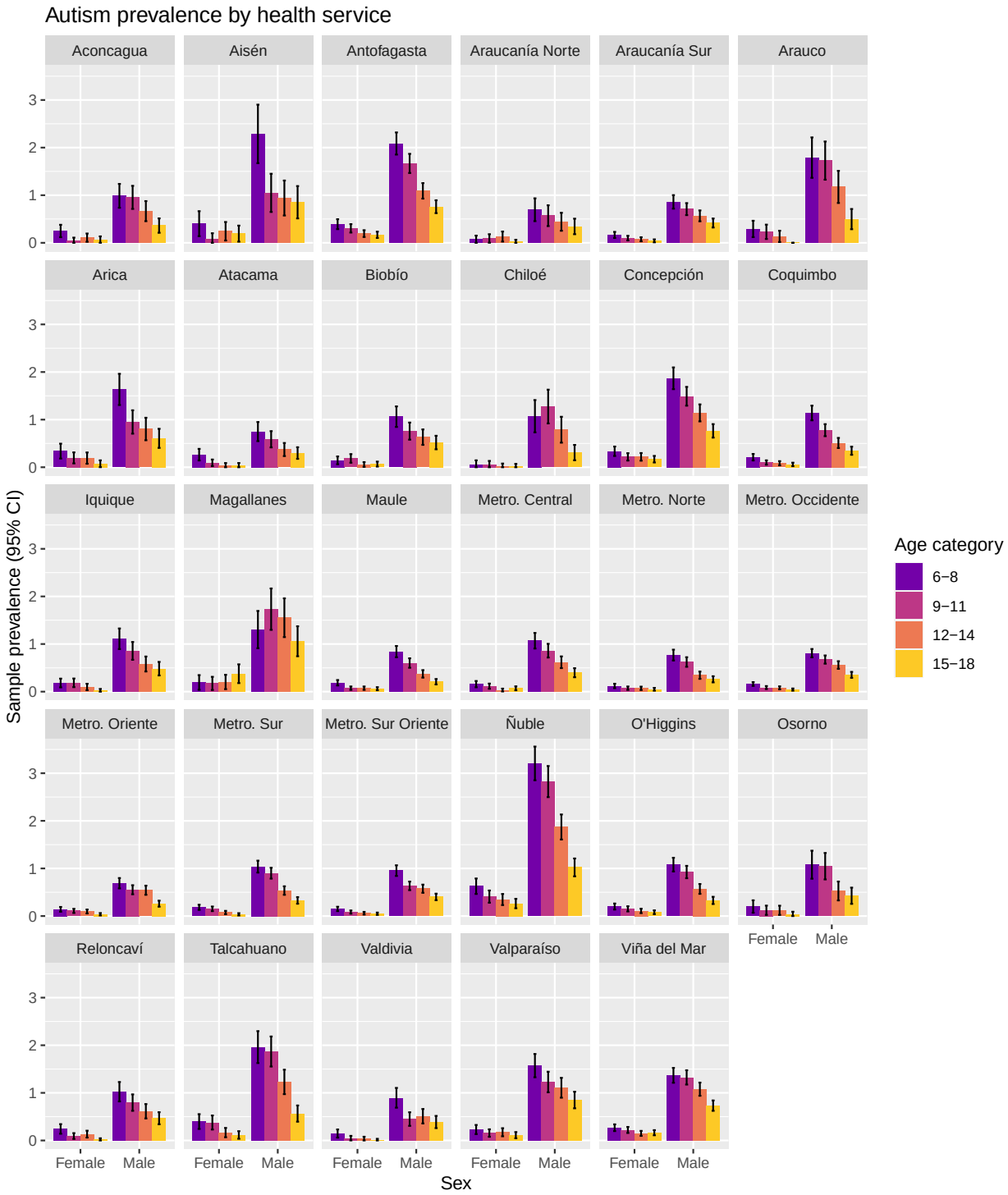

Figure 9: Sample prevalence of autism in school data by health service, age band and sex. Bars show 95% normal confidence intervals.

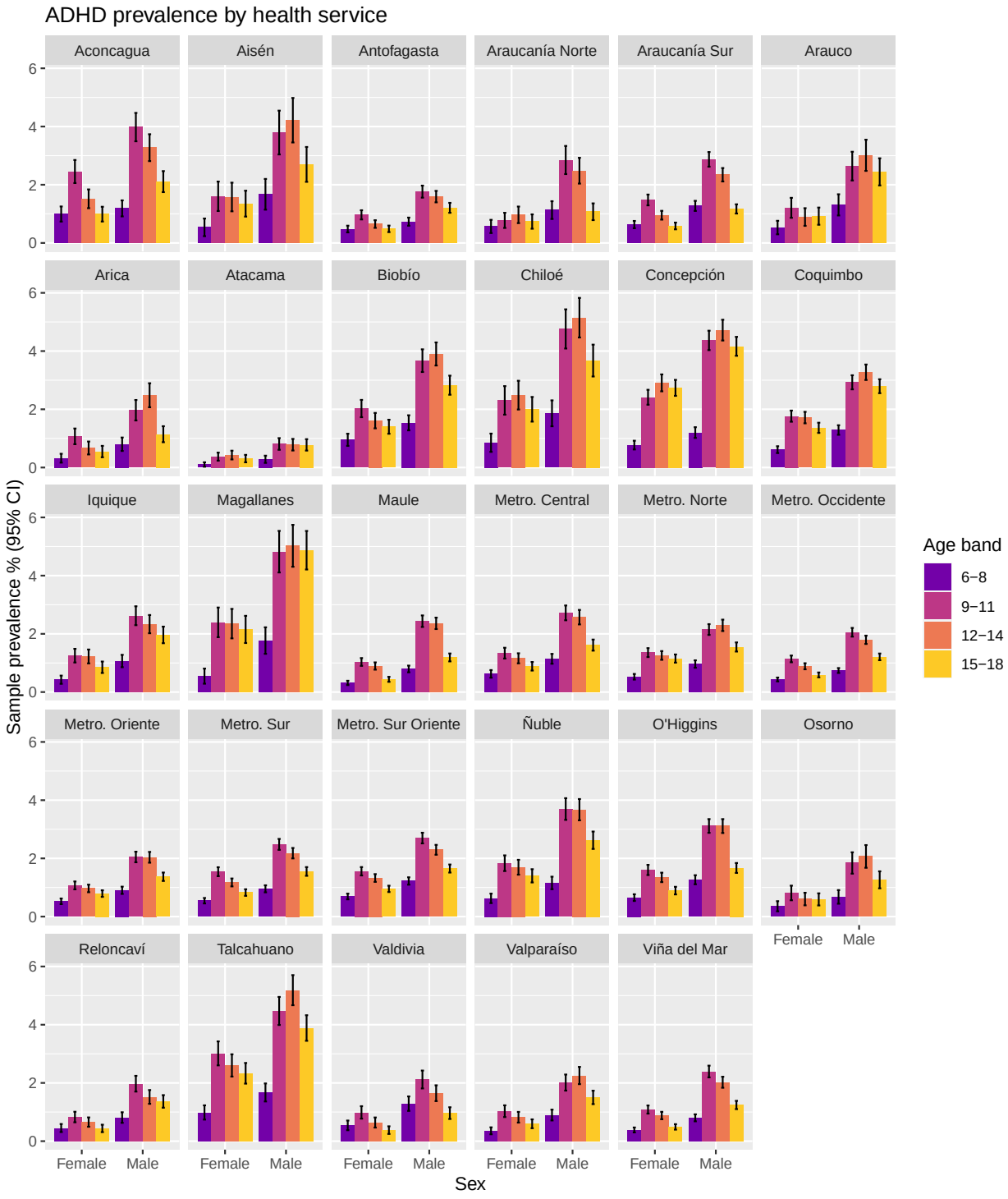

Figure 10: Sample prevalence of ADHD in school data by health service, age band and sex. Bars show 95% normal confidence intervals.

Table 8: Count and prevalence of Autism cases by school fee and age band in Chile school data for females and males with normal confidence intervals.

		Female		Male	
School fee	Age band	Autism cases	Prevalence % (95% CI)	Autism cases	Prevalence % (95% CI)
Free	6-8	655	0.25 (0.23, 0.27)	3799	1.37 (1.33, 1.42)
$1,000-$10,000	6-8	0	0.00 (0.00, 0.00)	2	1.17 (0.00, 2.78)
$10,001-$25,000	6-8	5	0.11 (0.01, 0.20)	20	0.44 (0.25, 0.64)
$25,001-$50,000	6-8	37	0.14 (0.10, 0.19)	168	0.70 (0.59, 0.80)
$50,001-$100,000	6-8	53	0.17 (0.13, 0.22)	296	0.94 (0.83, 1.04)
$100,001+	6-8	8	0.02 (0.01, 0.04)	30	0.08 (0.05, 0.11)
No information	6-8	16	0.23 (0.12, 0.35)	73	0.92 (0.71, 1.13)
Free	9-11	434	0.16 (0.15, 0.18)	3147	1.10 (1.06, 1.14)
$1,000-$10,000	9-11	0	0.00 (0.00, 0.00)	2	1.59 (0.00, 3.77)
$10,001-$25,000	9-11	1	0.02 (0.00, 0.06)	23	0.52 (0.31, 0.73)
$25,001-$50,000	9-11	26	0.10 (0.06, 0.14)	152	0.61 (0.52, 0.71)
$50,001-$100,000	9-11	45	0.14 (0.10, 0.18)	255	0.77 (0.67, 0.86)
$100,001+	9-11	6	0.02 (0.00, 0.03)	35	0.09 (0.06, 0.12)
No information	9-11	11	0.19 (0.08, 0.30)	75	1.11 (0.86, 1.36)
Free	12-14	332	0.13 (0.11, 0.14)	2246	0.81 (0.77, 0.84)
$1,000-$10,000	12-14	0	0.00 (0.00, 0.00)	4	3.67 (0.14, 7.20)
$10,001-$25,000	12-14	2	0.04 (0.00, 0.10)	9	0.21 (0.07, 0.34)
$25,001-$50,000	12-14	21	0.08 (0.04, 0.11)	126	0.53 (0.44, 0.63)
$50,001-$100,000	12-14	25	0.08 (0.05, 0.11)	195	0.56 (0.48, 0.64)
$100,001+	12-14	4	0.01 (0.00, 0.02)	31	0.09 (0.06, 0.12)
No information	12-14	7	0.14 (0.04, 0.25)	36	0.60 (0.41, 0.80)
Free	15-18	236	0.09 (0.08, 0.10)	1547	0.53 (0.50, 0.55)
$1,000-$10,000	15-18	0	0.00 (0.00, 0.00)	0	0.00 (0.00, 0.00)
$10,001-$25,000	15-18	3	0.06 (0.00, 0.13)	10	0.23 (0.09, 0.37)
$25,001-$50,000	15-18	16	0.05 (0.03, 0.08)	82	0.31 (0.24, 0.38)
$50,001-$100,000	15-18	29	0.08 (0.05, 0.11)	153	0.40 (0.33, 0.46)
$100,001+	15-18	2	0.01 (0.00, 0.01)	25	0.06 (0.04, 0.09)
No information	15-18	4	0.08 (0.00, 0.16)	30	0.44 (0.29, 0.60)

Table 9: Count and prevalence of ADHD cases by school fee and age band in Chile school data for females and males with normal confidence intervals.

		Female		Male	
School fee	Age band	ADHD cases	Prevalence % (95% CI)	ADHD cases	Prevalence % (95% CI)
Free	6-8	1543	0.59 (0.56, 0.62)	3117	1.13 (1.09, 1.17)
$1,000-$10,000	6-8	0	0.00 (0.00, 0.00)	1	0.58 (0.00, 1.73)
$10,001-$25,000	6-8	32	0.70 (0.46, 0.94)	32	0.71 (0.46, 0.95)
$25,001-$50,000	6-8	167	0.65 (0.55, 0.75)	267	1.11 (0.97, 1.24)
$50,001-$100,000	6-8	200	0.65 (0.56, 0.74)	437	1.38 (1.26, 1.51)
$100,001+	6-8	20	0.05 (0.03, 0.07)	32	0.08 (0.06, 0.11)
No information	6-8	30	0.44 (0.28, 0.59)	58	0.73 (0.54, 0.92)
Free	9-11	4061	1.52 (1.47, 1.57)	8323	2.91 (2.85, 2.97)
$1,000-$10,000	9-11	0	0.00 (0.00, 0.00)	0	0.00 (0.00, 0.00)
$10,001-$25,000	9-11	47	1.01 (0.72, 1.29)	102	2.29 (1.85, 2.73)
$25,001-$50,000	9-11	417	1.58 (1.43, 1.73)	616	2.49 (2.29, 2.68)
$50,001-$100,000	9-11	567	1.76 (1.62, 1.91)	1024	3.08 (2.89, 3.26)
$100,001+	9-11	62	0.17 (0.12, 0.21)	112	0.30 (0.24, 0.35)
No information	9-11	73	1.26 (0.97, 1.54)	145	2.15 (1.80, 2.50)
Free	12-14	3342	1.28 (1.24, 1.33)	7634	2.74 (2.68, 2.80)
$1,000-$10,000	12-14	0	0.00 (0.00, 0.00)	0	0.00 (0.00, 0.00)
$10,001-$25,000	12-14	48	1.02 (0.74, 1.31)	79	1.80 (1.41, 2.20)
$25,001-$50,000	12-14	364	1.36 (1.22, 1.50)	638	2.70 (2.49, 2.90)
$50,001-$100,000	12-14	509	1.55 (1.42, 1.68)	1083	3.12 (2.94, 3.30)
$100,001+	12-14	72	0.20 (0.15, 0.24)	147	0.41 (0.34, 0.47)
No information	12-14	50	1.03 (0.74, 1.31)	133	2.23 (1.85, 2.60)
Free	15-18	2579	0.97 (0.93, 1.00)	5529	1.88 (1.83, 1.93)
$1,000-$10,000	15-18	0	0.00 (0.00, 0.00)	1	0.39 (0.00, 1.14)
$10,001-$25,000	15-18	21	0.43 (0.25, 0.62)	25	0.57 (0.35, 0.80)
$25,001-$50,000	15-18	343	1.17 (1.04, 1.29)	482	1.84 (1.67, 2.00)
$50,001-$100,000	15-18	416	1.12 (1.01, 1.22)	898	2.33 (2.18, 2.48)
$100,001+	15-18	79	0.21 (0.16, 0.26)	151	0.39 (0.33, 0.45)
No information	15-18	37	0.73 (0.50, 0.97)	79	1.17 (0.91, 1.43)

Table 10: Count and prevalence of Autism cases by ethnicity and age band in Chile school data for females and males with normal confidence intervals. Regions with high Mapuche populations only.

		Female		Male	
Ethnicity	Age band	Autism cases	Prevalence % (95% CI)	Autism cases	Prevalence % (95% CI)
Mapuche	6-8	29	0.15 (0.10, 0.20)	186	0.93 (0.80, 1.07)
Other Indigenous group	6-8	0	0.00 (0.00, 0.00)	6	0.77 (0.16, 1.39)
No Indigenous group	6-8	377	0.18 (0.16, 0.20)	2220	1.02 (0.98, 1.06)
Mapuche	9-11	11	0.05 (0.02, 0.09)	155	0.72 (0.61, 0.84)
Other Indigenous group	9-11	1	0.17 (0.00, 0.50)	8	1.36 (0.42, 2.30)
No Indigenous group	9-11	264	0.13 (0.11, 0.14)	1852	0.84 (0.80, 0.87)
Mapuche	12-14	12	0.06 (0.03, 0.09)	110	0.52 (0.42, 0.62)
Other Indigenous group	12-14	0	0.00 (0.00, 0.00)	2	0.34 (0.00, 0.81)
No Indigenous group	12-14	192	0.09 (0.08, 0.11)	1385	0.64 (0.61, 0.67)
Mapuche	15-18	3	0.02 (0.00, 0.03)	71	0.35 (0.27, 0.43)
Other Indigenous group	15-18	0	0.00 (0.00, 0.00)	1	0.19 (0.00, 0.56)
No Indigenous group	15-18	136	0.06 (0.05, 0.07)	969	0.41 (0.39, 0.44)

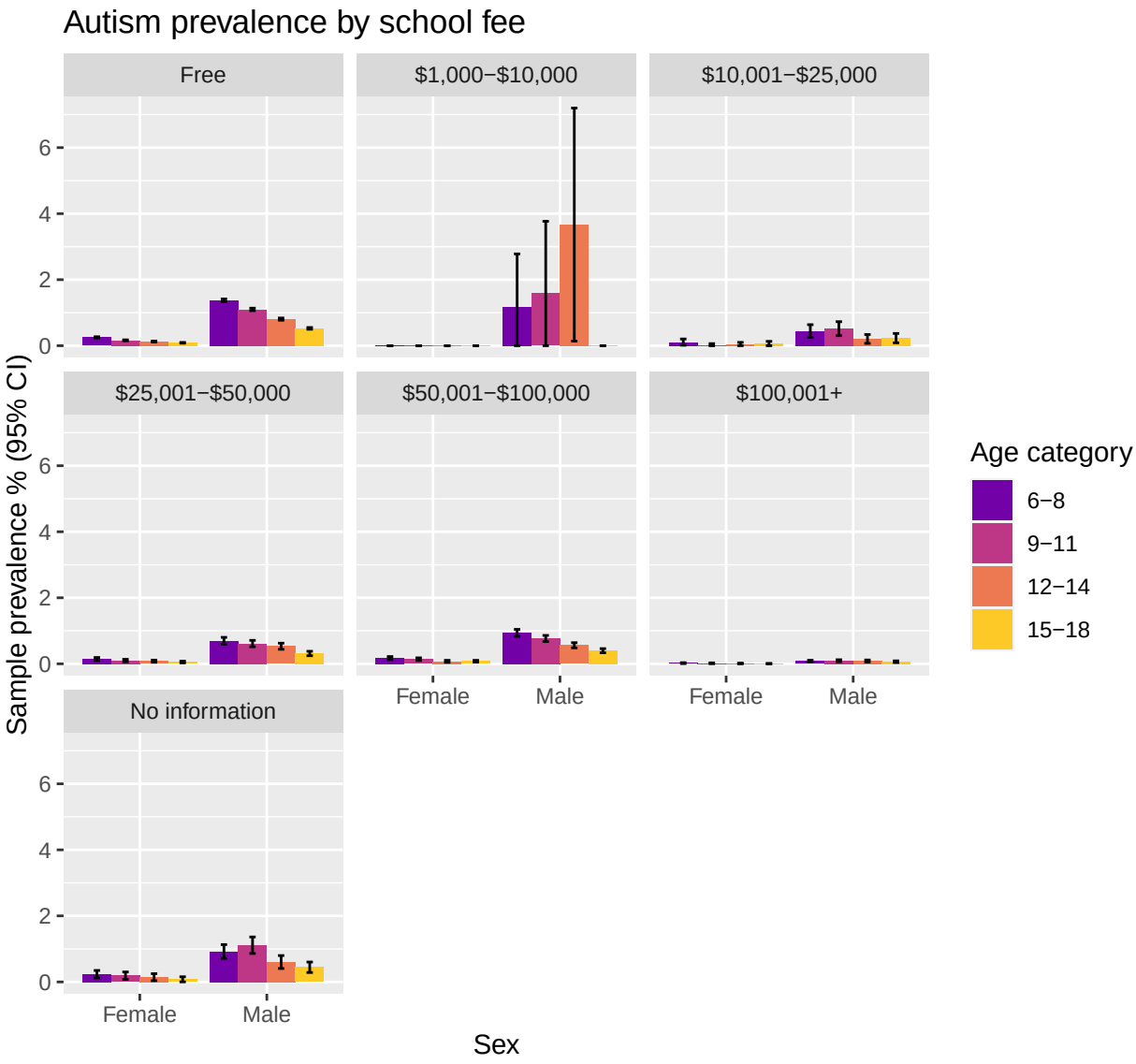

Figure 11: Sample prevalence of autism in school data by student's monthly school fee (Peso), age band and sex. Bars show 95% normal confidence intervals.

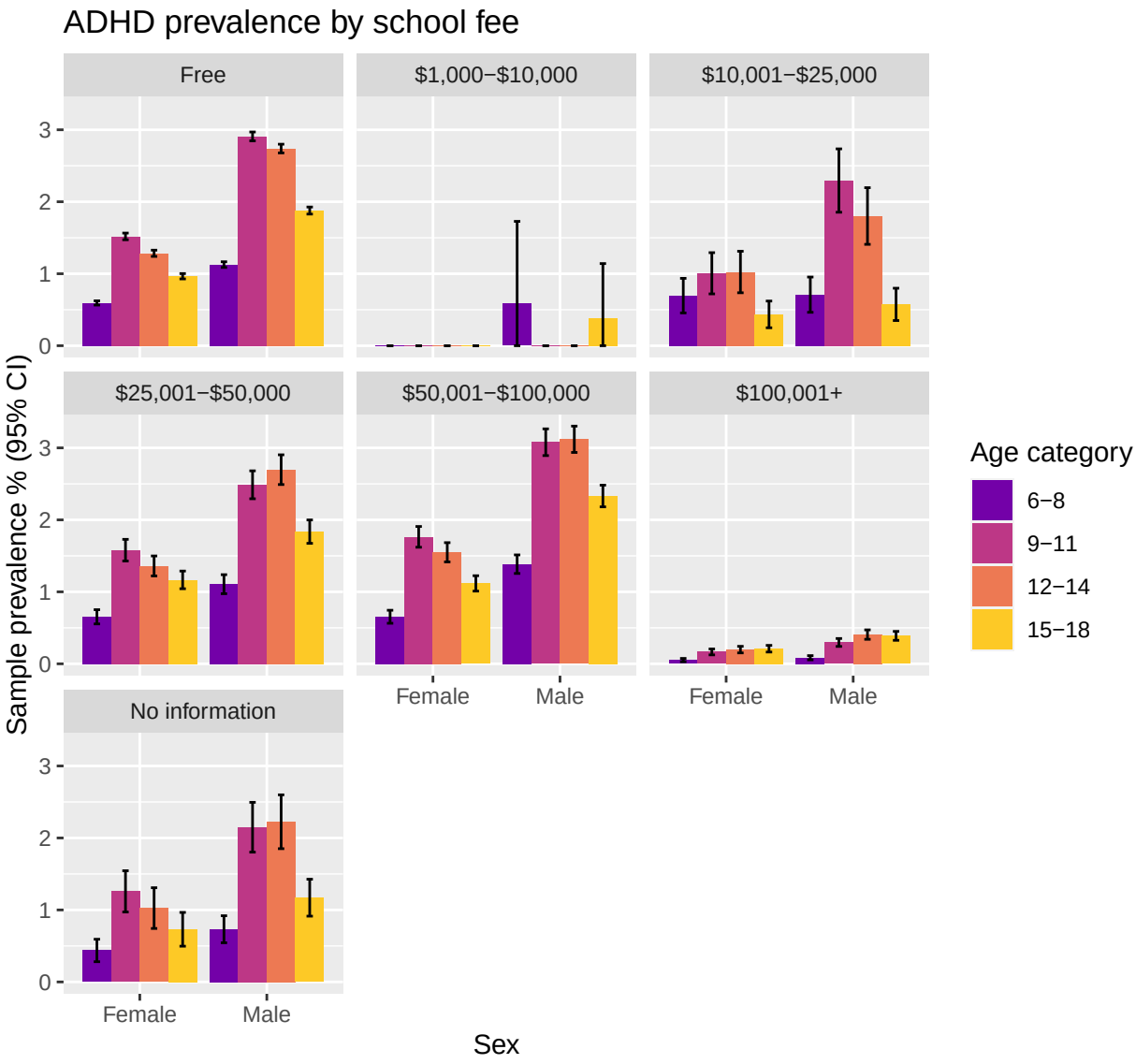

Figure 12: Sample prevalence of ADHD in school data by student's monthly school fee (Peso), age band and sex. Bars show 95% normal confidence intervals.

Table 11: Count and prevalence of ADHD cases by ethnicity and age band in Chile school data for females and males with normal confidence intervals. Regions with high Mapuche populations only.

		Female		Male	
Ethnicity	Age band	ADHD cases	Prevalence % (95% CI)	ADHD cases	Prevalence % (95% CI)
Mapuche	6-8	113	0.58 (0.48, 0.69)	218	1.09 (0.95, 1.24)
Other Indigenous group	6-8	3	0.39 (0.00, 0.84)	7	0.90 (0.24, 1.57)
No Indigenous group	6-8	1227	0.59 (0.56, 0.62)	2342	1.07 (1.03, 1.12)
Mapuche	9-11	253	1.24 (1.08, 1.39)	557	2.59 (2.38, 2.81)
Other Indigenous group	9-11	7	1.19 (0.31, 2.06)	11	1.87 (0.78, 2.97)
No Indigenous group	9-11	3113	1.48 (1.43, 1.53)	5950	2.68 (2.62, 2.75)
Mapuche	12-14	206	1.00 (0.87, 1.14)	479	2.26 (2.06, 2.46)
Other Indigenous group	12-14	4	0.72 (0.02, 1.42)	15	2.54 (1.27, 3.81)
No Indigenous group	12-14	2660	1.29 (1.24, 1.34)	5614	2.59 (2.52, 2.66)
Mapuche	15-18	131	0.70 (0.58, 0.81)	278	1.38 (1.22, 1.54)
Other Indigenous group	15-18	6	1.14 (0.23, 2.05)	2	0.38 (0.00, 0.90)
No Indigenous group	15-18	2268	1.03 (0.99, 1.08)	4353	1.86 (1.80, 1.91)

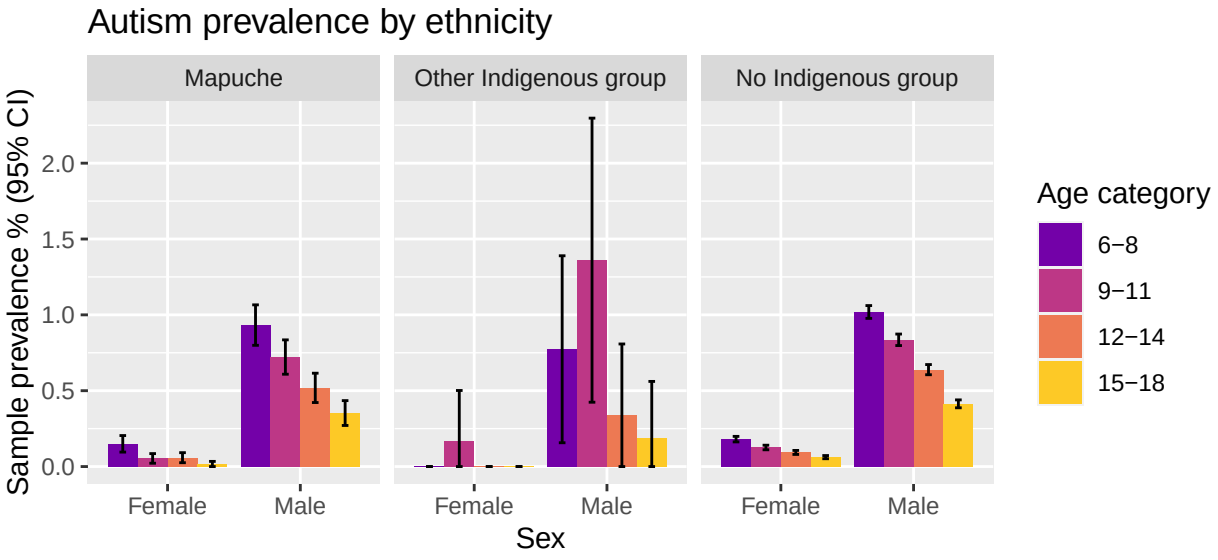

Figure 13: Sample prevalence of autism in school data by ethnicity, age band and sex. Bars show 95% normal confidence intervals. Regions with high Mapuche populations only.

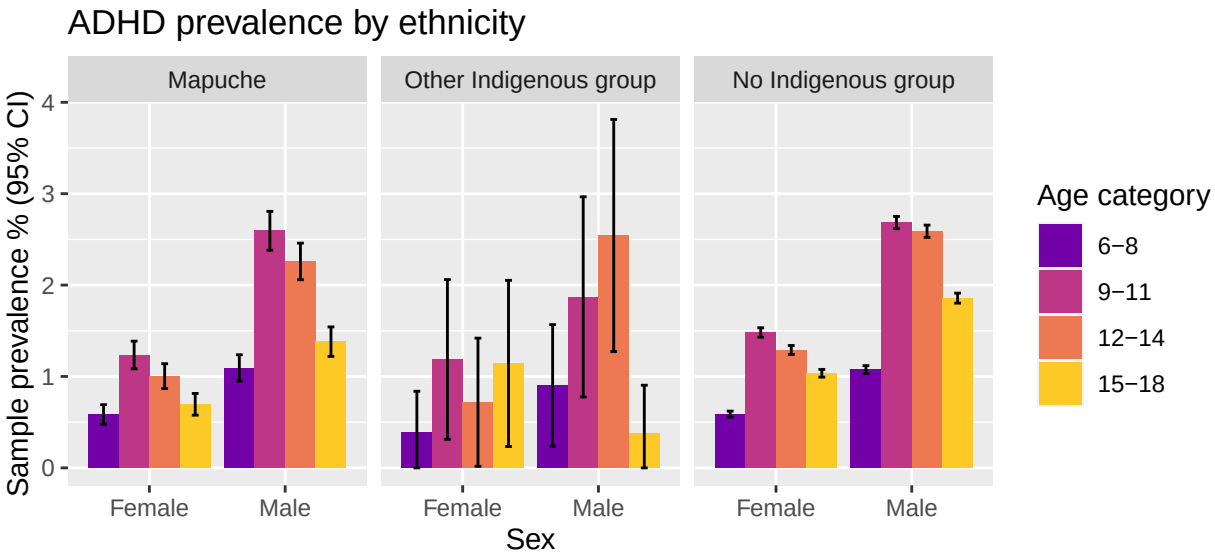

Figure 14: Sample prevalence of ADHD in school data by ethnicity, age band and sex. Bars show 95% normal confidence intervals. Regions with high Mapuche populations only.

Table 12: Count and prevalence of Autism cases by school's rurality and age band in Chile school data for females and males with normal confidence intervals.

		Female		Male	
School rurality	Age band	Autism cases	Prevalence % (95% CI)	Autism cases	Prevalence % (95% CI)
Rural	6-8	93	0.26 (0.21, 0.31)	519	1.33 (1.22, 1.45)
Urban	6-8	681	0.21 (0.19, 0.22)	3869	1.13 (1.09, 1.16)
Rural	9-11	61	0.16 (0.12, 0.20)	472	1.14 (1.03, 1.24)
Urban	9-11	462	0.14 (0.12, 0.15)	3217	0.91 (0.88, 0.95)
Rural	12-14	44	0.16 (0.12, 0.21)	291	0.93 (0.82, 1.03)
Urban	12-14	347	0.10 (0.09, 0.11)	2356	0.67 (0.64, 0.70)
Rural	15-18	9	0.08 (0.03, 0.13)	92	0.61 (0.49, 0.73)
Urban	15-18	281	0.08 (0.07, 0.08)	1755	0.45 (0.42, 0.47)

Table 13: Count and prevalence of ADHD cases by school's rurality and age band in Chile school data for females and males with normal confidence intervals.

		Female		Male	
School rurality	Age band	ADHD cases	Prevalence % (95% CI)	ADHD cases	Prevalence % (95% CI)
Rural	6-8	221	0.62 (0.53, 0.70)	503	1.29 (1.18, 1.40)
Urban	6-8	1771	0.54 (0.51, 0.56)	3441	1.00 (0.97, 1.03)
Rural	9-11	532	1.41 (1.29, 1.53)	1304	3.14 (2.97, 3.31)
Urban	9-11	4695	1.40 (1.36, 1.44)	9018	2.56 (2.51, 2.62)
Rural	12-14	323	1.21 (1.08, 1.34)	946	3.02 (2.83, 3.21)
Urban	12-14	4062	1.20 (1.16, 1.23)	8768	2.49 (2.44, 2.54)
Rural	15-18	112	0.97 (0.79, 1.15)	287	1.90 (1.68, 2.12)
Urban	15-18	3363	0.91 (0.88, 0.94)	6878	1.74 (1.70, 1.79)

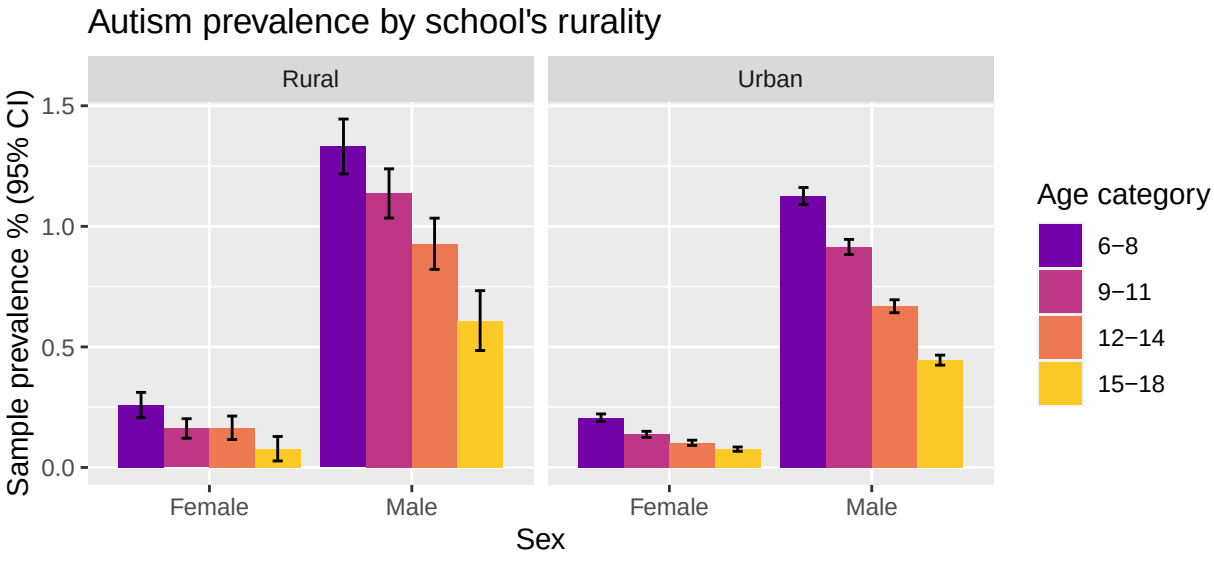

Figure 15: Sample prevalence of autism in school data by school's rurality, age band and sex. Bars show 95% normal confidence intervals.

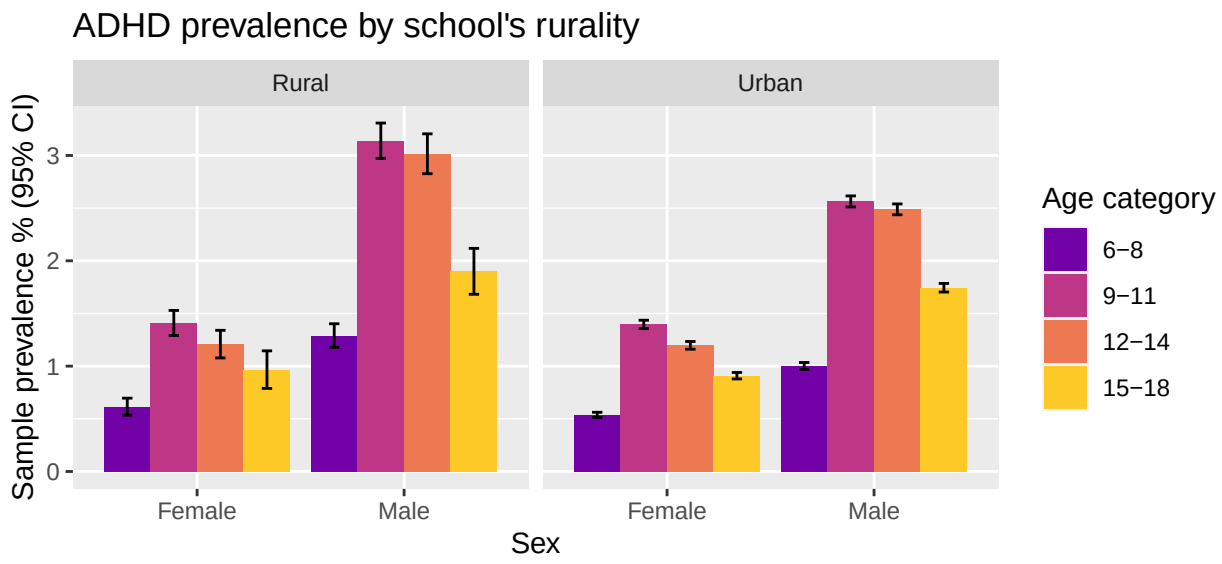

Figure 16: Sample prevalence of ADHD in school data by school's rurality, age band and sex. Bars show 95% normal confidence intervals.

Table 14: Crude and age- and sex-adjusted autism prevalence by health service in Chile school data. Crude prevalence has 95% normal confidence intervals and adjusted prevalence has 95% gamma confidence intervals.

Health service	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Aconcagua	0.44 (0.38, 0.50)	0.43 (0.37, 0.50)
Aisén	0.75 (0.63, 0.87)	0.75 (0.63, 0.90)
Antofagasta	0.84 (0.79, 0.89)	0.83 (0.77, 0.88)
Araucanía Norte	0.30 (0.24, 0.36)	0.30 (0.24, 0.38)
Araucanía Sur	0.37 (0.34, 0.40)	0.37 (0.34, 0.41)
Arauco	0.73 (0.64, 0.83)	0.72 (0.62, 0.82)
Arica	0.61 (0.54, 0.68)	0.61 (0.54, 0.70)
Atacama	0.31 (0.26, 0.35)	0.31 (0.27, 0.37)
Biobío	0.43 (0.38, 0.48)	0.42 (0.37, 0.47)
Chiloé	0.45 (0.38, 0.52)	0.43 (0.36, 0.52)
Concepción	0.78 (0.73, 0.84)	0.77 (0.72, 0.83)
Coquimbo	0.41 (0.38, 0.45)	0.40 (0.36, 0.43)
Iquique	0.45 (0.40, 0.50)	0.43 (0.38, 0.49)
Magallanes	0.83 (0.72, 0.94)	0.83 (0.72, 0.96)
Maule	0.31 (0.28, 0.33)	0.30 (0.28, 0.33)
Metro. Central	0.42 (0.38, 0.46)	0.42 (0.38, 0.46)
Metro. Norte	0.29 (0.27, 0.32)	0.29 (0.26, 0.31)
Metro. Occidente	0.36 (0.34, 0.38)	0.34 (0.32, 0.36)
Metro. Oriente	0.30 (0.28, 0.33)	0.30 (0.27, 0.33)
Metro. Sur	0.41 (0.39, 0.44)	0.40 (0.37, 0.43)
Metro. Sur Oriente	0.37 (0.34, 0.39)	0.36 (0.34, 0.39)
O'Higgins	0.43 (0.40, 0.47)	0.42 (0.39, 0.46)
Osorno	0.44 (0.38, 0.51)	0.43 (0.37, 0.51)
Reloncaví	0.42 (0.38, 0.47)	0.42 (0.37, 0.47)
Talcahuano	0.84 (0.76, 0.92)	0.81 (0.74, 0.90)

Health service	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Valdivia	0.31 (0.27, 0.35)	0.30 (0.26, 0.35)
Valparaíso	0.69 (0.63, 0.74)	0.68 (0.62, 0.74)
Viña del Mar	0.67 (0.63, 0.71)	0.66 (0.62, 0.70)
Ñuble	1.32 (1.24, 1.40)	1.29 (1.21, 1.37)

Table 15: Crude and age- and sex-adjusted ADHD prevalence by health service in Chile school data. Crude prevalence has 95% normal confidence intervals and adjusted prevalence has 95% gamma confidence intervals.

Health service	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Aconcagua	2.08 (1.95, 2.21)	2.04 (1.91, 2.19)
Aisén	2.23 (2.03, 2.44)	2.17 (1.97, 2.40)
Antofagasta	1.00 (0.94, 1.06)	0.98 (0.93, 1.04)
Araucanía Norte	1.33 (1.22, 1.45)	1.29 (1.18, 1.43)
Araucanía Sur	1.42 (1.36, 1.49)	1.38 (1.32, 1.45)
Arauco	1.64 (1.50, 1.78)	1.64 (1.50, 1.81)
Arica	1.14 (1.04, 1.24)	1.12 (1.02, 1.24)
Atacama	0.49 (0.44, 0.55)	0.49 (0.43, 0.56)
Biobío	2.27 (2.16, 2.38)	2.26 (2.15, 2.38)
Chiloé	2.96 (2.77, 3.15)	2.87 (2.68, 3.07)
Concepción	2.94 (2.84, 3.04)	3.00 (2.89, 3.11)
Coquimbo	1.98 (1.91, 2.05)	2.00 (1.92, 2.08)
Iquique	1.49 (1.40, 1.58)	1.50 (1.40, 1.60)
Magallanes	3.07 (2.87, 3.27)	3.06 (2.85, 3.29)
Maule	1.19 (1.14, 1.24)	1.15 (1.11, 1.21)
Metro. Central	1.53 (1.46, 1.59)	1.49 (1.43, 1.57)
Metro. Norte	1.42 (1.36, 1.47)	1.42 (1.36, 1.48)
Metro. Occidente	1.09 (1.05, 1.13)	1.12 (1.08, 1.17)
Metro. Oriente	1.22 (1.17, 1.27)	1.20 (1.15, 1.26)
Metro. Sur	1.42 (1.37, 1.48)	1.41 (1.35, 1.46)
Metro. Sur Oriente	1.56 (1.51, 1.61)	1.53 (1.48, 1.58)
O'Higgins	1.73 (1.66, 1.79)	1.69 (1.63, 1.76)
Osorno	1.05 (0.95, 1.14)	1.02 (0.92, 1.13)
Reloncaví	1.02 (0.95, 1.09)	0.99 (0.92, 1.07)
Talcahuano	3.07 (2.93, 3.22)	3.02 (2.87, 3.18)
Valdivia	1.08 (1.00, 1.16)	1.06 (0.98, 1.15)
Valparaíso	1.19 (1.12, 1.27)	1.19 (1.11, 1.27)
Viña del Mar	1.17 (1.12, 1.22)	1.15 (1.10, 1.20)
Ñuble	2.12 (2.02, 2.22)	2.11 (2.00, 2.22)

For both autism and ADHD, adjusted prevalences by school fees, ethnicity and school's rurality show similar patterns to crude prevalences, except for students at rural schools which have notably lower adjusted prevalence for both conditions. See tables 16 - 21. These results indicate that for autism and ADHD there are differences in prevalence across location and demographic features.

Table 16: Crude and age- and sex-adjusted autism prevalence by monthly school fee (Peso) in Chile school data. Crude prevalence has 95% normal confidence intervals and adjusted prevalence has 95% gamma confidence intervals.

School fee	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Free	0.57 (0.56, 0.58)	0.55 (0.54, 0.56)
$1,000-$10,000	0.71 (0.22, 1.21)	0.69 (0.29, 3.25)
$10,001-$25,000	0.20 (0.15, 0.25)	0.20 (0.16, 0.27)
$25,001-$50,000	0.30 (0.28, 0.33)	0.32 (0.29, 0.34)
$50,001-$100,000	0.39 (0.36, 0.41)	0.40 (0.37, 0.43)
$100,001+	0.05 (0.04, 0.05)	0.05 (0.04, 0.06)
No information	0.50 (0.44, 0.57)	0.46 (0.40, 0.52)

Table 17: Crude and age- and sex-adjusted ADHD prevalence by monthly school fee (Peso) in Chile school data. Crude prevalence has 95% normal confidence intervals and adjusted prevalence has 95% gamma confidence intervals.

School fee	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Free	1.65 (1.63, 1.67)	1.62 (1.60, 1.64)
$1,000-$10,000	0.18 (0.00, 0.43)	0.11 (0.01, 2.74)
$10,001-$25,000	1.06 (0.95, 1.16)	1.05 (0.94, 1.17)
$25,001-$50,000	1.59 (1.54, 1.65)	1.65 (1.59, 1.72)
$50,001-$100,000	1.90 (1.84, 1.95)	1.90 (1.84, 1.96)
$100,001+	0.22 (0.21, 0.24)	0.23 (0.21, 0.25)
No information	1.21 (1.11, 1.31)	1.22 (1.13, 1.33)

Table 18: Crude and age- and sex-adjusted autism prevalence by ethnicity in Chile school data. Crude prevalence has 95% normal confidence intervals and adjusted prevalence has 95% gamma confidence intervals.

Ethnicity	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Mapuche	0.36 (0.33, 0.39)	0.34 (0.31, 0.37)
Other Indigenous group	0.37 (0.20, 0.54)	0.38 (0.21, 0.75)
No Indigenous group	0.43 (0.42, 0.44)	0.42 (0.41, 0.43)

Table 19: Crude and age- and sex-adjusted ADHD prevalence by ethnicity in Chile school data. Crude prevalence has 95% normal confidence intervals and adjusted prevalence has 95% gamma confidence intervals.

Ethnicity	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Mapuche	1.38 (1.32, 1.44)	1.33 (1.28, 1.39)
Other Indigenous group	1.12 (0.82, 1.41)	1.10 (0.83, 1.55)
No Indigenous group	1.59 (1.57, 1.61)	1.58 (1.56, 1.60)

Table 20: Crude and age- and sex-adjusted autism prevalence by school's rurality in Chile school data. Crude prevalence has 95% normal confidence intervals and adjusted prevalence has 95% gamma confidence intervals.

School rurality	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Rural	0.66 (0.63, 0.69)	0.57 (0.54, 0.61)
Urban	0.46 (0.45, 0.47)	0.46 (0.45, 0.46)

Table 21: Crude and age- and sex-adjusted ADHD prevalence by school's rurality in Chile school data. Crude prevalence has 95% normal confidence intervals and adjusted prevalence has 95% gamma confidence intervals.

School rurality	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
Rural	1.77 (1.72, 1.82)	1.67 (1.61, 1.74)
Urban	1.49 (1.48, 1.50)	1.48 (1.47, 1.50)

5.3 Clinical data

The small clinical dataset has data on 1,570 appointments for 253 unique patients aged 1-18 in 2021, of which 247 patients have autism and have lived in a commune in the SSAS catchment area during the period the data covers. Among the patients with autism that live in SSAS, 55 (22.27%) are female, 4 (1.62%) are Mapuche, 221 (89.47%) live in an urban area, 12 (4.86%) have a disability and 2 (0.81%) have experienced foster care. 17 (6.88%) have an intellectual disability as well as autism.

5.4 Machine learning with clinical data

Multiple correspondence analysis was first conducted with all features thought to be associated with autism diagnosis with no imputation. Figure 17 shows approximately 14.6% of the variance in this data can be captured by the first two dimensions of MCA. Disability and foster care status are well separated by the first dimension but Figures 26 and27 show that this separation is primarily driven by whether information is available for these features. Ethnicity, age band and commune of residence are well separately by the second dimension of MCA, as shown in Figure 17, and are somewhat separated by the first dimension. Figures 18 and 19 further shows the importance of categories within the foster care, disabilities, ethnicity, age band and commune features for explaining the variance in this data. In particular, categories that explain more of the variance include not having disability or foster care status, or not having information on these, age 0-2 and 3-5, being foreign or Chilean, living in Toltén and having private health insurance. Examining the clustering of individual patients by the first two dimensions of MCA, Figure 20 demonstrates that patients in age bands 0-2 and 3-5 cluster well and those in older age bands do not. There is some clustering by ethnicity in Figure 21 but it is obscured by the separation of points into the two larger clustered defined by having or not having information on disability and foster care status. Figure 22 shows clear separation of Toltén and Nueva Imperial communes, however it is important to know here that these and several other communes are represented by only one patient, see Table 22. There is decent clustering of patients in Temuco and Pitrufquén communes. Figure 23 shows possible separation for patients with private health insurance and Figures 24 and 25 show little clustering by sex or rurality of residence.

Exchanging disability and foster care status for their imputed versions leads to the MCA capturing approximately 14.1% of the variance in that data with its first two dimensions, as shown in Figure 28. Disability and foster care status are no longer well separated by the first dimension and Figures 37 and 38 show that the patients who have experienced foster care do cluster well but the patients with a disability do not. In Figure 28, age band and ethnicity are now well separated by both dimensions and commune mostly by the second.

Table 22: Count and percentage of features' values in the small clinical dataset.

Feature	Available values	Count (%)
Sex	Female	55 (22.27%)
Sex	Male	192 (77.73%)
Age band	Age 0-2	13 (5.26%)
Age band	Age 3-5	55 (22.27%)
Age band	Age 6-8	37 (14.98%)
Age band	Age 9-11	56 (22.67%)
Age band	Age 12-14	45 (18.22%)
Age band	Age 15-18	41 (16.60%)
Private health level	FONASA - A	99 (40.08%)
Private health level	FONASA - B	67 (27.13%)
Private health level	FONASA - C	35 (14.17%)
Private health level	FONASA - D	38 (15.38%)
Private health level	Private Health Insurance	8 (3.24%)
Commune	Curarrehue	6 (2.43%)
Commune	Freire	2 (0.81%)
Commune	Gorbea	1 (0.40%)
Commune	Loncoche	39 (15.79%)
Commune	Nueva Imperial	1 (0.40%)
Commune	Pitrufquén	2 (0.81%)
Commune	Pucón	50 (20.24%)
Commune	Temuco	7 (2.83%)
Commune	Teodoro Schmidt	1 (0.40%)
Commune	Toltén	1 (0.40%)
Commune	Villarrica	137 (55.47%)
Rurality	Rural	26 (10.53%)
Rurality	Urban	221 (89.47%)
Ethnicity	Mapuche	4 (1.62%)
Ethnicity	Chilean	131 (53.04%)
Ethnicity	Foreign	32 (12.96%)
Ethnicity	No ethnicity information	80 (32.39%)
Disability	Yes disability	12 (4.86%)
Disability	No disability	78 (31.58%)
Disability	No disability information	157 (63.56%)
Foster care	Yes foster care	2 (0.81%)
Foster care	No foster care	88 (35.63%)
Foster care	No foster care information	157 (63.56%)
Intellectual disability	Yes intellectual disability	17 (6.88%)
Intellectual disability	No intellectual disability	230 (93.12%)

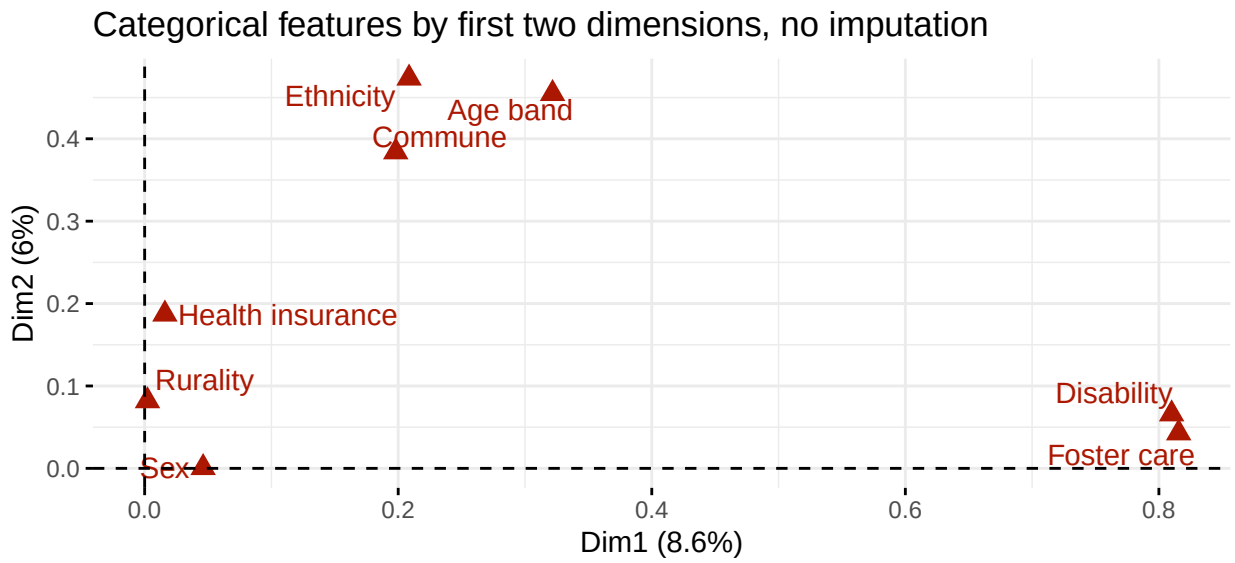

Figure 17: Categorical features by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation.

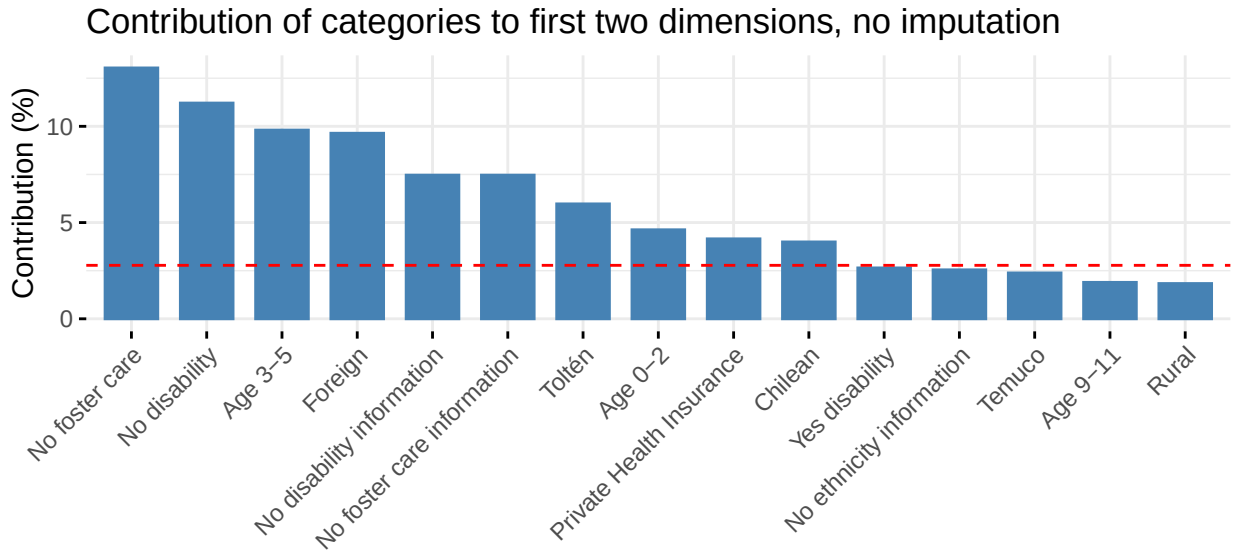

Figure 18: Contribution of the top 15 categories to the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation. The red line shows the expected average if contributions were uniform.

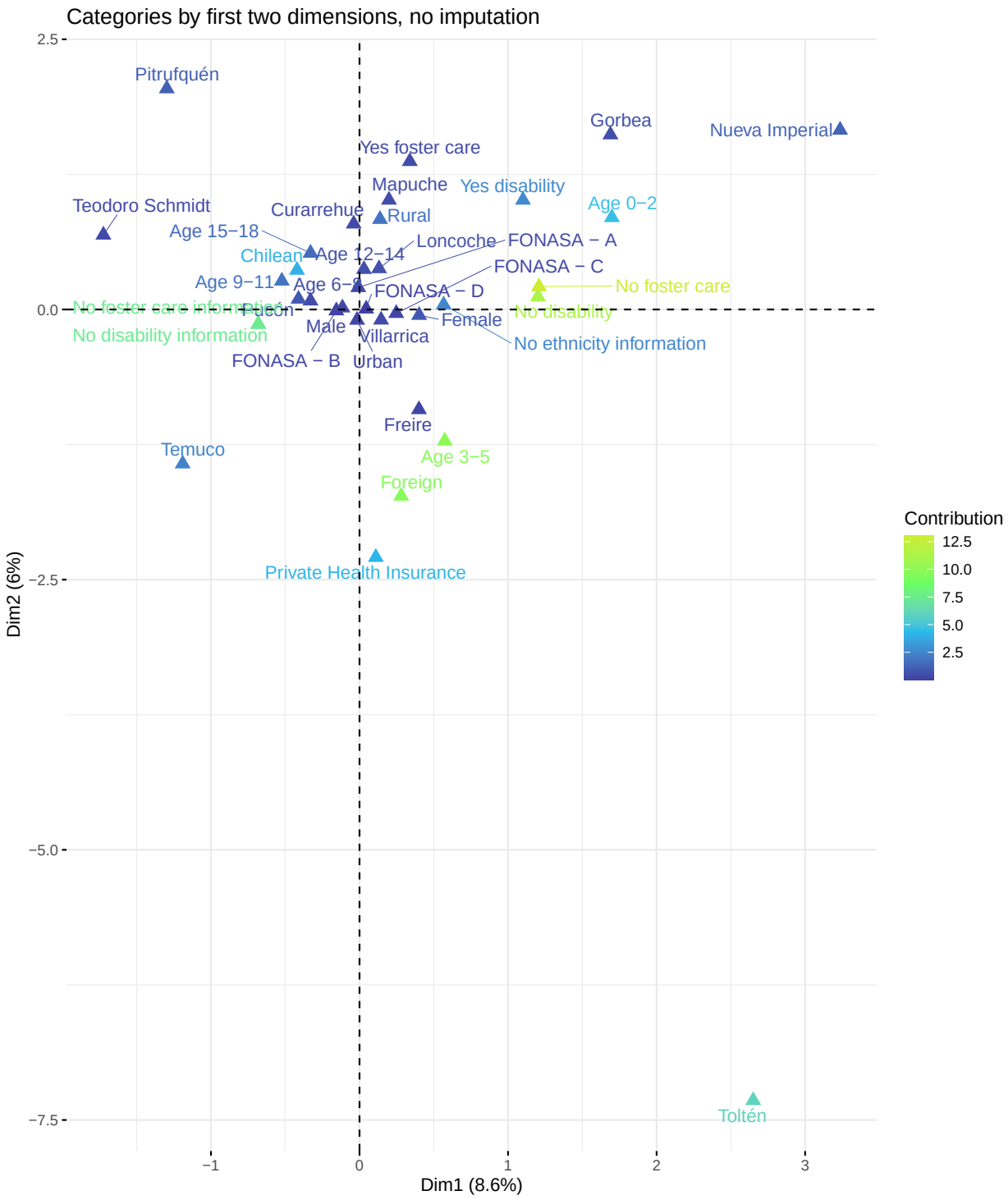

Figure 19: Available categories by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation. Brigther, more yellow colours indicate larger contribution to the first two dimensions.

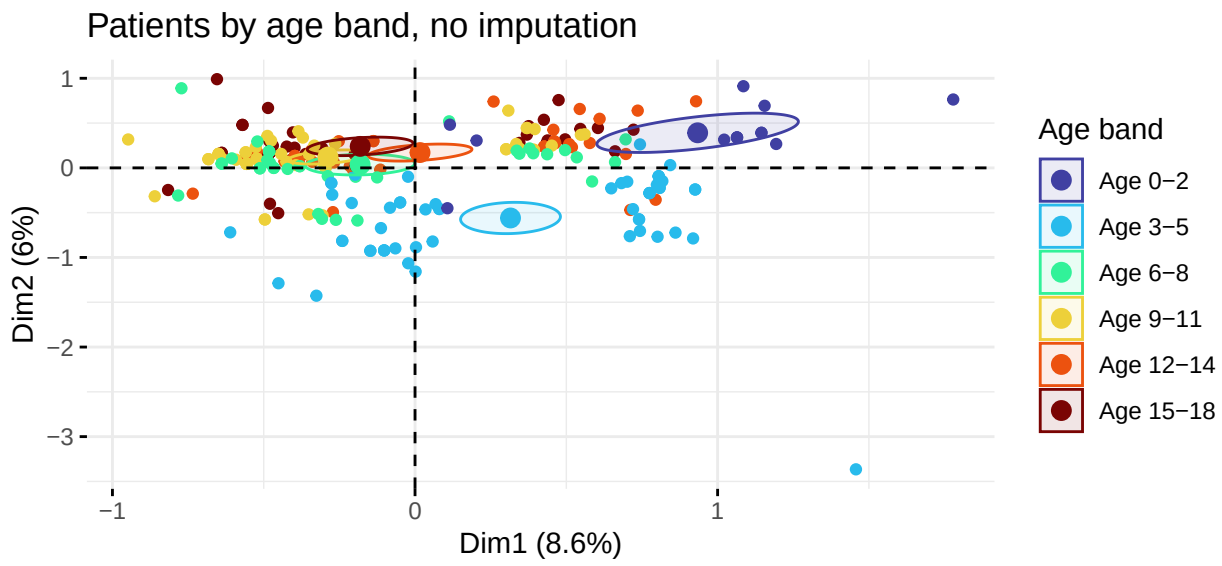

Figure 20: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation, coloured by age band.

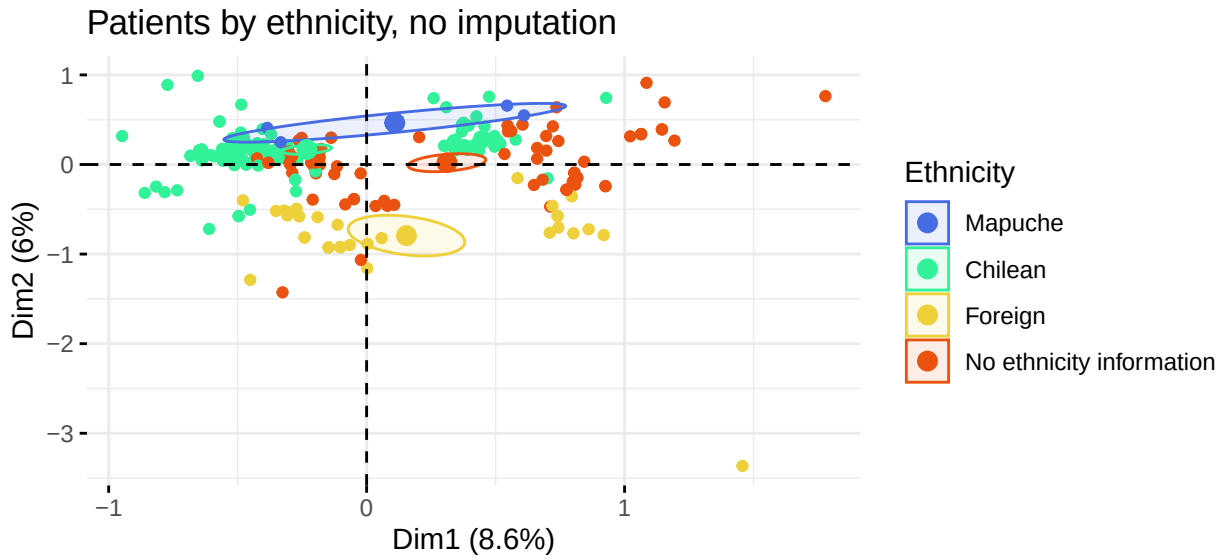

Figure 21: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation, coloured by ethnicity.

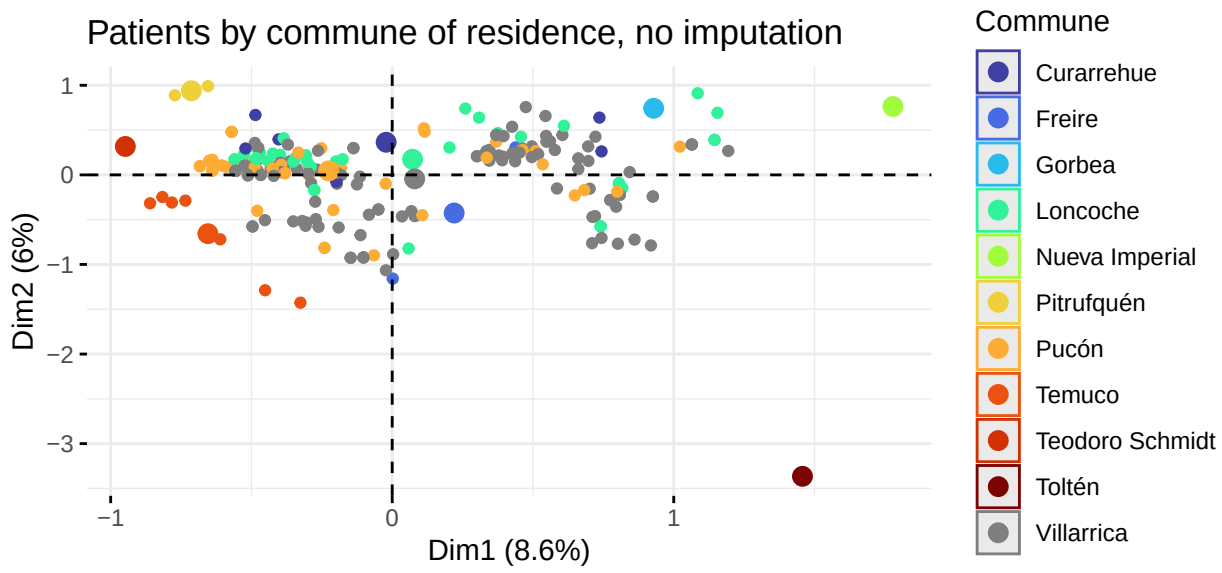

Figure 22: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation, coloured by commune of residence.

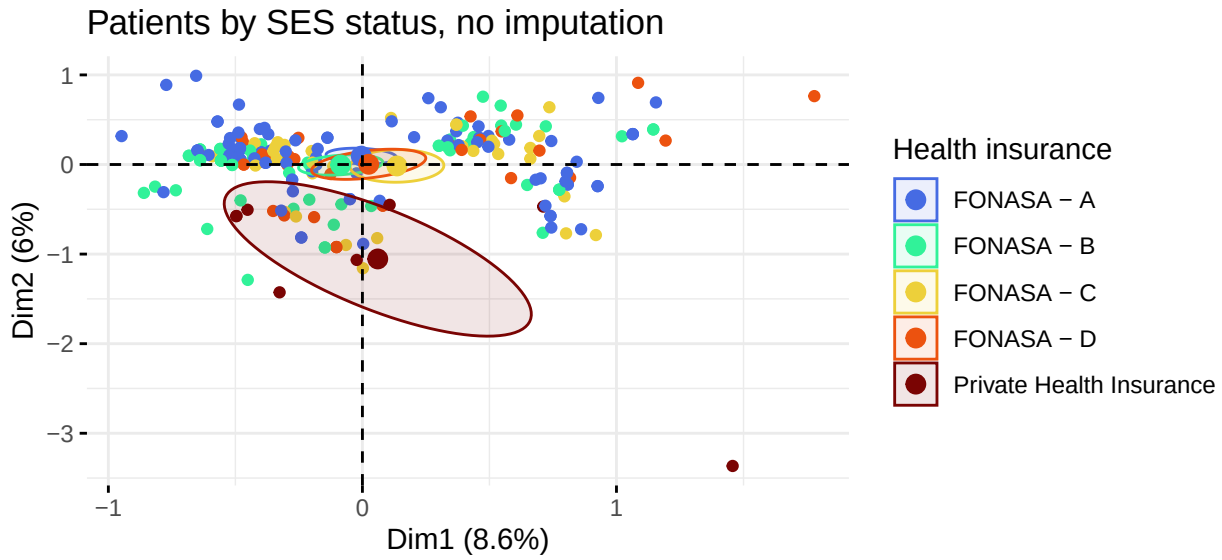

Figure 23: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation, coloured by health insurance level.

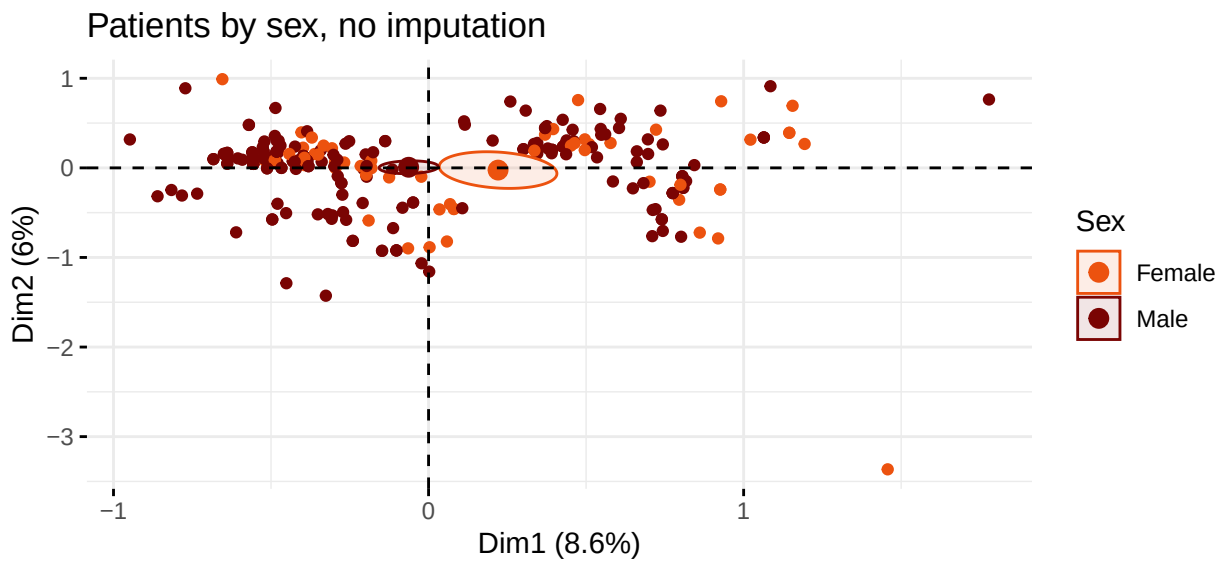

Figure 24: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation, coloured by sex.

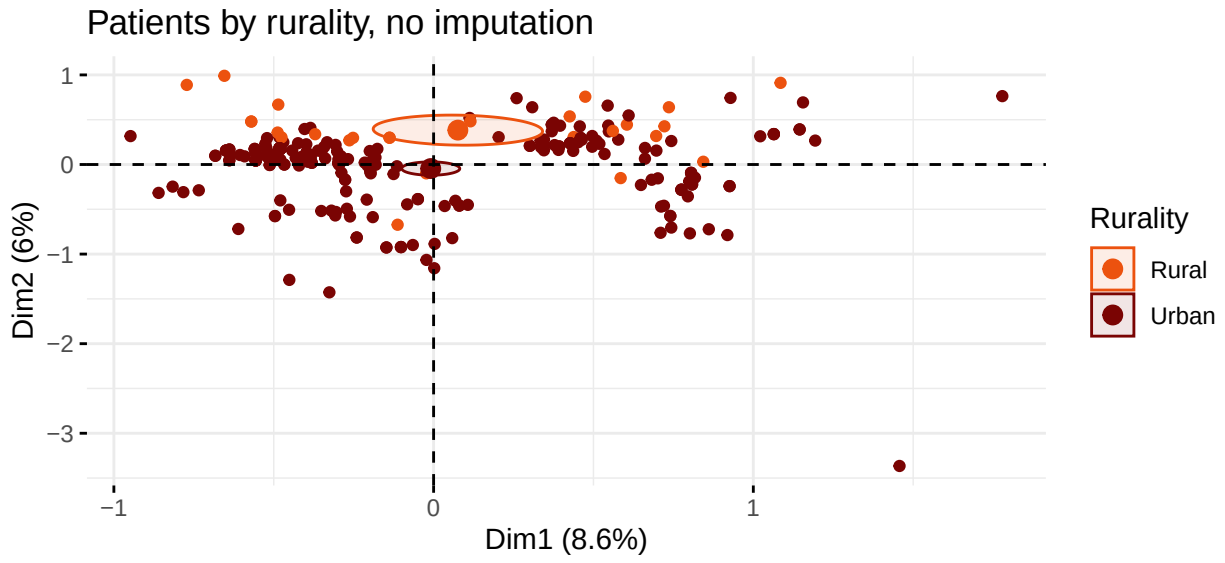

Figure 25: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation, coloured by rurality of residence.

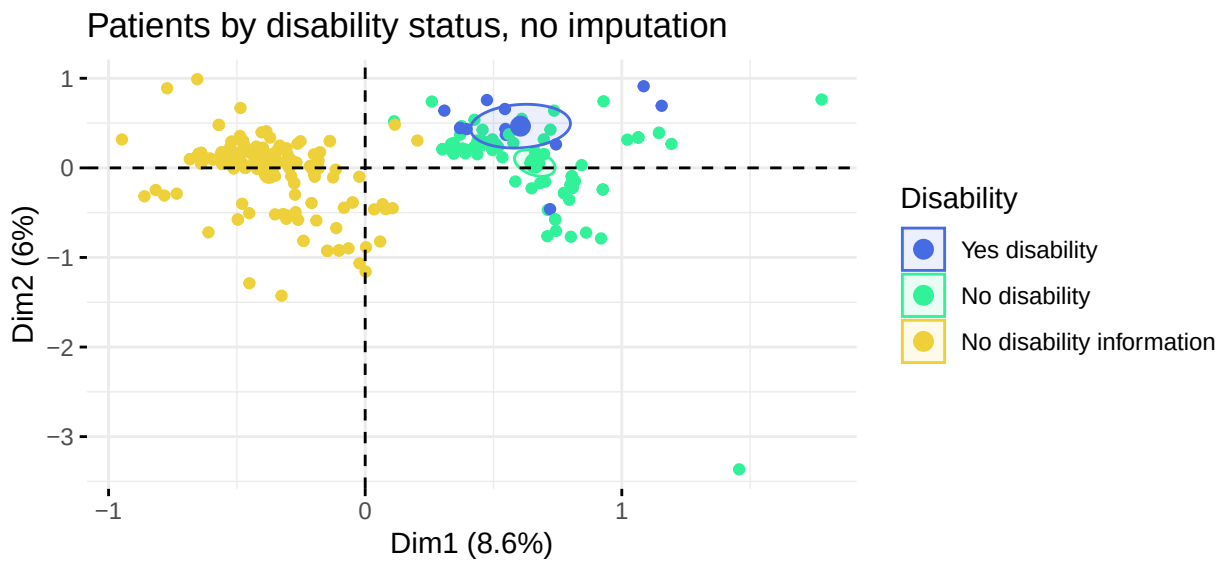

Figure 26: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation, coloured by disability status.

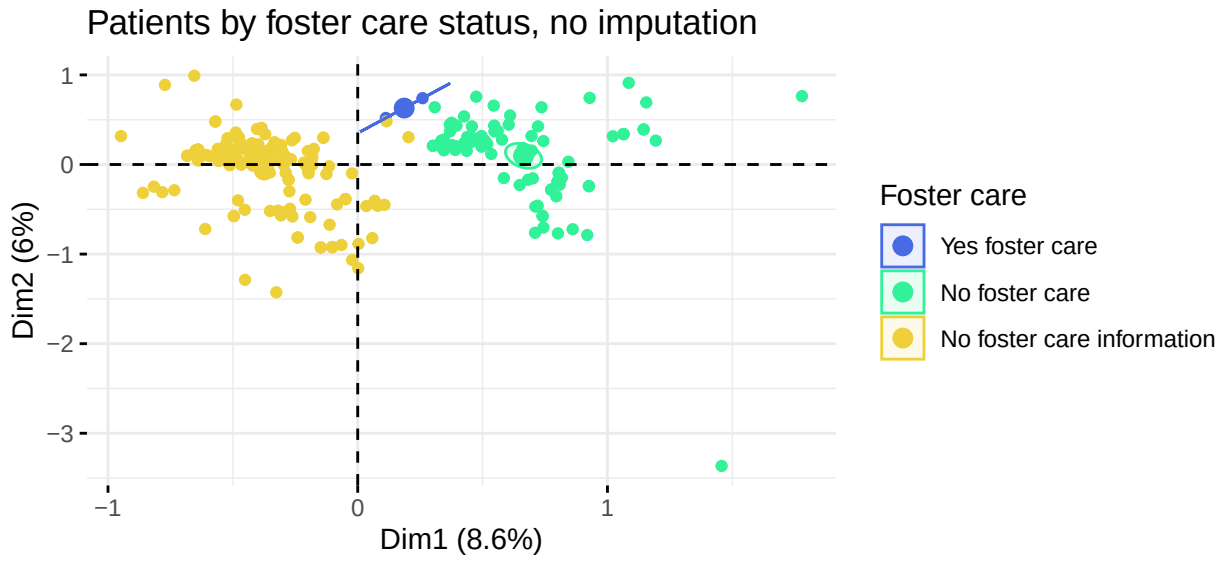

Figure 27: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features without imputation, coloured by foster care status.

With the reduced importance of disability and foster care, age bands 0-2 and 3-5, foreign, no information and Chilean ethnicity and Toltén and Nueva Imperial communes contribute most to the first two dimensions, see Figures 29 and 30. Again patients with age bands 0-2 and 3-5 cluster well and older age bands do not, as shown in Figure 31. Figure 32 shows much clearer clustering of ethnicity than before. For communes with more than one patient resident, clustering is less clear than before with only Pitrufquén well separated by these dimensions, see Figure 33. In Figures 34, 35 and 36, clustering by feature is weak.

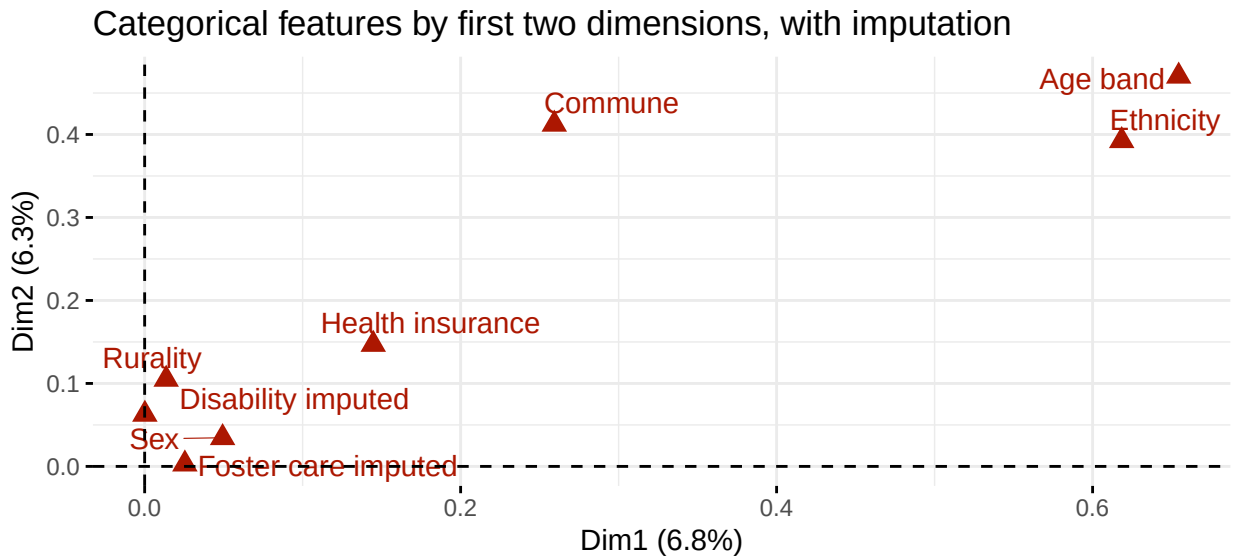

Figure 28: Categorical features by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation.

MCA on the small clinical data using only age band, ethnicity and commune results in the first two dimensions capturing 17.8% of the variance in that data, see Figure 39, and these features are fairly well represented by the first two dimensions. Again age bands 0-2 and 3-5, foreign, no information and Chilean ethnicity and Toltén and Nueva Imperial communes contribute most to the first two dimensions, see Figures 40 and 41. By patient, age bands 0-2 and 3-5 still cluster well (Figure 42), ethnicity clusters are very distinct (Figure 43), and communes show more structure than previously (Figure 44).

5.5 Linkage of school and patient records

In the school records, 132,242 students live in SSAS health service catchment, of which 488 (0.37%) have autism.

Aggregating the combined clinical data to patient-level data for linkage resulted in the patient dataset with 1,376 records for 1,365 unique patients as 9 patients lived in 2 communes and 1 lived in 3 during the period covered by the data.

5.5.1 Manual record linkage

Using perfect match on sex, date of birth, commune of residence and the proxies for SES, 79 matches can be found between the SSAS school and patient records. Of these, 77 unique SSAS school records can be perfectly matched to SSAS patient records, and all perfect matches were for unique patients. When mismatch on SES proxy is allowed, 197 matches can be manually found between the SSAS school and patient records. Of these, 188 unique school records can be perfectly matched to SSAS patient records and 193 SSAS patients can be perfectly matched to SSAS school records.

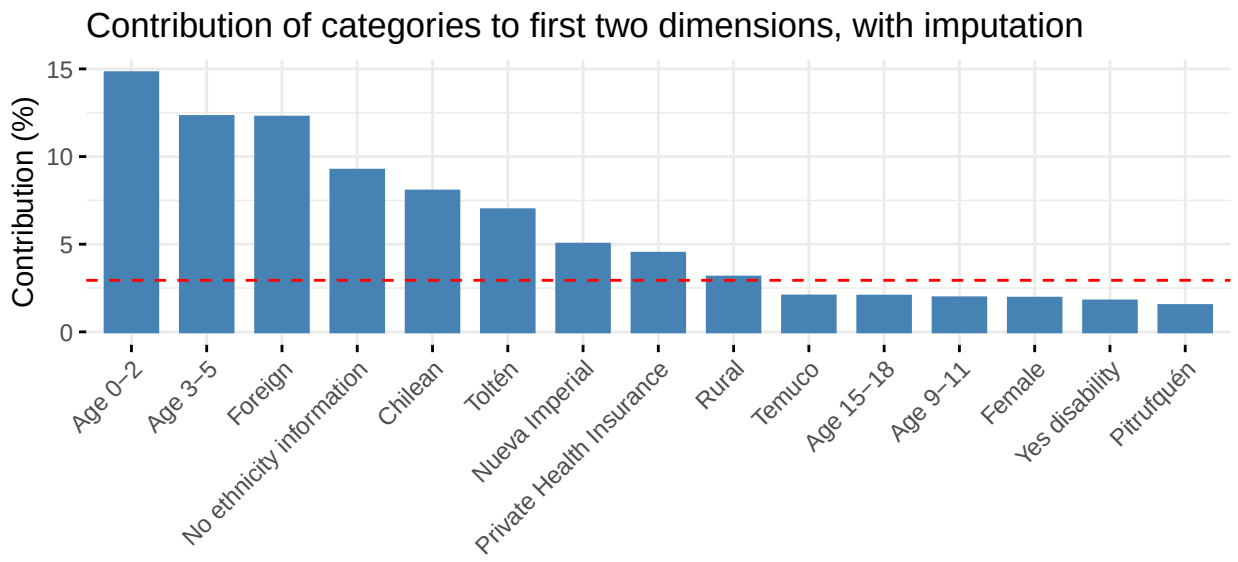

Figure 29: Contribution of the top 15 categories to the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation. The red line shows the expected average if contributions were uniform.

5.5.2 Probabilistic record linkage

Blocking on sex and date of birth, resulted in 293 blocked pairs. Probabilistic matching on sex, date of birth, commune of residence and the proxies for SES with selection of possible matches to create a bijective set of matches resulted in 233 matches of unique SSAS school and patient records. This corresponds to 47.65% of the school records for students with autism in SSAS having a match in the SSAS patient records, 16.93% of the patient records having a match in the SSAS school records and 17.07% of the unique patients having a match in the SSAS school records. For each patient that had lived in more than one commune and therefore appeared more than once in the patient data, only one match to an SSAS school record was made, meaning the matching was bijective for SSAS school records and unique patients.

Analysis of differences between matched and unmatched records in the SSAS school data and in the patient data by sex, commune and proxy for SES are provided in the Supplementary Figures. Kolmogorov-Smirnov permutation tests found no significant difference in frequency of sexes between matched and unmatched SSAS school records, see Figure 61. They found a strongly significant difference in the frequency of sexes between matched and unmatched patient records, see Figure 62. This difference is likely due to the sex ratios differing across the datasets: the SSAS school data is 12.47% female, the patient data is 20.06% female and the matches are 12.5% females. Kolmogovo-Smirnov permutation testing found matched and unmatched records differed significantly by commune for the SSAS school data, see Figure 63, and that there was no significant difference between matched and unmatched records by commune for the patient data, see Figure 63. This appears to be driven by the matchability of students and patients living in Temuco, the most prevalent commune. By proxy SES, there was a strongly significant difference between matched and unmatched records in the SSAS school and a somewhat significant difference for the patient data, see Figure 65 and Figure 66 respectively, likely reflecting different frequencies in these values across datasets. Kolmogorov-Smirnov permutation testing was not conducted for date of birth as this feature contains too many categories for results to be meaningful.

5.6 Updated autism prevalence estimates and delta

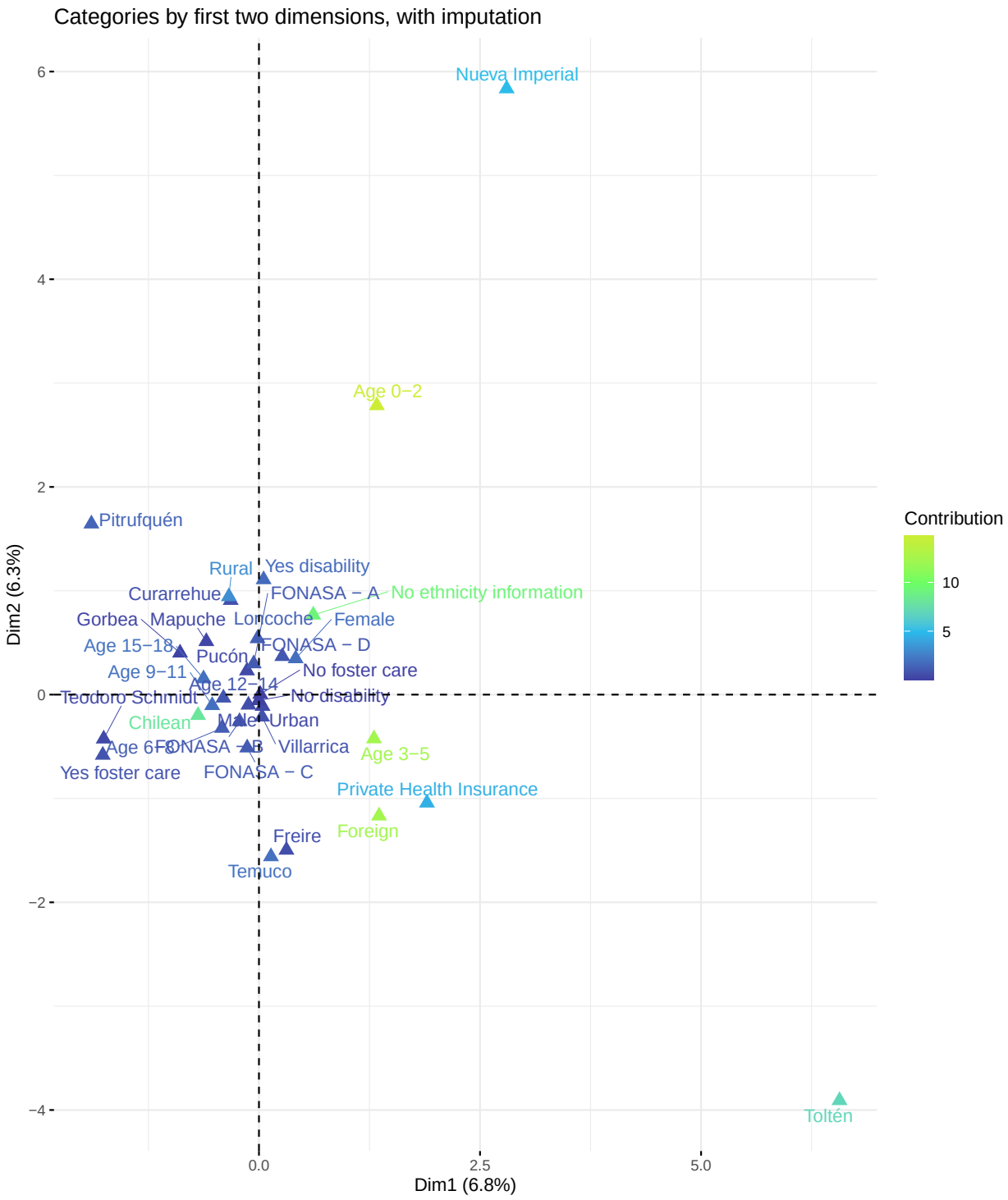

Figure 30: Available categories by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation. Brigther, more yellow colours indicate larger contribution to the first two dimensions.

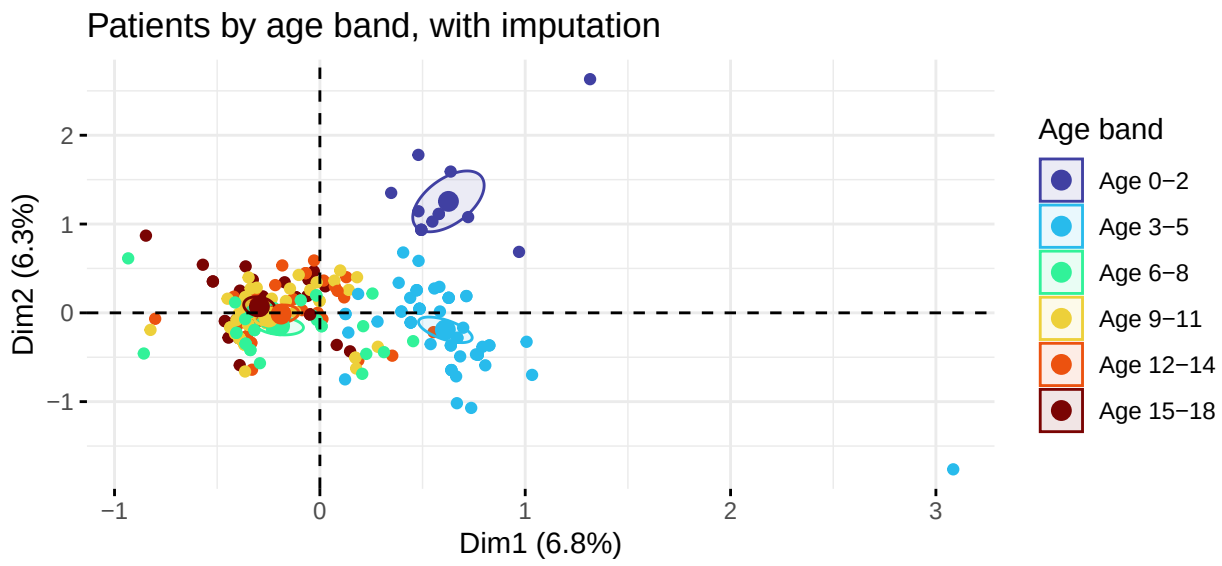

Figure 31: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation, coloured by age band.

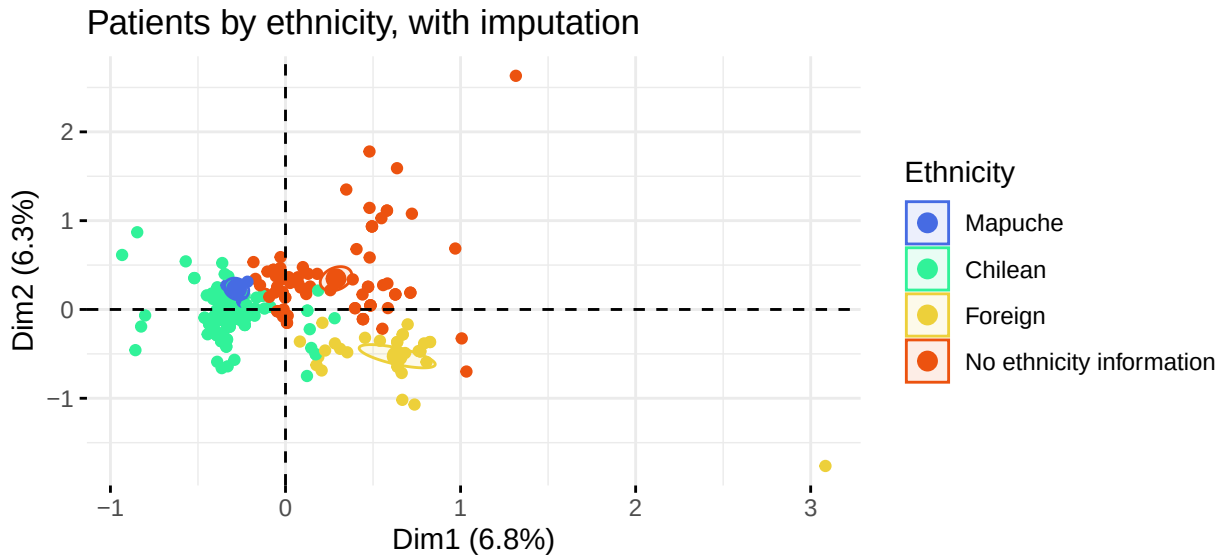

Figure 32: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation, coloured by ethnicity.

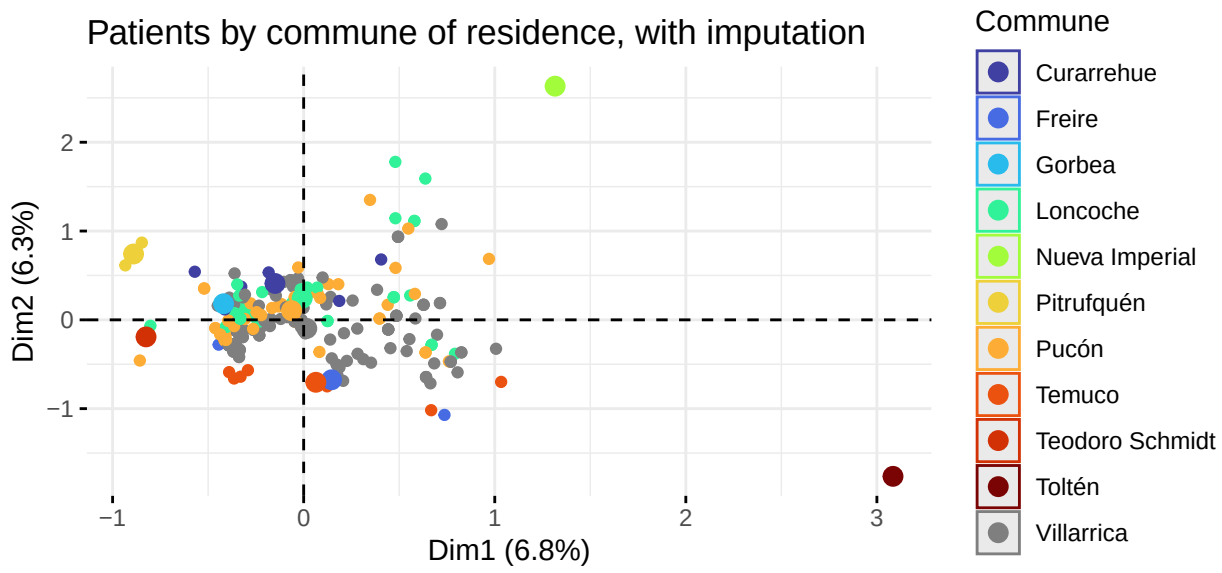

Figure 33: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation, coloured by commune of residence.

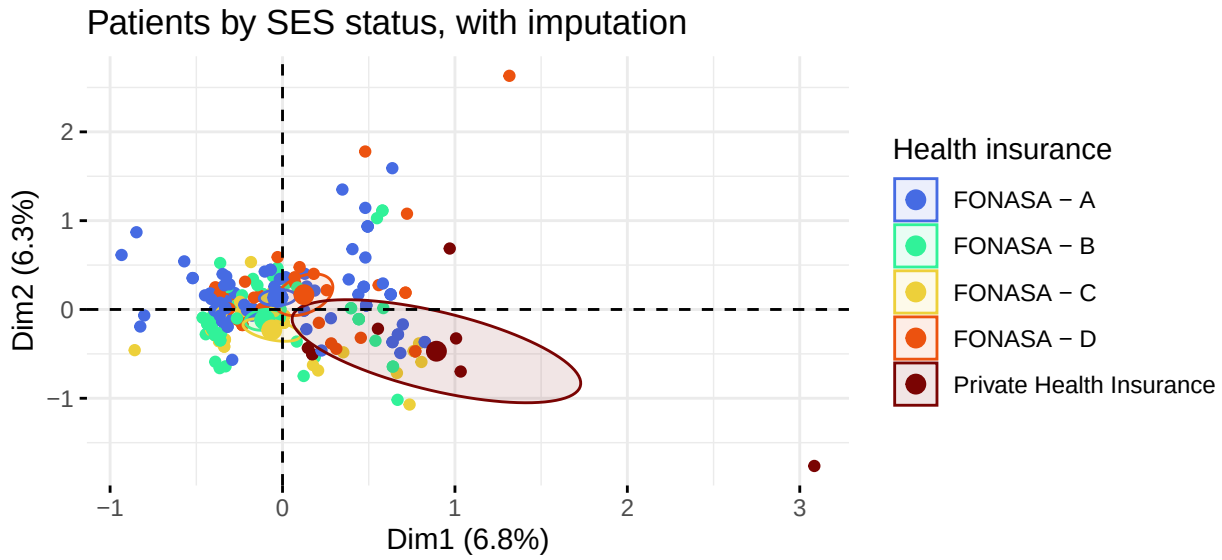

Figure 34: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation, coloured by health insurance level.

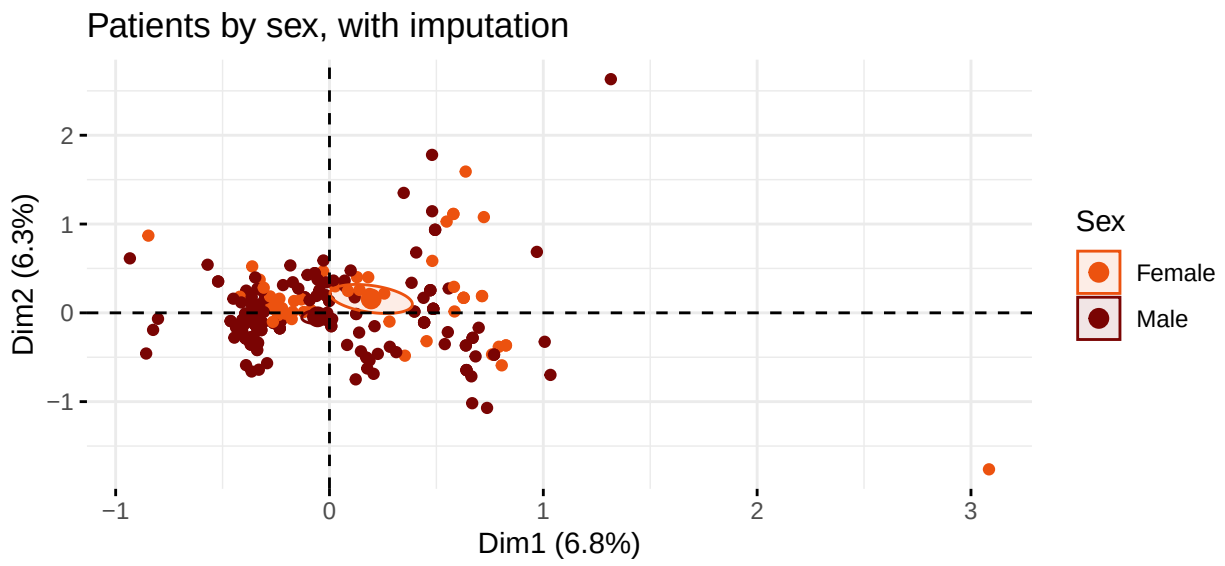

Figure 35: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation, coloured by sex.

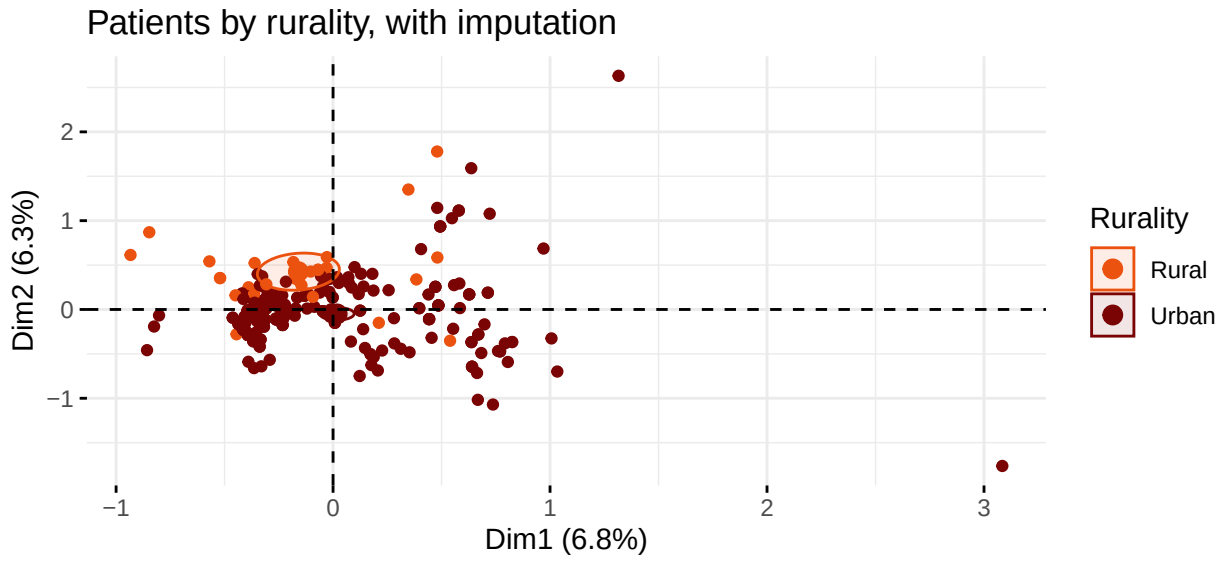

Figure 36: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation, coloured by rurality of residence.

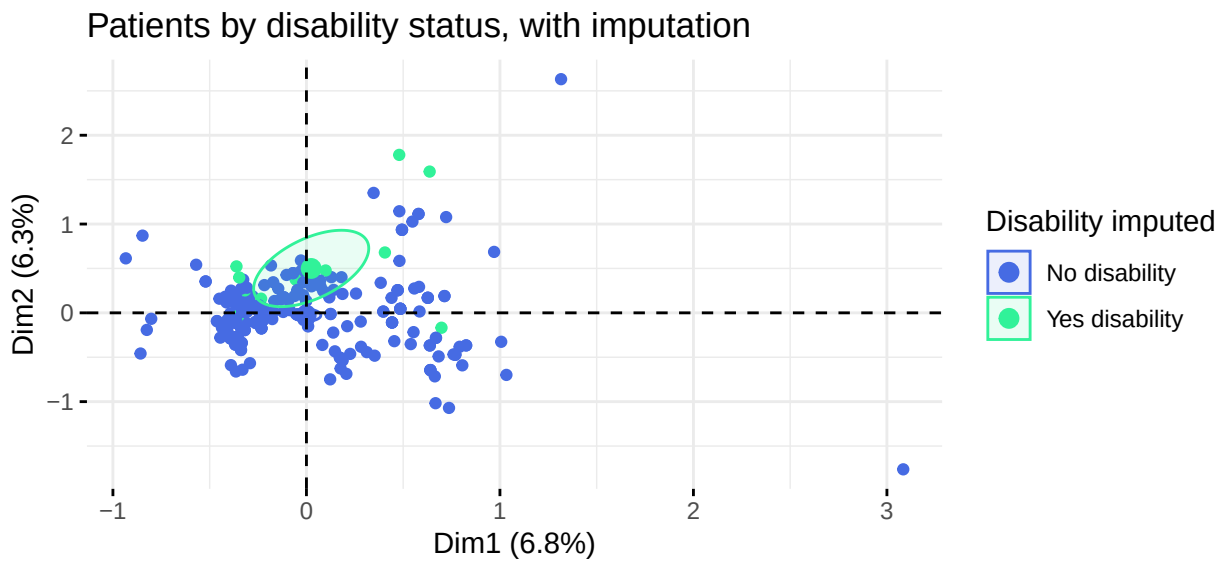

Figure 37: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation, coloured by disability status.

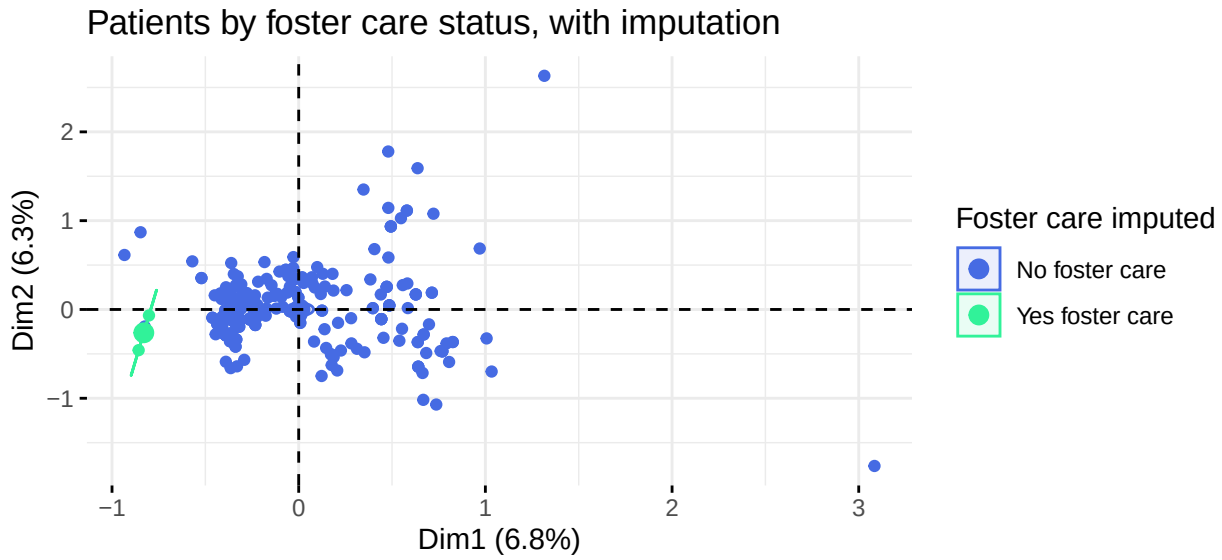

Figure 38: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using all features with imputation, coloured by foster care status.

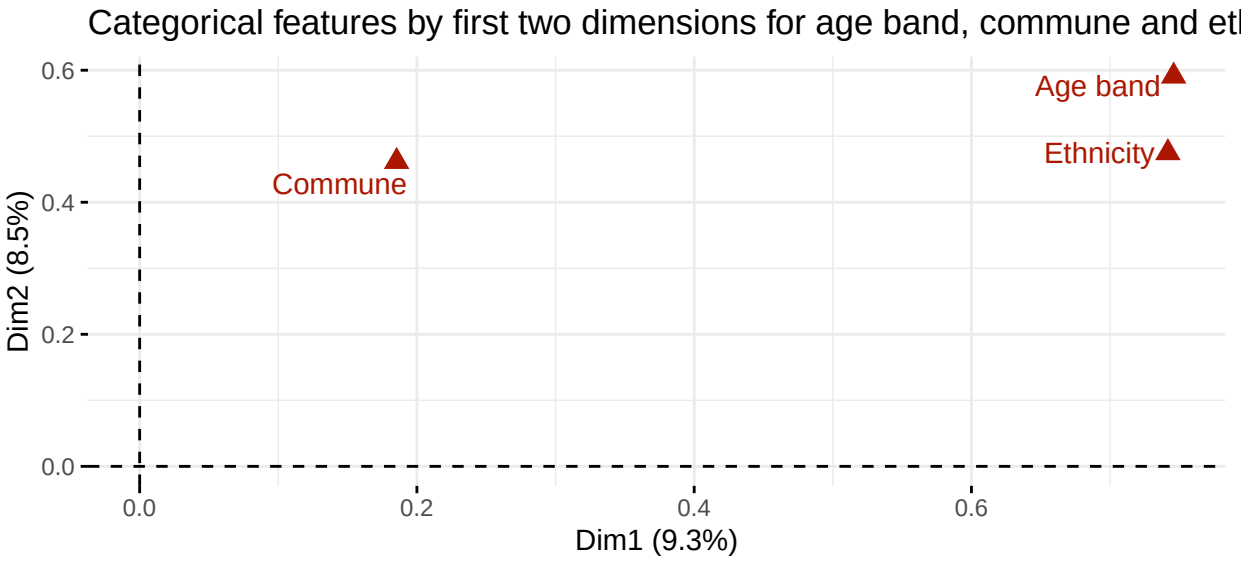

Figure 39: Categorical features by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using patients' age band, commune of residence and ethnicity.

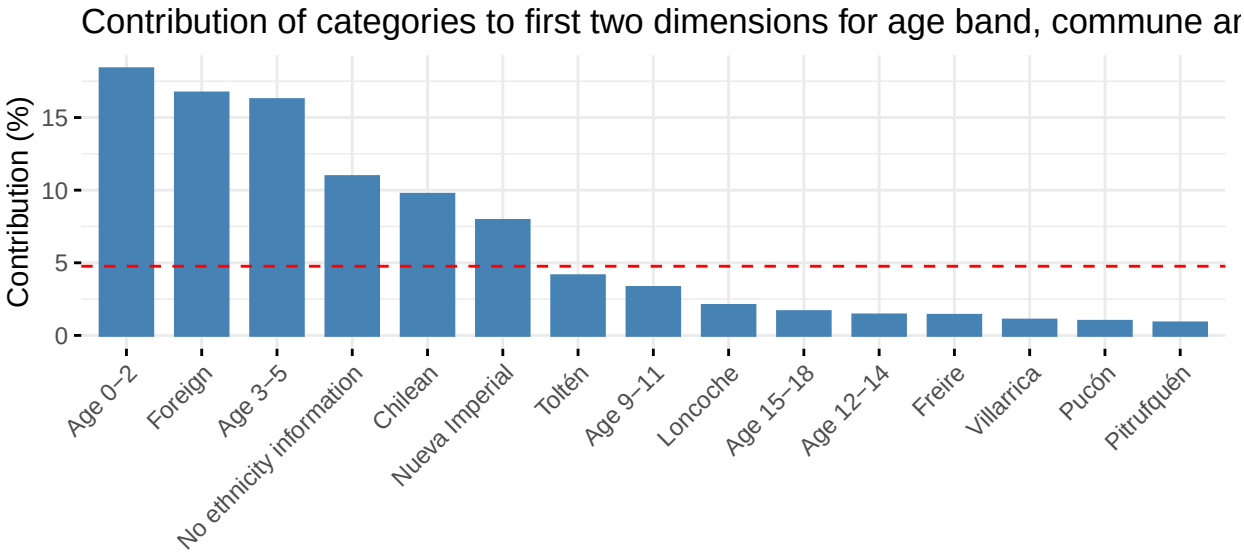

Figure 40: Contribution of the top 15 categories to the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using patients' age band, commune and ethnicity. The red line shows the expected average if contributions were uniform.

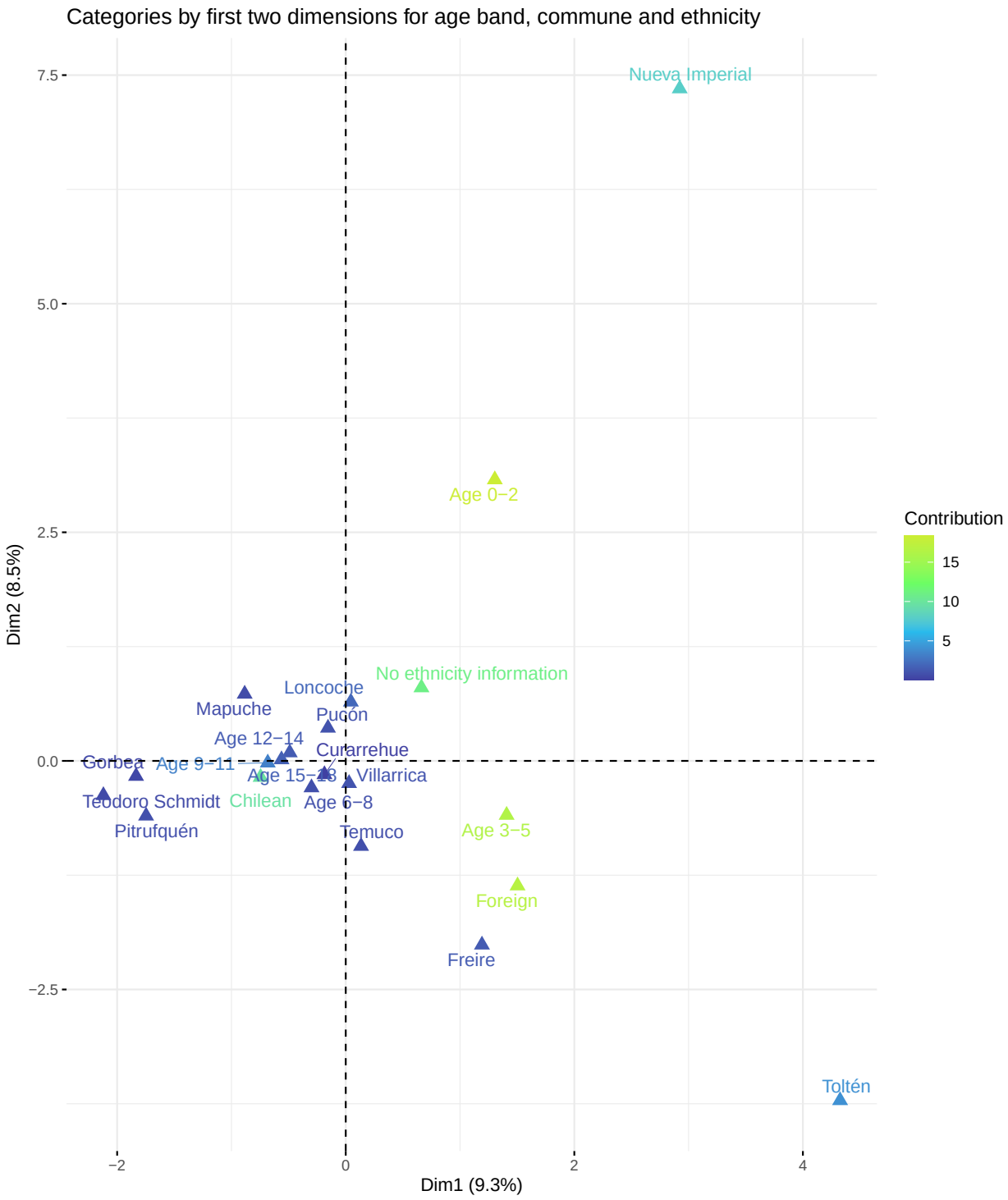

Figure 41: Available categories by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using patients' age band, commune and ethnicity. Brigther, more yellow colours indicate larger contribution to the first two dimensions.

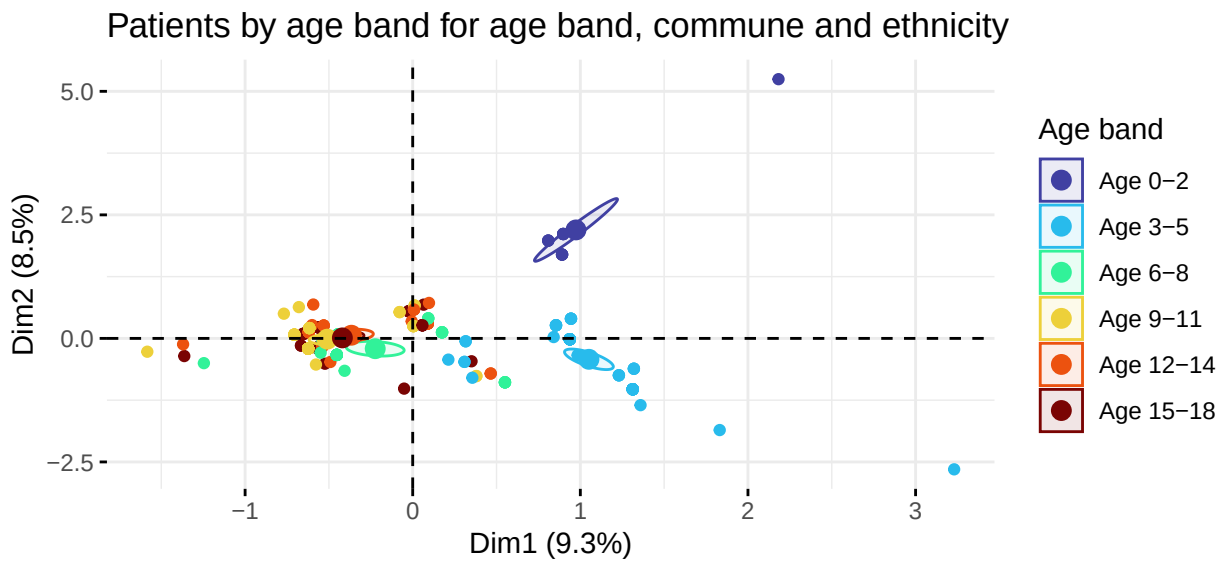

Figure 42: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using patients' age band, commune and ethnicity, coloured by age band.

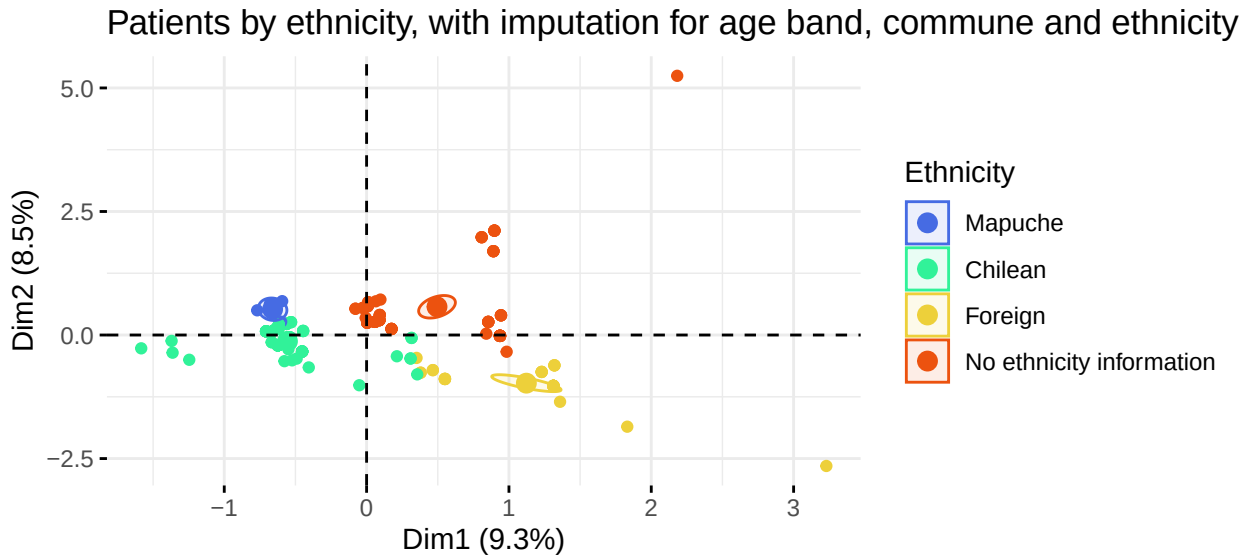

Figure 43: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using age band, commune and ethnicity, coloured by ethnicity.

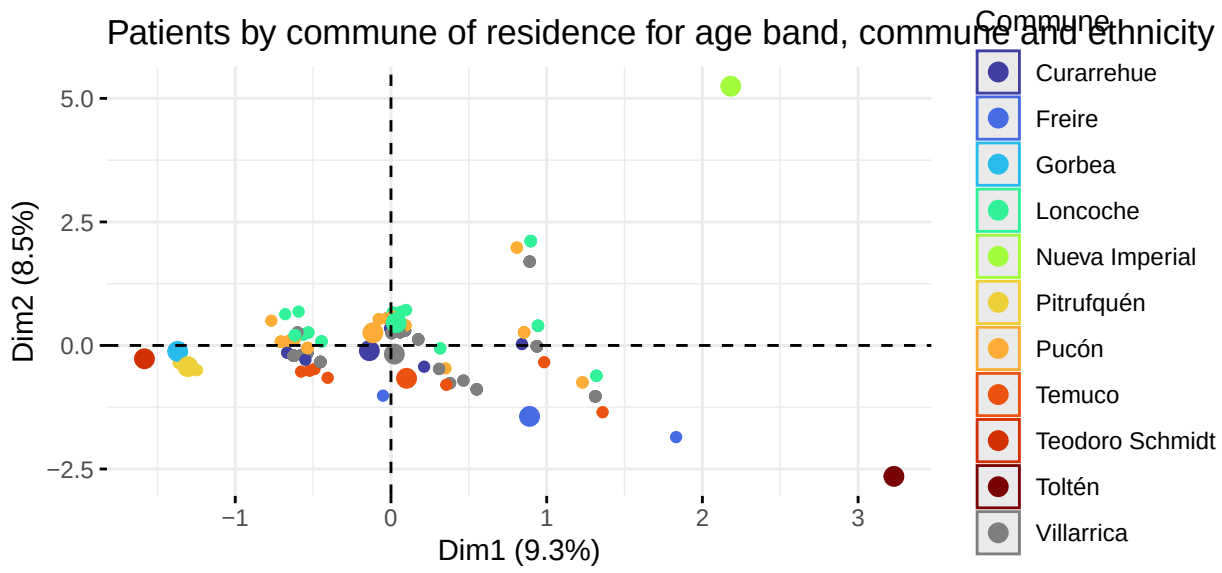

Figure 44: Patients by the first two dimensions of multiple correspondence analysis on autism patients in the small clinical data using age band, commune and ethnicity, coloured by commune of residence.

Table 23: Age- and sex-adjusted updated autism prevalence from data linkage in SSAS by age band with 95% gamma confidence intervals.

Age band	Crude prevalence (95% CI)	Adjusted prevalence (95% CI)
6-8	1.54 (1.40, 1.67)	1.54 (1.41, 1.68)
9-11	1.34 (1.21, 1.46)	1.33 (1.21, 1.46)
12-14	1.08 (0.97, 1.19)	1.08 (0.97, 1.20)
15-18	0.96 (0.86, 1.07)	0.98 (0.87, 1.11)

The crude updated prevalence of autism in SSAS is 1.23% (1.17-1.28%) and the age- and sex-adjusted prevalence of autism is 1.22% (1.16-1.28%). For females, crude updated prevalence is 0.47% (0.42-0.52%) and the adjusted updated prevalence is 0.47% (0.41-0.53%). For males, crude updated prevalence is 1.95% (1.84-2.05%) and the adjusted updated prevalence is 1.95% (1.84-2.06%). This gives an adjusted male to female ratio of 4.15. Updated autism prevalence is highest among individuals aged 6-8 with crude prevalence of 1.54% (1.40-1.67%) and age- and sex- adjusted prevalence of 1.54% (1.41-1.68%) and decreases with age, see Table 23.

Table 24: Adjusted prevalence and adjusted updated prevalence of autism by health service in Chile. Adjusted prevalence is from school data only. Adjusted updated prevalence is from linkage of school data and patient data. Prevalences for Servicio de Salud Araucanía Sur were calculated directly from linkage results. Prevalences for other health services were calcuated by adding the adjusted prevalence delta to each health service's adjusted prevalence from the school data only. Adjusted prevalence has 95% gamma confidence intervals and adjusted updated prevalence has confidence intervals equal in width to the school data adjusted prevalence intervals for each health service, except for SSAS which has the 95% gamma confidence intervals found earlier.

Health service	Adjusted prevalence (95% CI)	Adjusted updated prevalence (Equivalent CI)
Aconcagua	0.43 (0.37, 0.50)	1.28 (1.21, 1.34)
Aisén	0.75 (0.63, 0.90)	1.60 (1.47, 1.73)
Antofagasta	0.83 (0.77, 0.88)	1.67 (1.62, 1.73)
Araucanía Norte	0.30 (0.24, 0.38)	1.15 (1.08, 1.21)
Araucanía Sur	0.37 (0.34, 0.41)	1.22 (1.16, 1.28)
Arauco	0.72 (0.62, 0.82)	1.56 (1.46, 1.66)
Arica	0.61 (0.54, 0.70)	1.46 (1.38, 1.54)
Atacama	0.31 (0.27, 0.37)	1.16 (1.11, 1.21)
Biobío	0.42 (0.37, 0.47)	1.27 (1.22, 1.32)
Chiloé	0.43 (0.36, 0.52)	1.28 (1.20, 1.36)
Concepción	0.77 (0.72, 0.83)	1.62 (1.56, 1.67)
Coquimbo	0.40 (0.36, 0.43)	1.24 (1.21, 1.28)
Iquique	0.43 (0.38, 0.49)	1.28 (1.23, 1.33)
Magallanes	0.83 (0.72, 0.96)	1.68 (1.56, 1.80)
Maule	0.30 (0.28, 0.33)	1.15 (1.12, 1.18)
Metro. Central	0.42 (0.38, 0.46)	1.26 (1.22, 1.30)
Metro. Norte	0.29 (0.26, 0.31)	1.13 (1.11, 1.16)
Metro. Occidente	0.34 (0.32, 0.36)	1.19 (1.16, 1.21)
Metro. Oriente	0.30 (0.27, 0.33)	1.15 (1.12, 1.17)
Metro. Sur	0.40 (0.37, 0.43)	1.25 (1.22, 1.27)
Metro. Sur Oriente	0.36 (0.34, 0.39)	1.21 (1.19, 1.24)
O'Higgins	0.42 (0.39, 0.46)	1.27 (1.24, 1.31)
Osorno	0.43 (0.37, 0.51)	1.28 (1.21, 1.35)
Reloncaví	0.42 (0.37, 0.47)	1.26 (1.22, 1.31)
Talcahuano	0.81 (0.74, 0.90)	1.66 (1.58, 1.74)
Valdivia	0.30 (0.26, 0.35)	1.15 (1.10, 1.19)
Valparaíso	0.68 (0.62, 0.74)	1.52 (1.46, 1.58)
Viña del Mar	0.66 (0.62, 0.70)	1.51 (1.47, 1.55)
Ñuble	1.29 (1.21, 1.37)	2.13 (2.05, 2.21)

This gives an adjusted prevalence delta for SSAS of 0.85%. Table 24 shows the projection of adjusted updated prevalence from the data linkage for SSAS onto the other health services. The patterns of prevalence across health services are retained and Ñuble has the highest adjusted updated prevalence at 2.13% (2.05, 2.21%).

5.7 Bayesian prevalence projection

Bayesian prevalence projections by health service with common global autism prevalence prior is shown in Figure 45. The differences in adjusted sample prevalence across health services are evident in the red bands and the posterior predictive distributions have been pulled towards the common prior. For regions like Ñuble

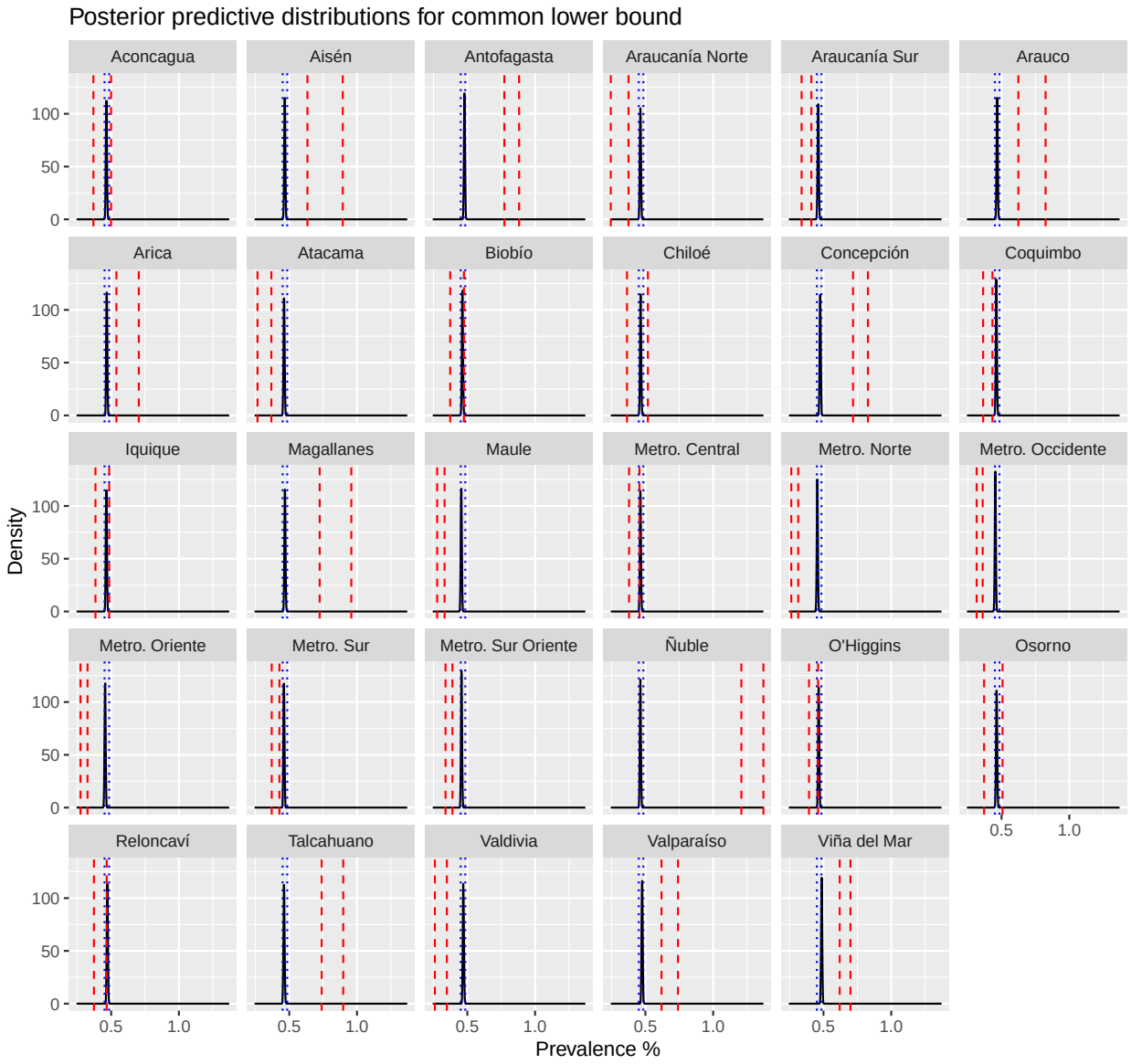

Figure 45: Posterior predictive distributions for autism prevalence using adjusted case counts from the school data with a random effect on student's health service. Beta conjugate prior of age- and sex-adjusted global autism prevalence from school data. Red dashed lines show the adjusted sample prevalence 95% gamma confidence intervals and blue dotted lines show the posterior 95% credible interval.

which have adjusted sample prevalence in the school data higher than the global adjusted prevalence, these posterior densities are not plausible.

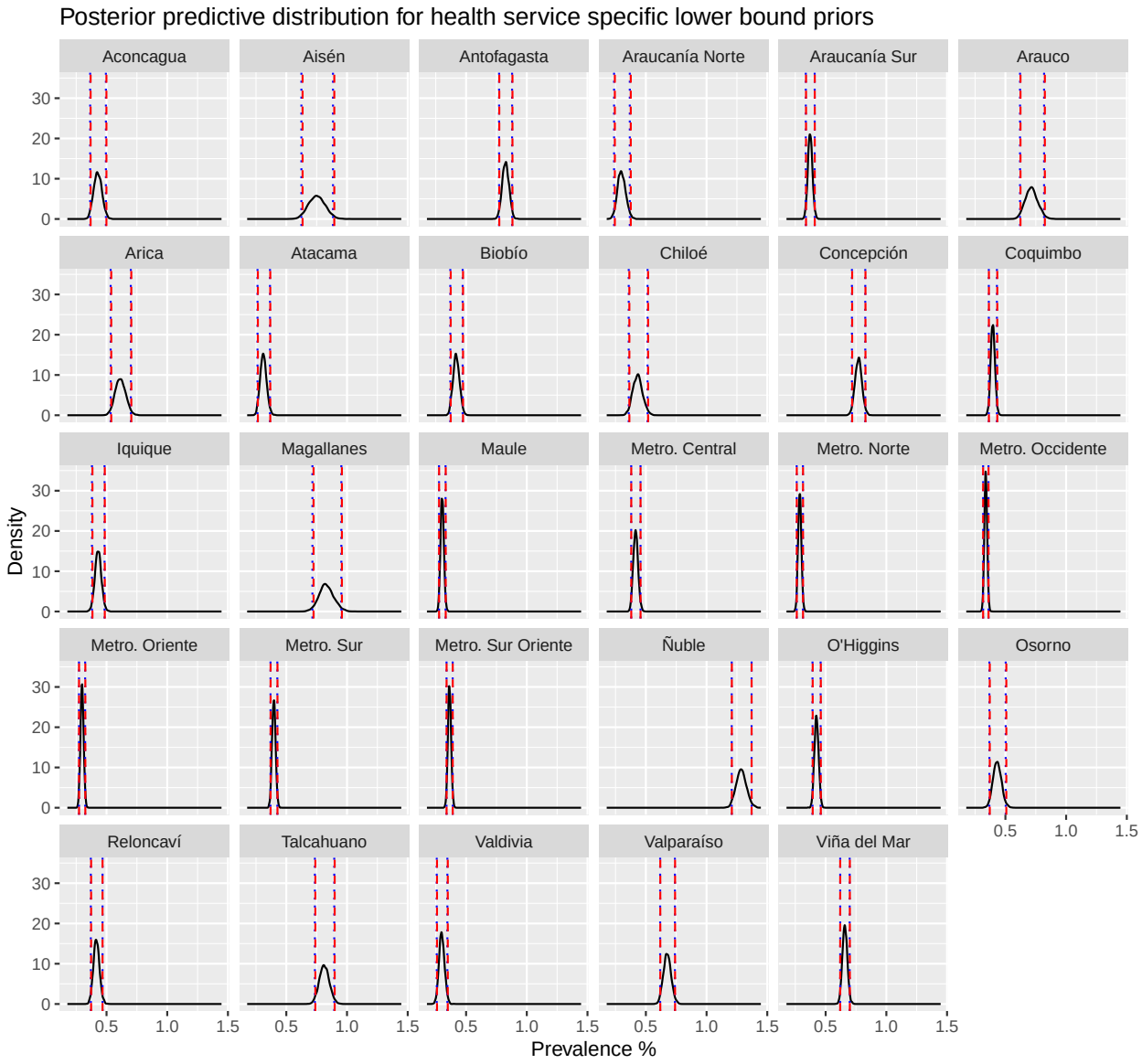

Figure 46: Posterior predictive distributions for autism prevalence using adjusted case counts from the school data with a random effect on student's health service. Beta conjugate prior of health service specific age- and sex-adjusted autism prevalence from school data. Red dashed lines show the adjusted sample prevalence 95% gamma confidence intervals and blue dotted lines show the posterior 95% credible interval.

Figure 46 shows Bayesian prevalence projections by health service with health service specific priors. As expected, the posterior 95% credible intervals are coincident with the adjusted sample prevalence 95% confidence intervals. The posterior prevalence peaks can be considered lower bounds for the true autism prevalence in each health service.

With projected prevalence priors, the Bayesian prevalence projections shown in Figure 47 are pulled upward toward their priors by approximately delta. For regions with high sample prevalence such as Ñuble, this addition of delta results in a posterior prevalence projects up to 1.5 percentage points higher than the global adjusted prevalence for the school data only which may not be plausible.

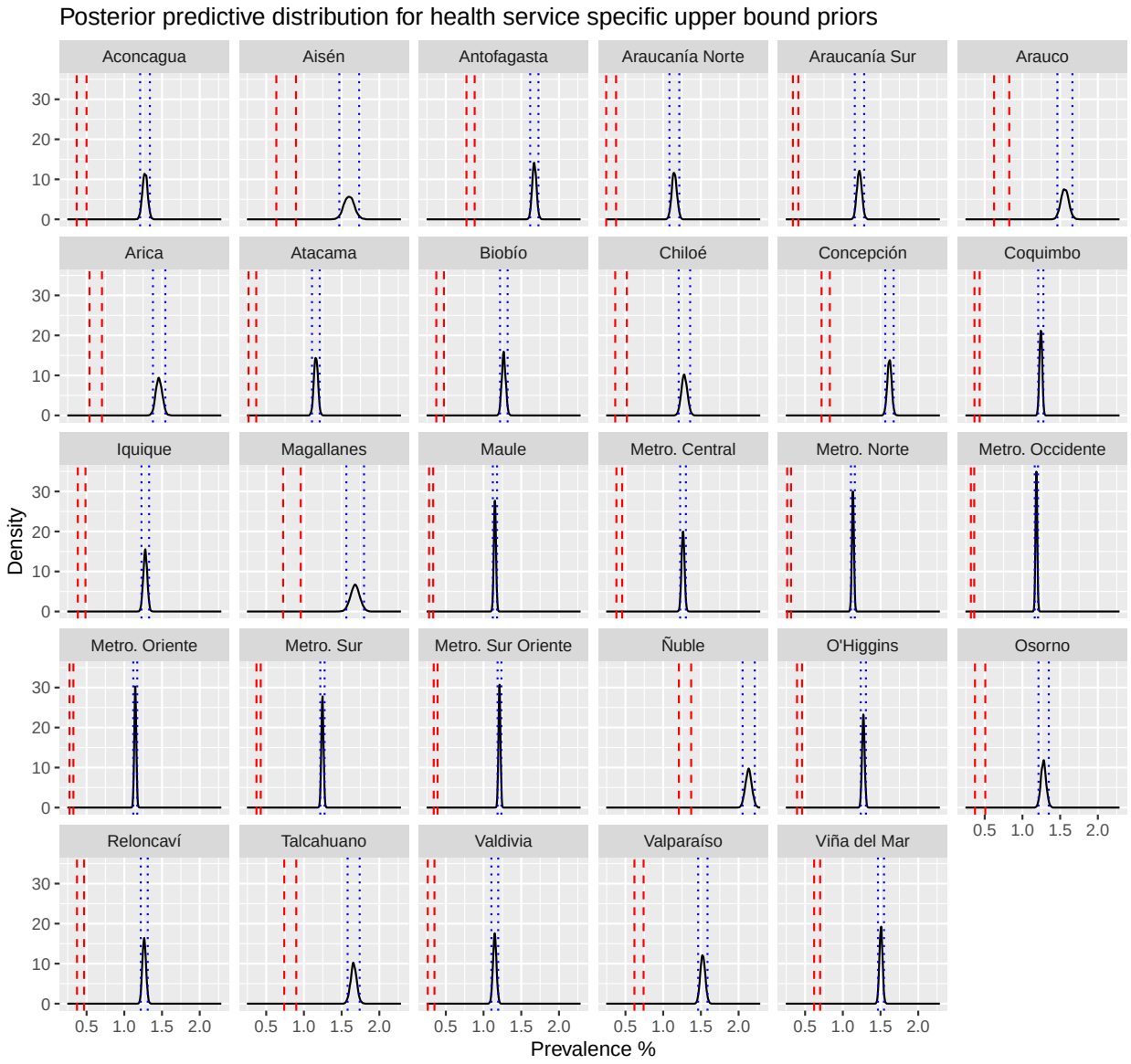

Figure 47: Posterior predictive distributions for autism prevalence using adjusted case counts from the school data with a random effect on student's health service. Beta conjugate prior of health service specific age- and sex-adjusted updated autism prevalence from data linkage. Red dashed lines show the adjusted sample prevalence 95% gamma confidence intervals and blue dotted lines show the posterior 95% credible interval.

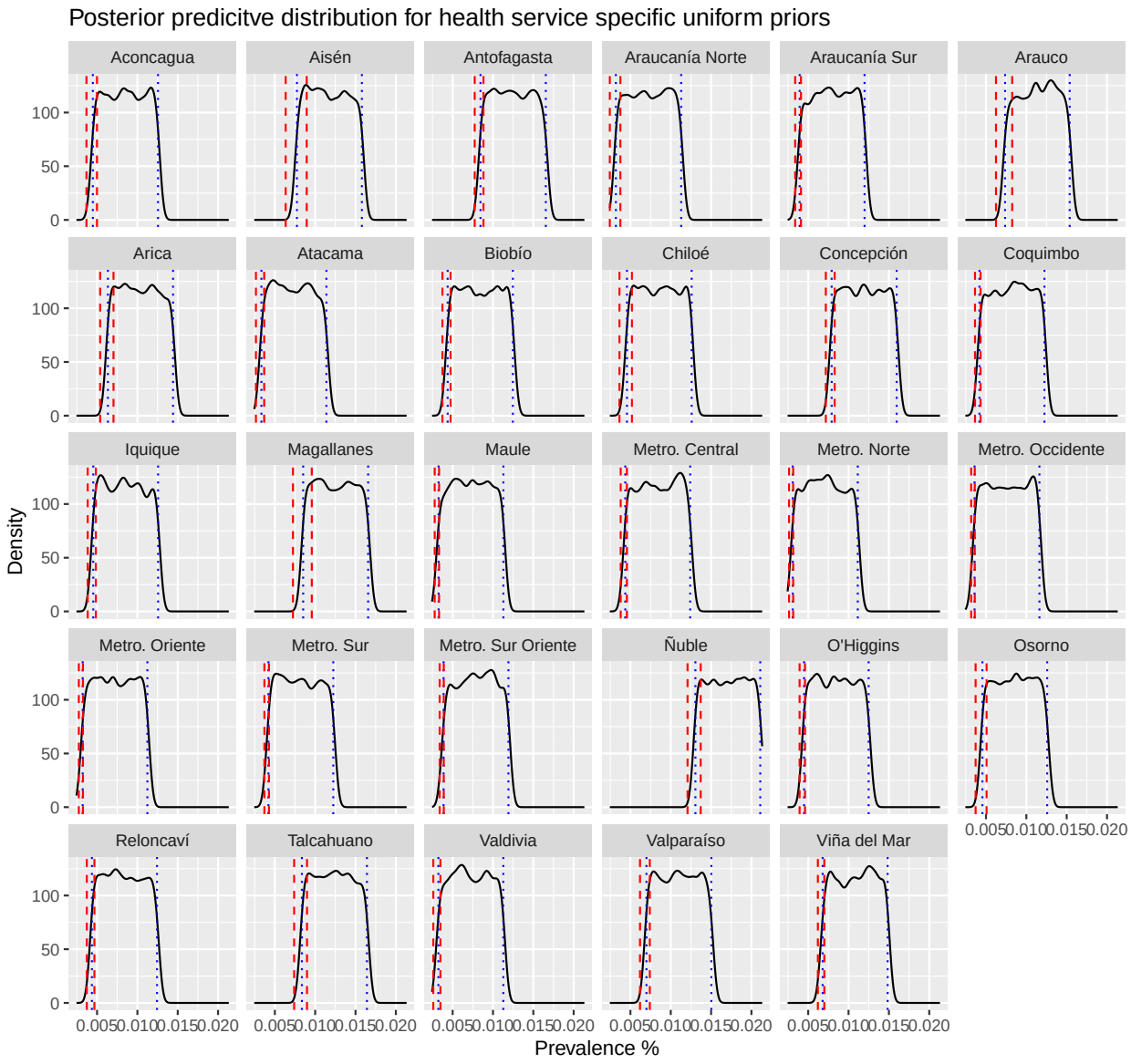

Figure 48: Posterior predictive distributions for autism prevalence using adjusted case counts from the school data with a random effect on student's health service. Uniform prior bounded by health service specific age- and sex-adjusted autism prevalence from school data, and health service specific age- and sex-adjusted updated autism prevalence from data linkage. Red dashed lines show the adjusted sample prevalence 95% gamma confidence intervals and blue dotted lines show the posterior 95% credible interval.

In Figure 48, the posterior distributions are reflective of their uniform priors with posterior credible intervals slightly within the prior bounds (Table 24). These predictive densities show a considerable departure from the adjusted sample prevalence from the school data and provide a window within which the true prevalence of recorded autism in each health service would plausibly fall.

5.8 Projected unmet need in school support

Do unmet need if delta is true across all regions and if araucS adjusted updated prev is true for all regions except Nuble which can keeps its observed prev.

6 Discussion

6.1 Findings

6.1.1 Aim 1: Use school data and frequentist method to find a lower bound on autism and ADHD prevalence

In Chilean 6-18 year old students, the observed crude prevalence of autism is highest in young students and decreases with age, and this pattern is preserved across demographic features. Assuming true autism prevalence increases with age as individuals can be diagnosed well into adulthood, this result shows that it is easier for young children to access SEN funding in Chile. This could be the result of improved awareness of the SEED programme and policy changes in Chile to increase accessibility and participation for students with autism (37), and could also indicate that students with autism need fewer schooling interventions as they mature.

The crude prevalence of ADHD in Chilean students is highest for early teenagers then decreases somewhat and this pattern is also preserved across demographic features. This suggests ADHD is diagnosed later than autism which is somewhat consistent with others' work including Sainsbury et al. who note that autism is typically diagnosed at age 5 and ADHD at age 6 (38). The observed decrease in ADHD prevalence in later teenage years could indicate reduced ability to access schooling interventions with age, or reduced need for them.

After age- and sex-adjustment, the lower bound on autism prevalence of xxx% is certainly lower than in comparable countries ref. The male to female ratio is more skewed towards males ref.

Econ

Prevalence of autism and ADHD is lower among students of the Mapuche and other Indigenous groups than among non-Indigenous students (confidence intervals are non-overlapping except for autism prevalence for students from other Indigenous groups) which agrees with Lindblom's findings that Indigenous children with autism in Canada's British Colombia are underrepresented in school data (39). However it is important to note here that people Indigenous to different regions are not necessarily comparable and very little research has been done on autism and ADHD in Indigenous people in Latin America.

Autism and ADHD prevalence is higher in rural schools that urban school which is somewhat surprising as urban areas generally have more resources to diagnose and provide adjustments for these conditions, but may be the result of successful implementation of policy interventions described by Núñez and Manzano that aimed to improve healthcare for disadvantaged areas and population groups (40).

6.1.2 Aim 2: Use clinical data and machine learning to identify autism diagnosis characteristics

MCA revealed that age, ethnicity and commune of residence, and to some extent disability status, foster care status and health insurance level, are important features for understanding the distribution of autism diagnosis. The clustering of patients based on whether information was available for disability and foster care status shows that missingness in demographic data is important when understanding autism diagnoses. The variance explained by the available features is not large enough to draw strong conclusions about possible

composite components associated with autism diagnosis, but it does indicate that further investigation with larger dataset and more demographic characteristics are likely to lead to interesting results.

6.1.3 Aim 3a: Use machine learning to link school and clinical records

Both manual and probabilistic linkage of SSAS school and patient data of individuals with autism were able to identify a reasonable number of matches. Probabilistic matching was more effective than manual matching here because matching on commune of residence was desirable but not essential, proxy SES features were known to be imprecise so requiring perfect matches would miss many correct matches, and probabilistic matching was able to identify more likely matches.

The proportions of matched and unmatched records in the SSAS school and patient datasets did differ by sex, commune of residence and proxy SES. The differences appear to be driven by differences in the frequency of these features' categories across the two datasets. In particular, the male to female ratio of people with autism in the patient data is 3.99:1 which is very consistent with Yáñez et al's estimate of 4:1 in the Estación Central commune of Servicio de Salud Metropolitano Central and Santiago commune of Servicio de Salud Metropolitano Occidente, both in Santiago (41). In the SSAS school data, the male to female ratio is 7.00:1, and in the school data for SSAS for students with and without autism it is 1.04:1. This suggests females with autism are underrepresented in the school data and may indicate that there are barriers to females accessing SEED for autism.

6.1.4 Aim 3b: Accurately estimate autism prevalence and project prevalence bounds across health services using Bayesian prevalence prediction

The updated estimate for the prevalence of autism in SSAS is 1.22% with a male to female ratio of 4.15. This sex ratio is consistent with the 4:1 ratio found by Yáñez et al (41). The is updated prevalence estimate found here is likely to be an underestimate due to not having private sector diagnoses in the clinical data and it is somewhat smaller than Yáñez et al's estimate of 1.96% (0.81-4.63) in Santiago communes but well within their estimate's confidence interval (41). This investigation therefore provides further evidence that the prevalence of autism and its sex distribution in Chile is comparable to other countries.

Projecting the adjusted prevalence delta across all regions resulted in updated estimates for all health services that are comparable to Yáñez et al's estimate (41). Ñuble's standard adjusted prevalence estimate is considerably higher than for other communes which may indicate that Ñuble uses a different approach for autism SEED funding assessment and this figure may be more indicative of Chile's true autism prevalence. Using the same adjusted prevalence delta for all regions may not have been appropriate given the regions differ considerably in prevalence observed from the school data. Additionally, SSAS is in the bottom third of adjusted prevalence estimates, meaning it is likely to have a larger difference between adjusted prevalence from the school data and true prevalence. If we can assume the adjusted updated prevalence for SSAS is the true autism prevalence of Chile then the delta to reach this will be larger for SSAS than for regions with higher original adjusted prevalence from the school data. Therefore using SSAS's delta in higher prevalence regions will produce overestimates. Ñuble's adjusted updated prevalence is 2.13% which is biologically possible but there is a phenotypic upper bound on true autism prevalence and it is more likely that a larger proportion of people with autism in Ñuble are accessing SEED than in other regions.

The Bayes prevalence projection was able to circumvent issues with the applicability of the adjusted prevalence delta by modelling prevalence within plausible bounds provided by the school and clinical dataset. Using uniform priors specific to the health services that spanned the lower bounds found through frequentist methods to the possibly overestimated adjusted updated prevalences found through linkage, the Bayesian modelling found that the true prevalence of autism in each commune is likely to be considerably higher than the prevalences observed from the school data alone. This analysis has determined a plausible range within which the true prevalence for each region is quantifiably likely to fall.

Unmet need.

6.2 Limitations

This investigation is limited by the quality of available data. Firstly, the school data is limited by having only one special needs code available per student as it is based on their SEED status and not all their extant special needs diagnoses. This limitation prompted the use of clinical data to supplement the available diagnoses, however this data is also somewhat limited as validated diagnoses and demographic information were only available for some patients. Additionally, it only includes patients treated in the public health system and the number of patients with autism that are treated privately is unknown. Autism prevalence estimates using clinical data may therefore be underestimates.

During MCA, the disability and foster care status features were very unbalanced with few patients having a recorded disability and even fewer having experienced foster care. For both features, nearly two thirds of records had no information available and while reasonable imputation was use, any imputation adds bias to the results and here the bias is likely to be considerable as so many values were imputed. This means the MCA results should be taken as indicative and more research done on larger, more complete dataset to confirm them.

Linkage of school and clinical data for SSAS used matching on proxy SES. In the school data, SES was mapped from school fees paid and in the clinical data it was mapped from health insurance contributions. These underlying features may not be comparable and the mapping to SES used here for each was imprecise. To account for this, a high error rate was declared when allocating weights during probabilistic linkage, however some incorrect matches based on SES may have been made or correct matches been missed. Linkage also assumed that individuals did not move into or out of SSAS communes between collection of the school and patient data. This is unlikely to be true and a small number of correct matches are likely to have been missed due to individuals who moved not being included in one of the SSAS school data or the patient data being linked. Missed correct matches would cause the updated prevalence estimate for SSAS to be slightly too high and therefore projected updated estimates to also be slightly too high.

The Bayesian modelling did use information from the sample data, the full school dataset, twice during modelling, to varying extents for each prior. For all four analyses, random effects models were fit to health services. For the second prior, which used health service specific autism prevalences, this meant the sample data was directly used twice which generally would not be informative but here was valuable to evidence the lower bound on the true prevalences. For the first prior, the sample data was used indirectly to calculate the common global prevalence and was valuable to demonstrate that the prevalences observed from the school data can only be underestimates of the true prevalence. For the third prior, using the updated prevalences found using data linkage, the sample data was used indirectly to give the prevalence values to which delta was added, and the fourth prior combined the second and third. Although it is usually not appropriate to use sample data twice in Bayesian modelling, it was considered acceptable here as the intention was to conduct exploratory analysis to project the bounds on prevalence rather than predict the true autism prevalence across the other health services.

6.3 Extensions

While this investigation has progressed the understanding of autism and ADHD prevalence in Chile, it has also uncovered new avenues for exploration. If clinical data on ADHD diagnoses were available for a region of Chile, the MCA analysis, data linkage and Bayesian projection conducted here for autism could be extended to ADHD. If clinical data on autism for other health service regions were available, the delta calculation could be validated and thus more accurate autism prevalence estimates could be found.

This investigation has considered autism and ADHD in isolation from each other. However, Mannion and Leader found that 14-78% of children with autism have a co-diagnosis of ADHD (42) and it would therefore be valuable to consider these conditions together. MCA of autism patient could be used to assess the contribution of ADHD and other co-diagnoses such as depression and anxiety, physical disability and intellectual disability.

This investigation used MCA to assess the contribution of feature categories among patients with autism in SSAS which required casting the age feature to categorical age bands. However age is a continuous variable and the information contained in it would be better captured using a continuous variable analysis technique

such as principle component analysis. The commune feature would need to be excluded because it cannot be meaningfully encoded as a continuous or ordered categorical variable. The remaining features could be one-hot encoded to pseudo-continuous variables before performing PCA. This would allow the contribution of patient age to be better characterised but one-hot encoding would somewhat reduce the power of inferences about the other features used. Cluster-based analysis of a larger dataset containing patients with and without autism would also be valuable and would allow further identification of the characteristics associated with autism diagnosis.

During data linkage, this investigation assumed that unmatched patients in the clinical data existed in the school data but did not have an autism diagnosis in the school data. It would be valuable to test this assumption by attempting to link the patient data for SSAS to all school records for SSAS, including those students that do not have an autism diagnosis. This would allow the number of patients that do not exist in the school data to be estimated which would provide a more accurate sample size for use as the denominator in prevalence calculation.

This investigation assumed the adjusted prevalence delta found for SSAS was directly applicable to other regions which may not be the case. Further analysis of the Bayesian prevalence projection could use health service specific deltas that are inversely proportional to the sample prevalence observed for each health service. This would decrease the delta for services with high observed prevalence and thus lessen the increase in the updated prevalence estimate, and would increase the updated prevalence estimate for health services with low observed prevalence. This would provide more biologically plausible prevalence estimates, however subsequent Bayesian analysis would be using information from the school data twice – as the sample data and to inform the health service specific prior through use of observed prevalence to scale delta.

Future work could replicate the prevalence estimation conducted here using the UK Department for Education's National Pupil Database which is similar to the school data used here. This would provide an interesting comparison of autism and ADHD prevalence in two large countries with different population structures and ethnic profiles.

7 Conclusions

This investigation has furthered the understanding of autism and ADHD prevalence in Chile. The patterns of autism and ADHD prevalence in Chile are consistent with other countries. A lower bound on prevalence of autism in Chile is 0.46% (0.13% in females and 0.79% in males) and a lower bound on the prevalence of ADHD is 1.5% (1.01% in females and 1.97% in males). In Servicio de Salud Araucanía Sur, age, ethnicity and commune of residence were found to be important components for explaining the variance in data on patients with autism. School data and patient data on individuals with autism in SSAS were successfully linked and used to calculate that a more accurate estimate of the prevalence of autism in this region is 1.23% with male to female ratio of 4.00 (0.47% for females and 1.95% for males). This corresponds to an unmet need of 1132 additional students in SSAS with autism that do not receive SEED funding for autism. Bayesian prevalence projection extended prevalence projections across health services and demonstrated that the prevalence of recorded autism cases in Chile could up to 2.13%.

8 Supplementary materials

Table 25: Features available in the school data

Feature type	Features
Student	Sex
	Date of birth
	Ethnicity, nationality and immigration status
	SEED special needs status
	Special need eligible for SEED, e.g. autism, ADHD, blindness
	Residential commune and region
	Monthly school fees paid
	Academic grade for 2021 school year
School	Name
	Region and commune
	Rurality

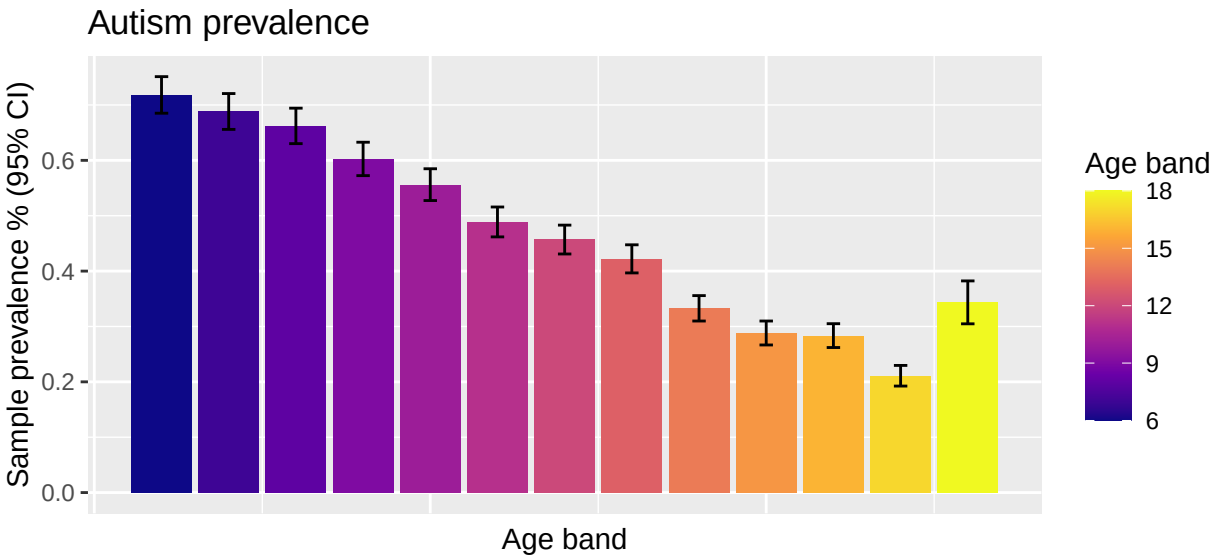

Figure 49: Sample prevalence of autism in school data by age band. Bars show 95% normal confidence intervals.

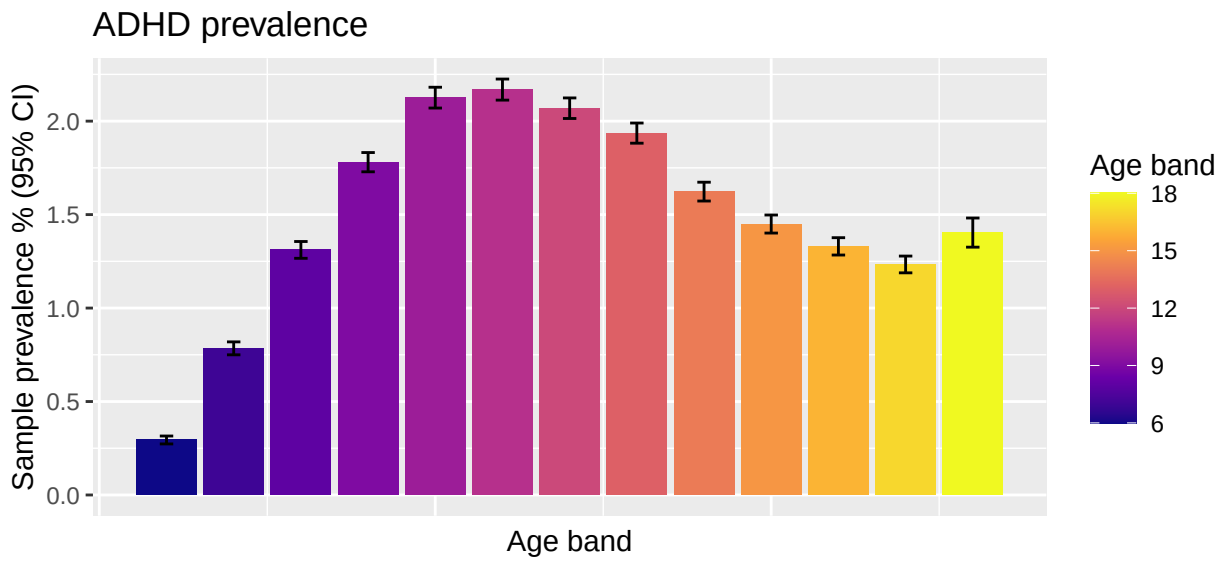

Figure 50: Sample prevalence of ADHD in school data by age band. Bars show 95% normal confidence intervals.

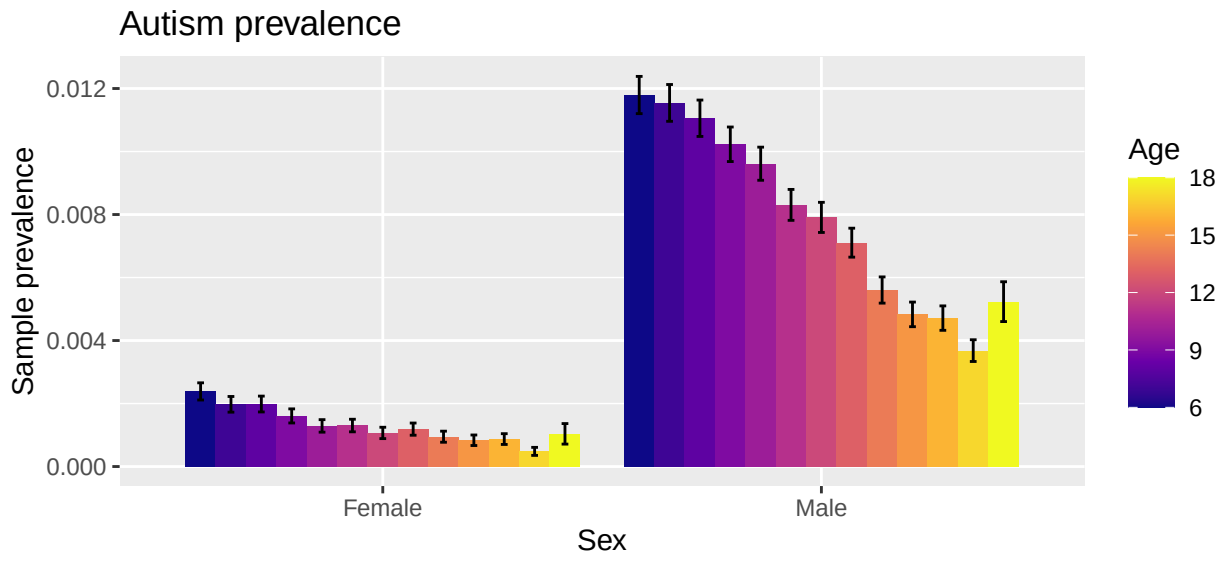

Figure 51: Sample prevalence of autism in school data by age and sex. Bars show 95% normal confidence intervals.

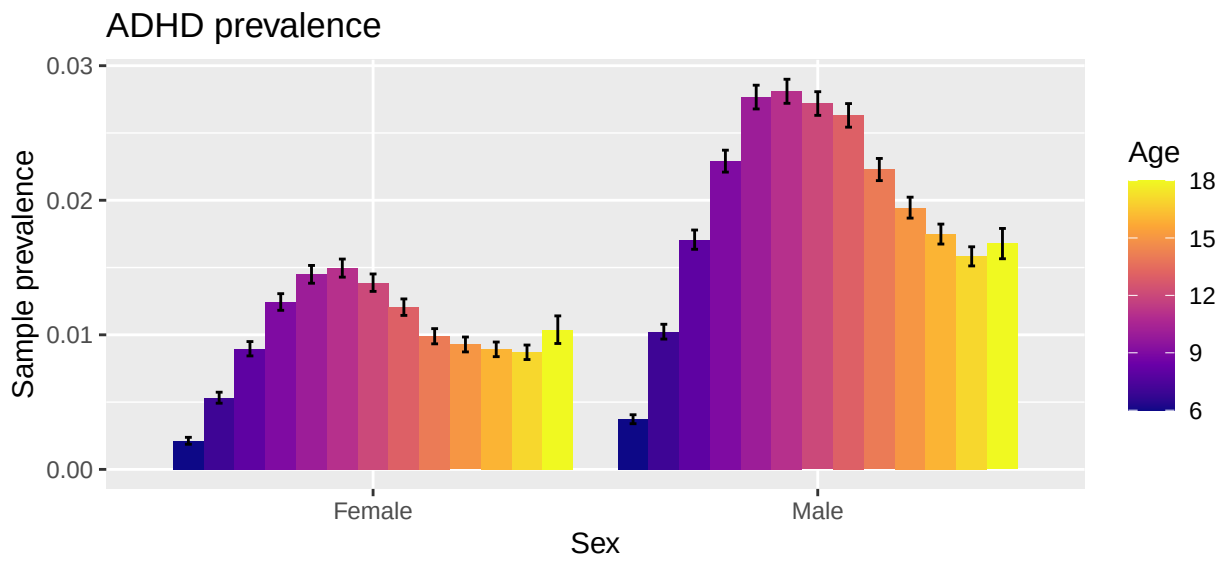

Figure 52: Sample prevalence of ADHD in school data by age and sex. Bars show 95% normal confidence intervals.

9 References

TODO

Figure 53: Sample prevalence of autism in school data by health service, age and sex. Bars show 95% normal confidence intervals.

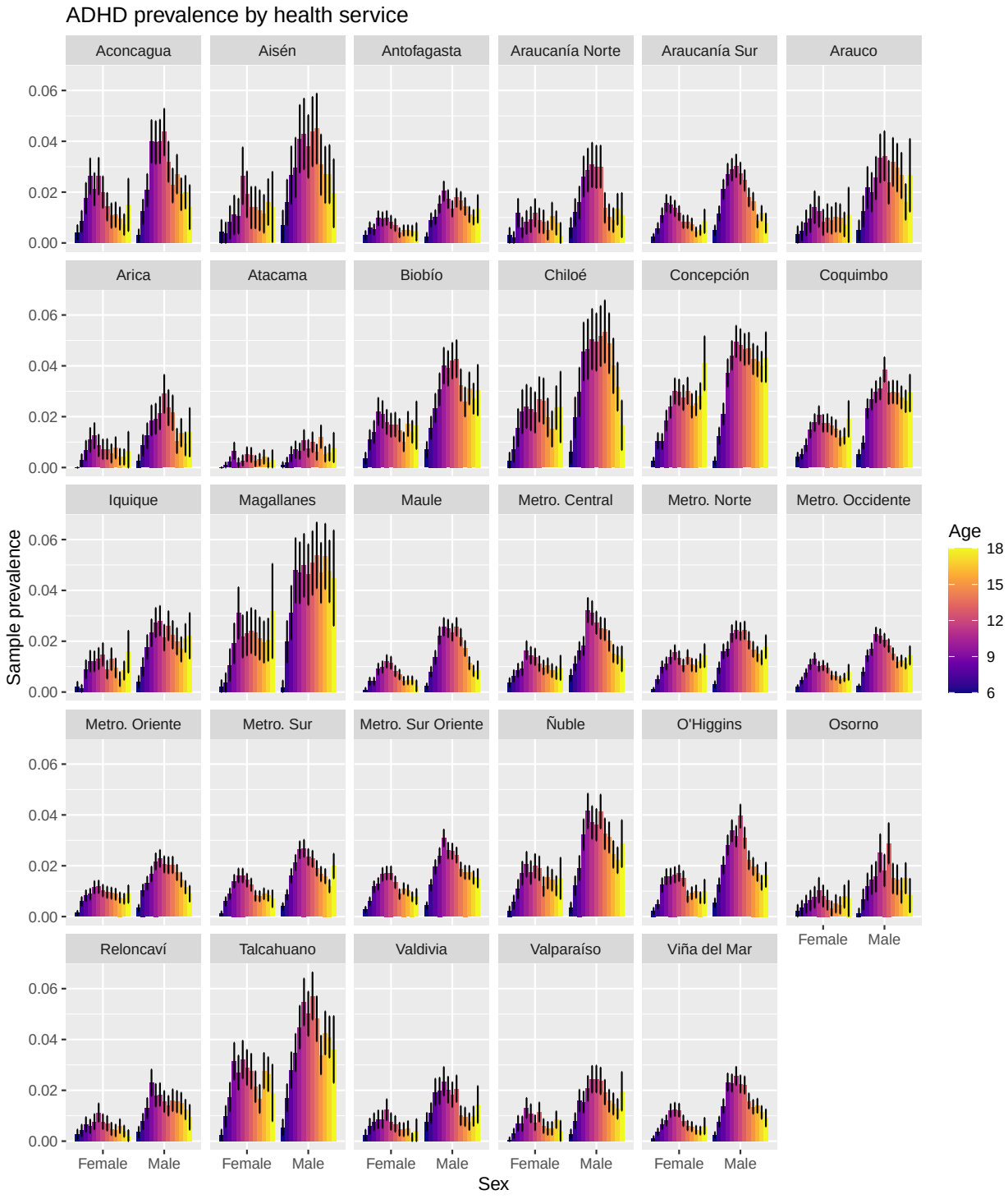

Figure 54: Sample prevalence of ADHD in school data by health service, age and sex. Bars show 95% normal confidence intervals.

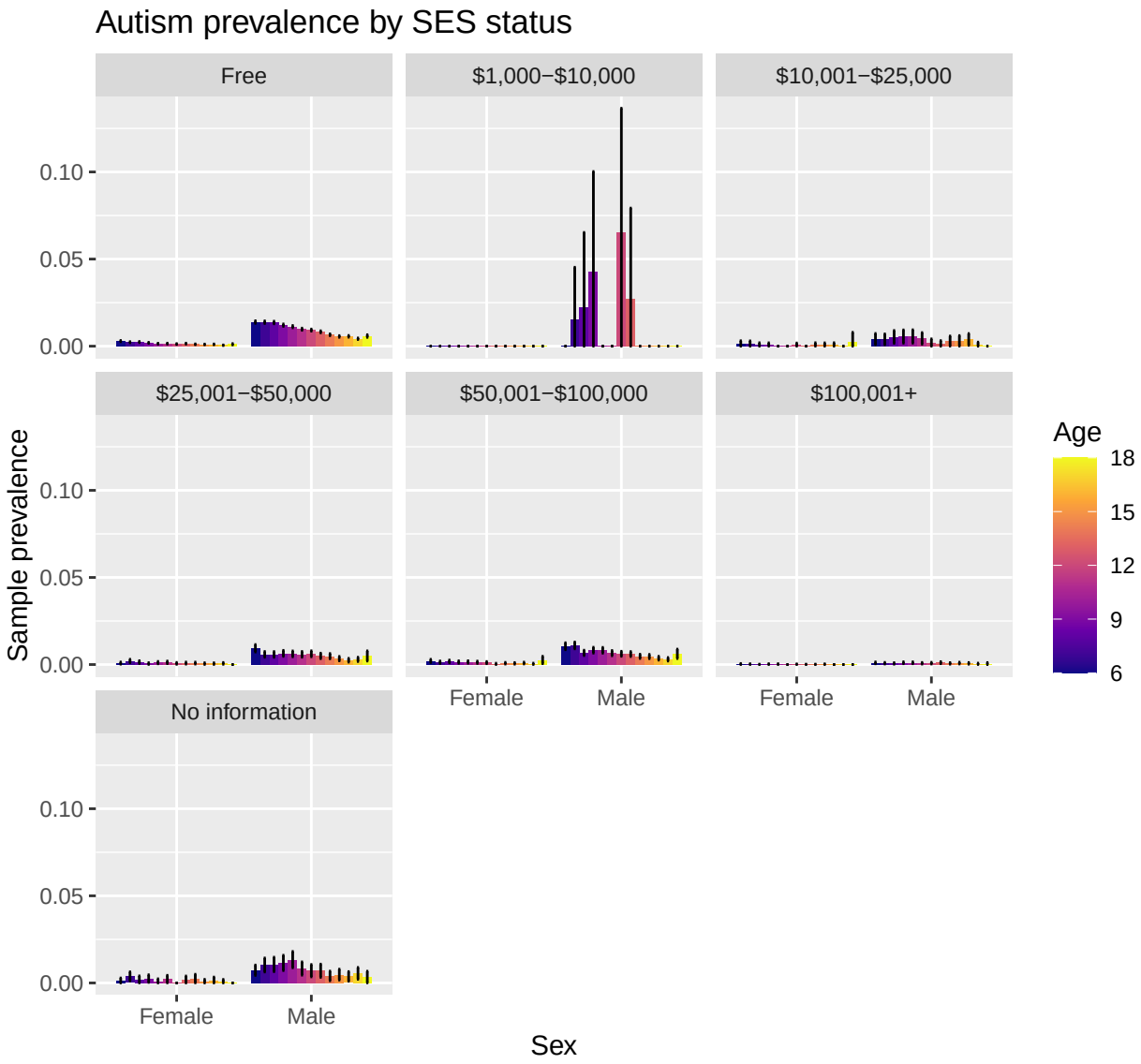

Figure 55: Sample prevalence of autism in school data by socio-economic (SES) status of student's family, age and sex. Bars show 95% normal confidence intervals.

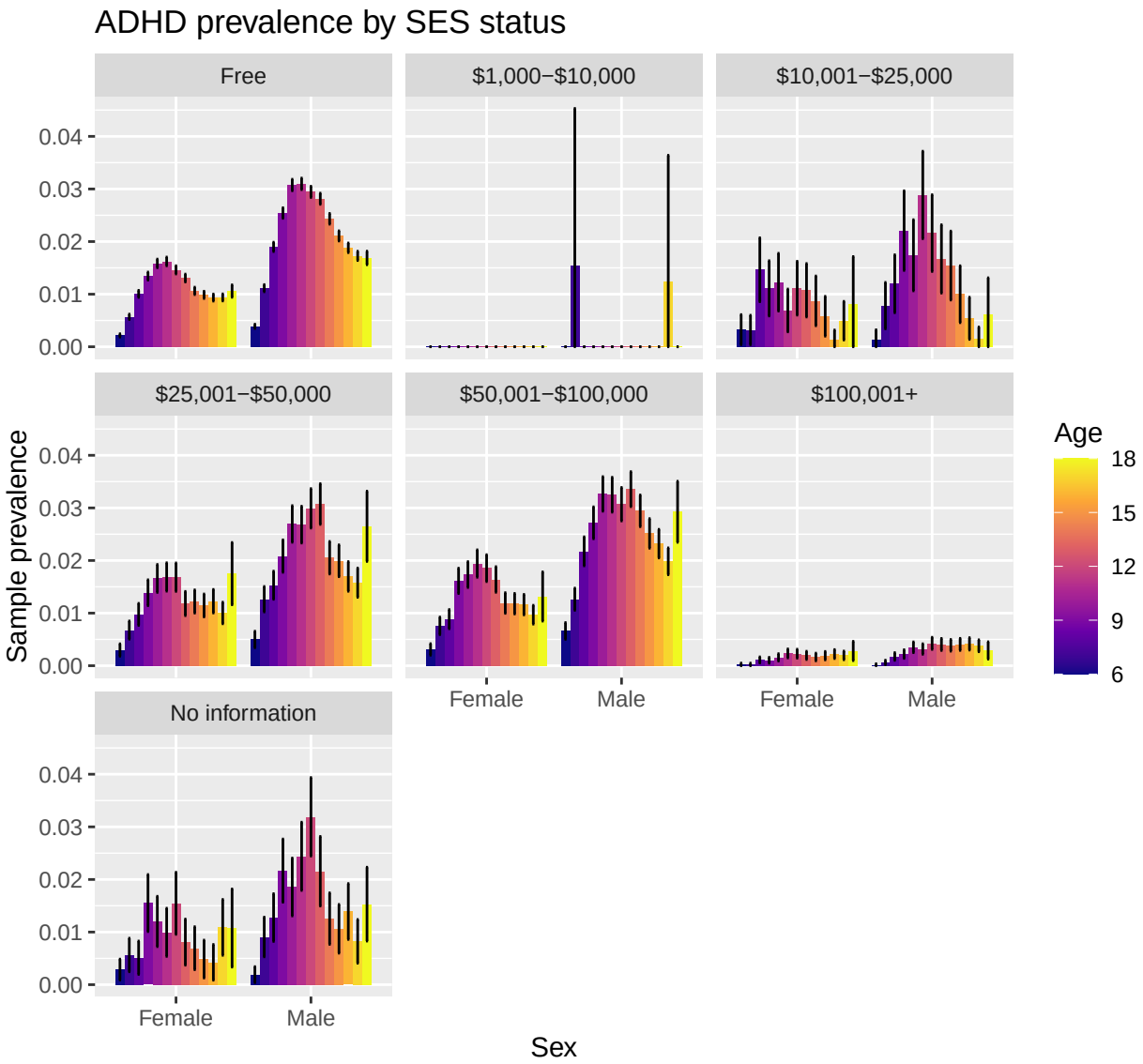

Figure 56: Sample prevalence of ADHD in school data by socio-economic (SES) status of student's family, age and sex. Bars show 95% normal confidence intervals.

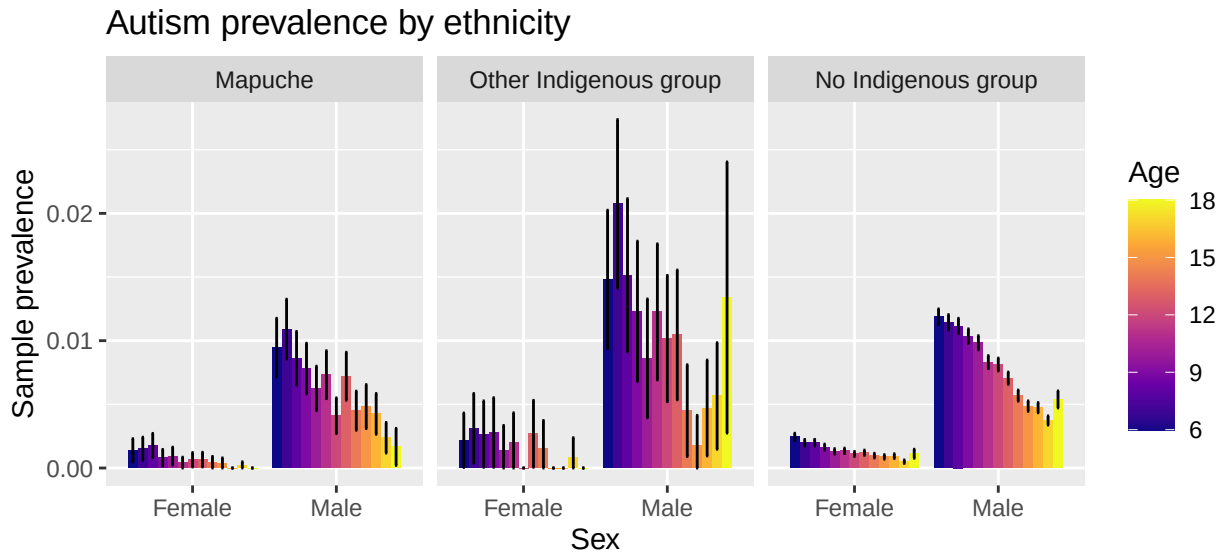

Figure 57: Sample prevalence of autism in school data by ethnicity, age and sex. Bars show 95% normal confidence intervals.

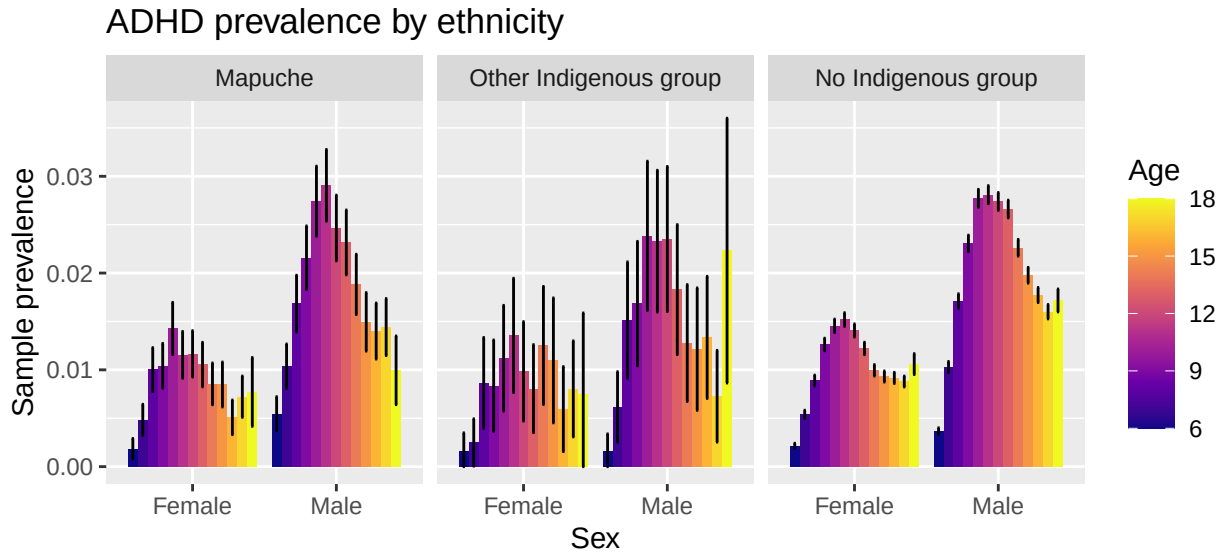

Figure 58: Sample prevalence of ADHD in school data by ethnicity, age and sex. Bars show 95% normal confidence intervals.

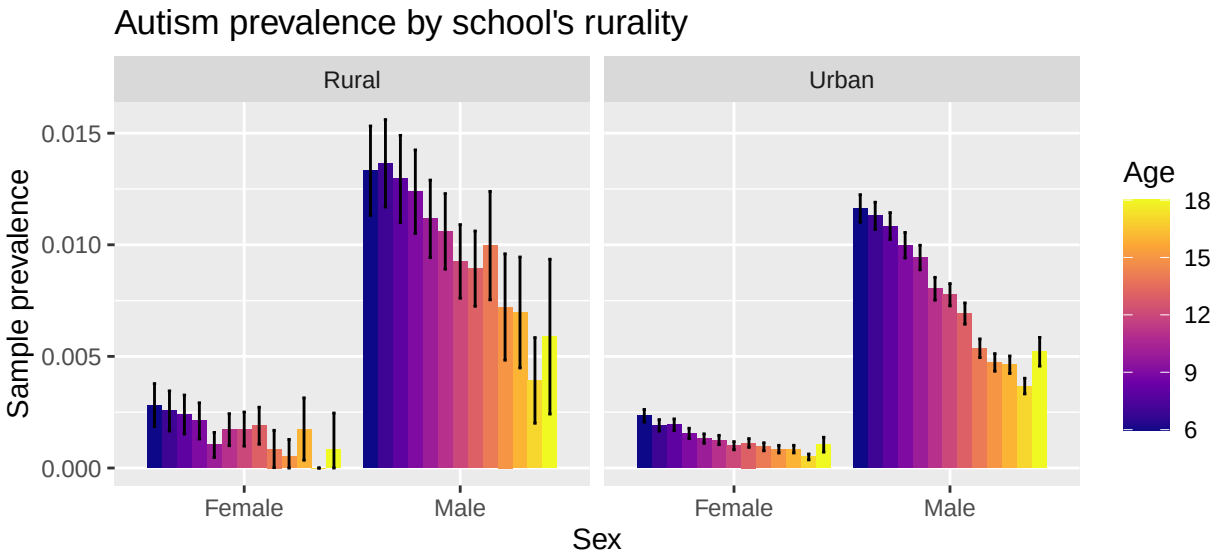

Figure 59: Sample prevalence of autism in school data by school's rurality, age and sex. Bars show 95% normal confidence intervals.

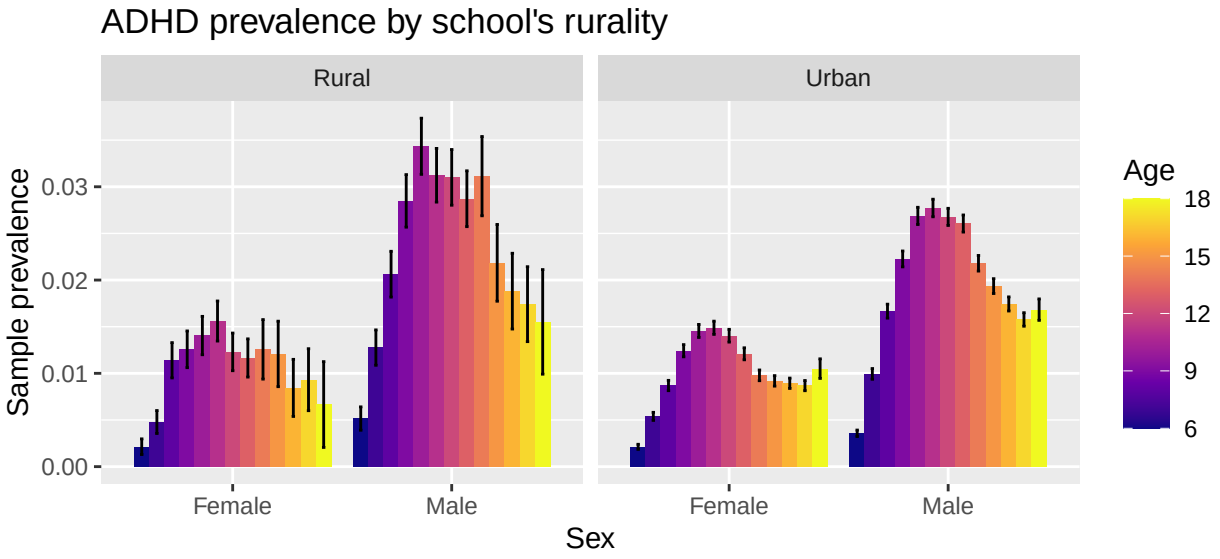

Figure 60: Sample prevalence of ADHD in school data by school's rurality, age and sex. Bars show 95% normal confidence intervals.

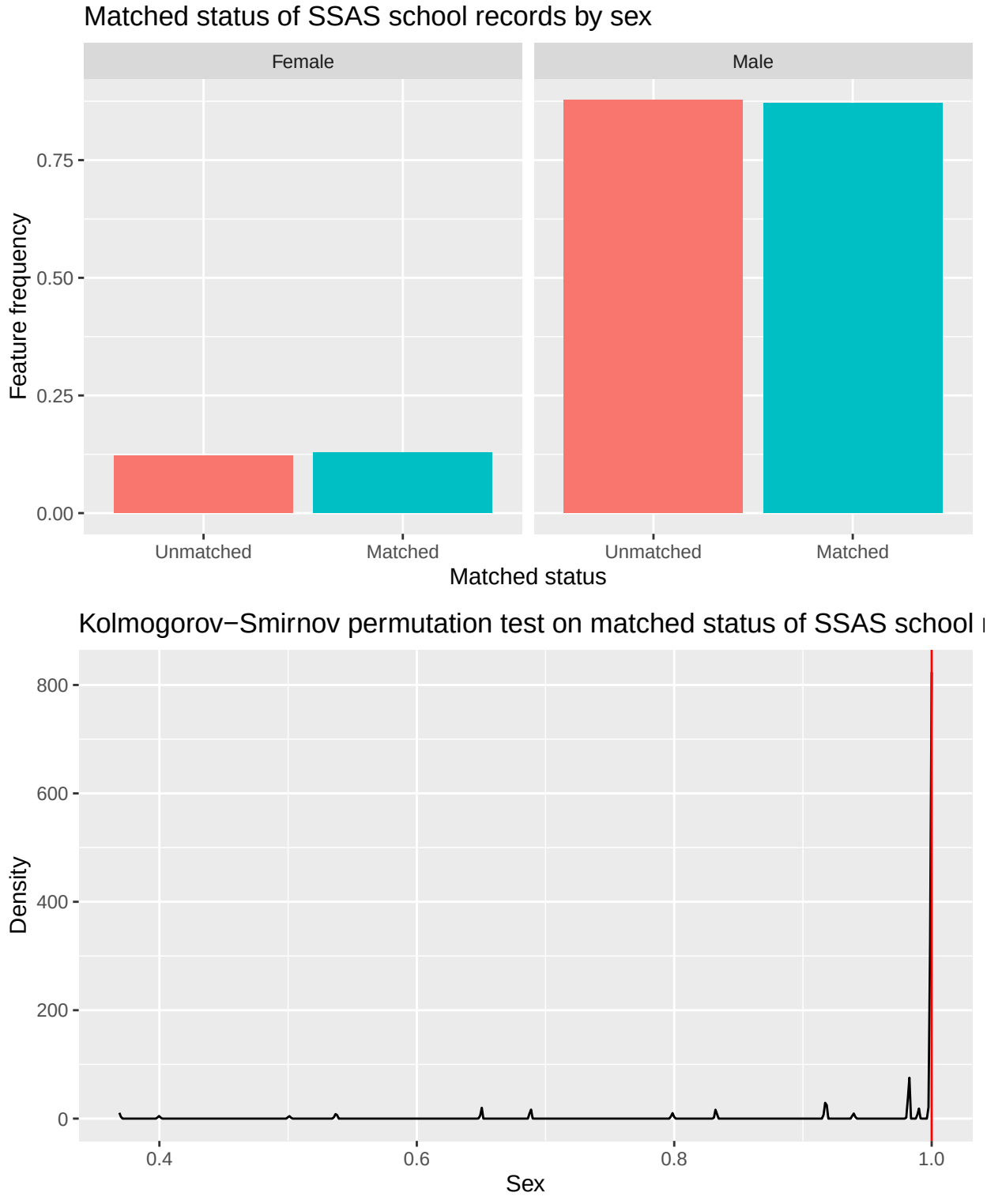

Figure 61: a) Difference in frequency of sexes among school records that matched to patient records and school records that did not match. b) Density of Kolmogorov-Smirnov p-values for 2000 permutations of matched status for school records by sex with observed p-value shown in red.

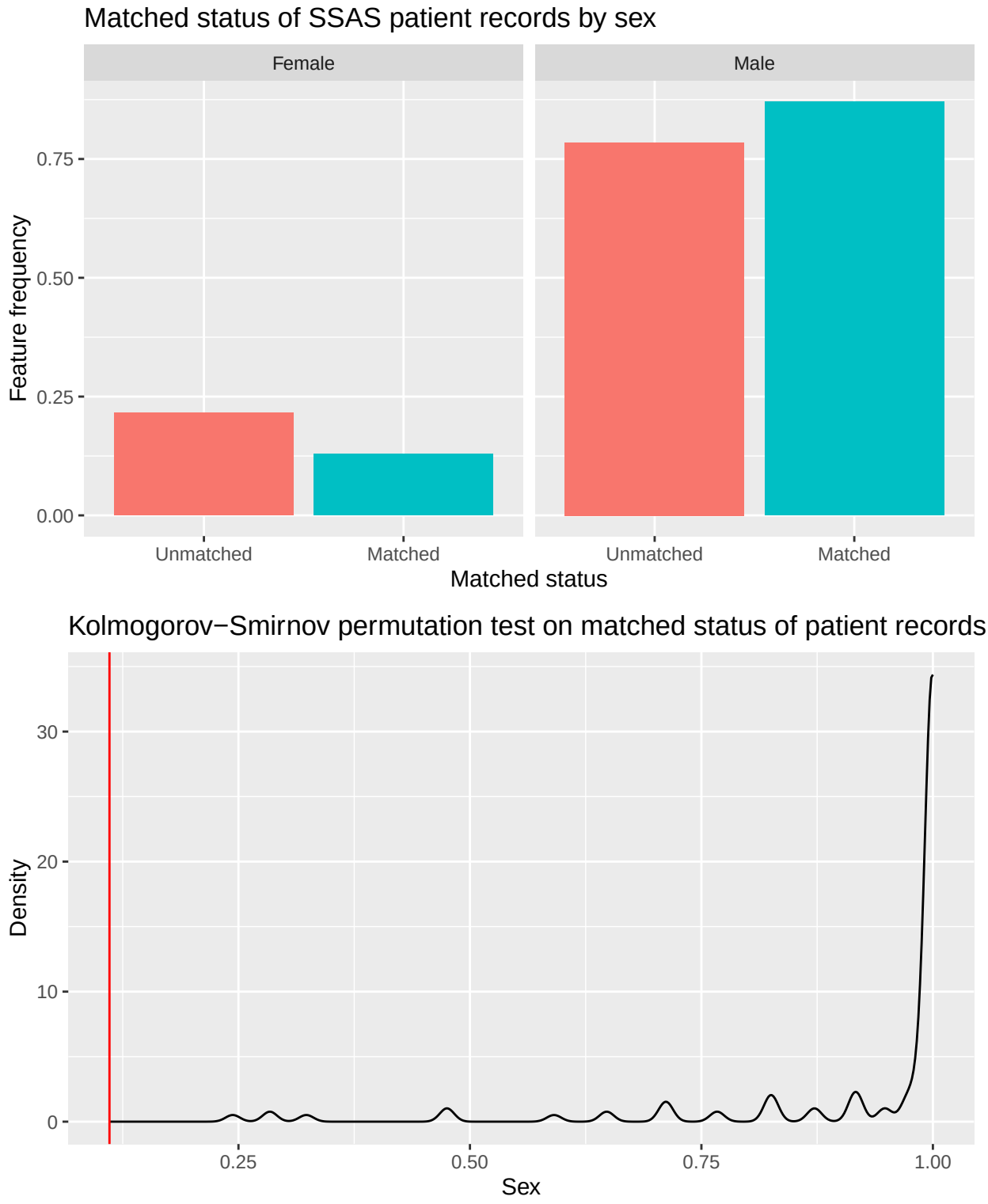

Figure 62: a) Difference in frequency of sexes among patient records that matched to school records and patient records that did not match. b) Density of Kolmogorov-Smirnov p-values for 2000 permutations of matched status for patient records by sex with observed p-value shown in red.

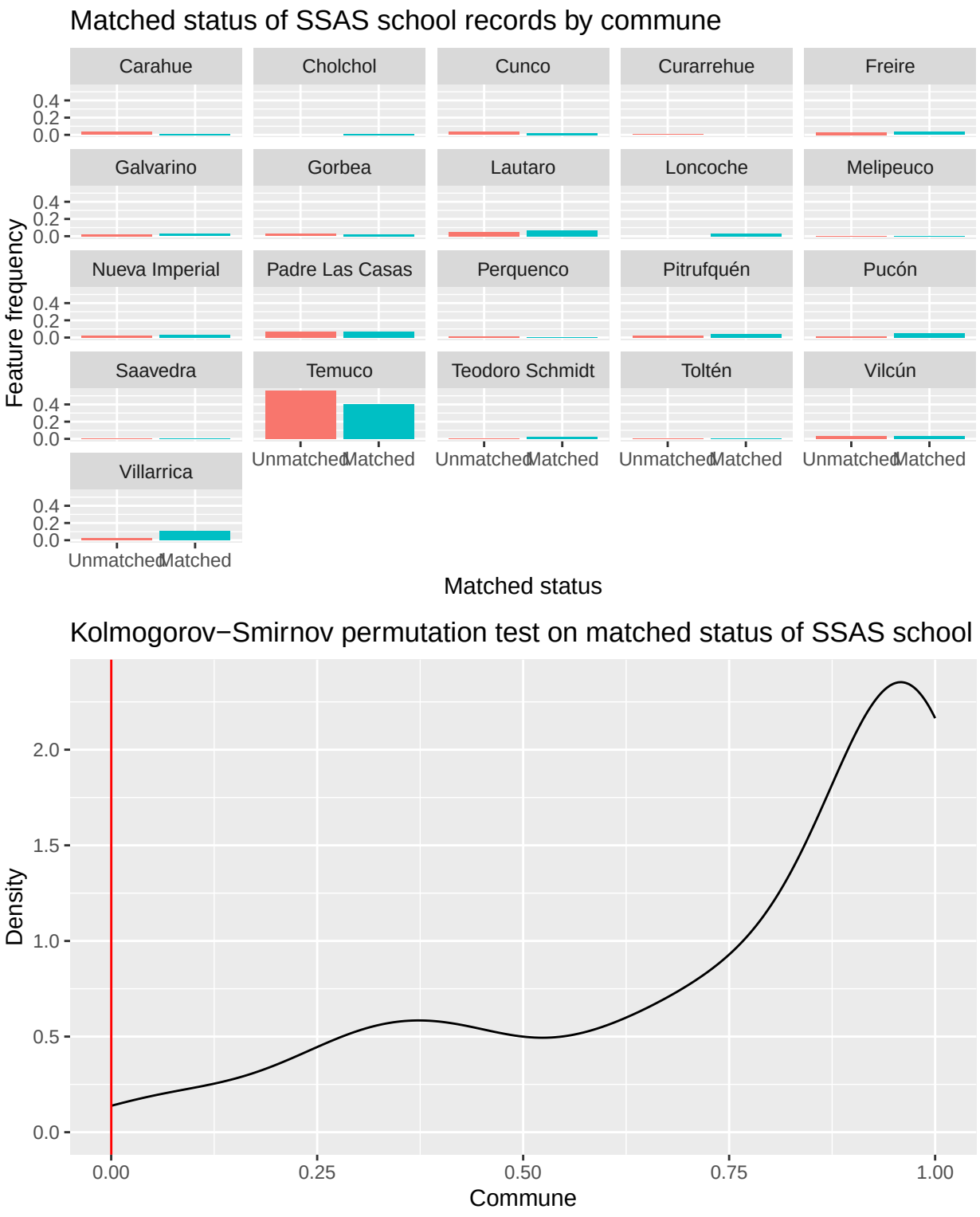

Figure 63: a) Difference in frequency of communes of residence among school records that matched to patient records and school records that did not match. b) Density of Kolmogorov-Smirnov p-values for 2000 permutations of matched status for school records by commune of residence with observed p-value shown in red.

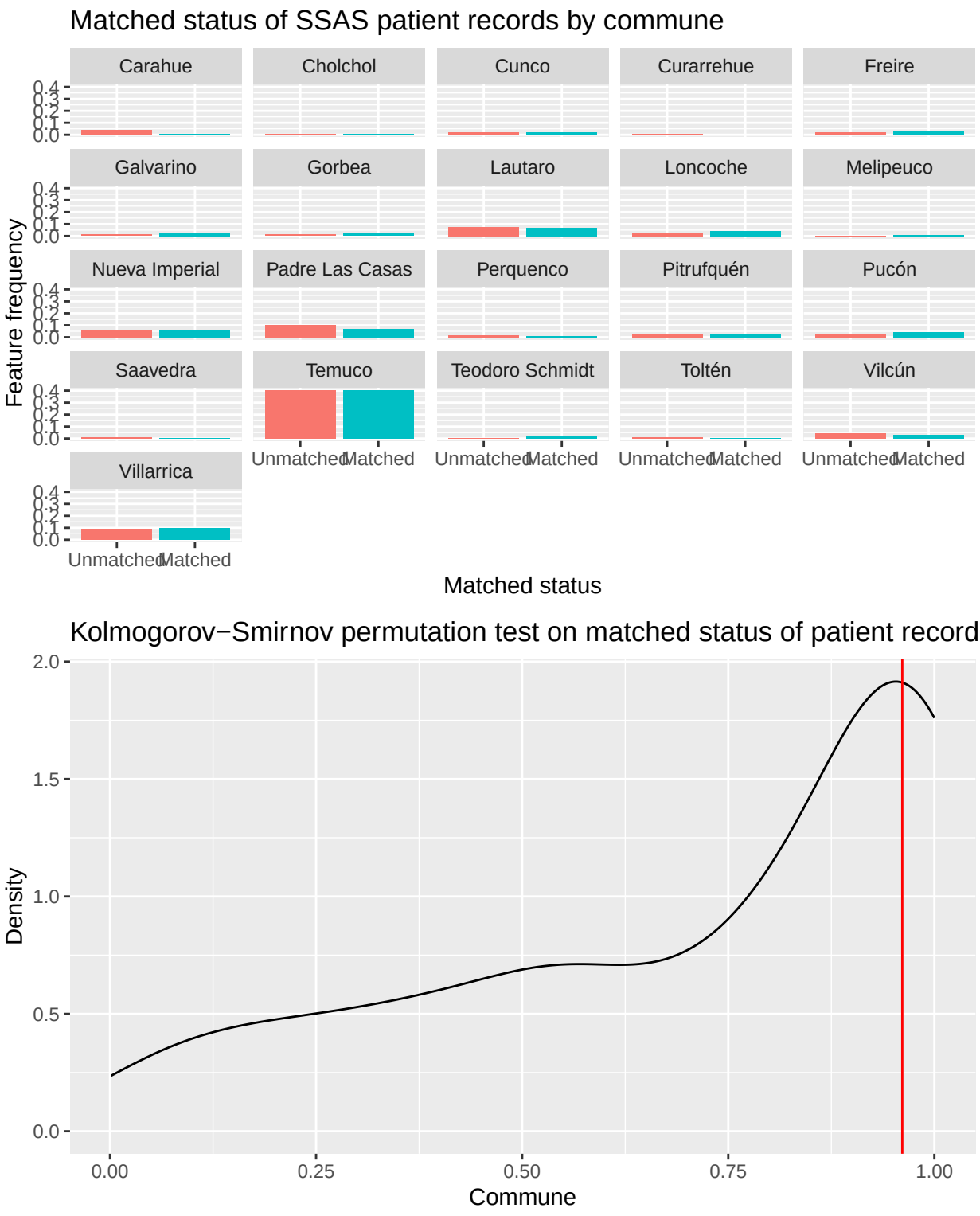

Figure 64: a) Difference in frequency of commune of residence among patient records that matched to school records and patient records that did not match. b) Density of Kolmogorov-Smirnov p-values for 2000 permutations of matched status for patient records by commune of residence with observed p-value shown in red.

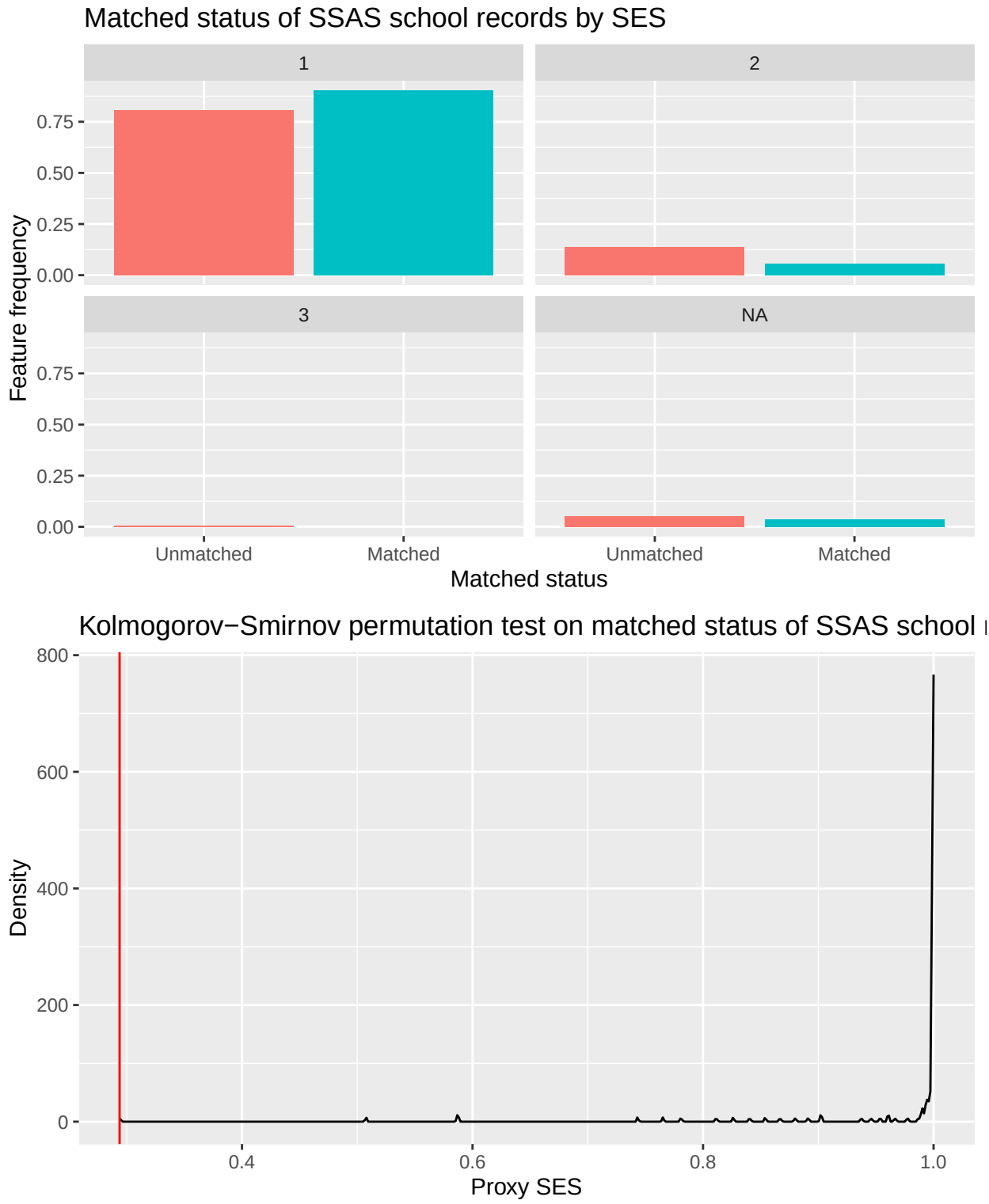

Figure 65: a) Difference in frequency of proxy SES among school records that matched to patient records and school records that did not match. b) Density of Kolmogorov-Smirnov p-values for 2000 permutations of matched status for school records by proxy SES with observed p-value shown in red.

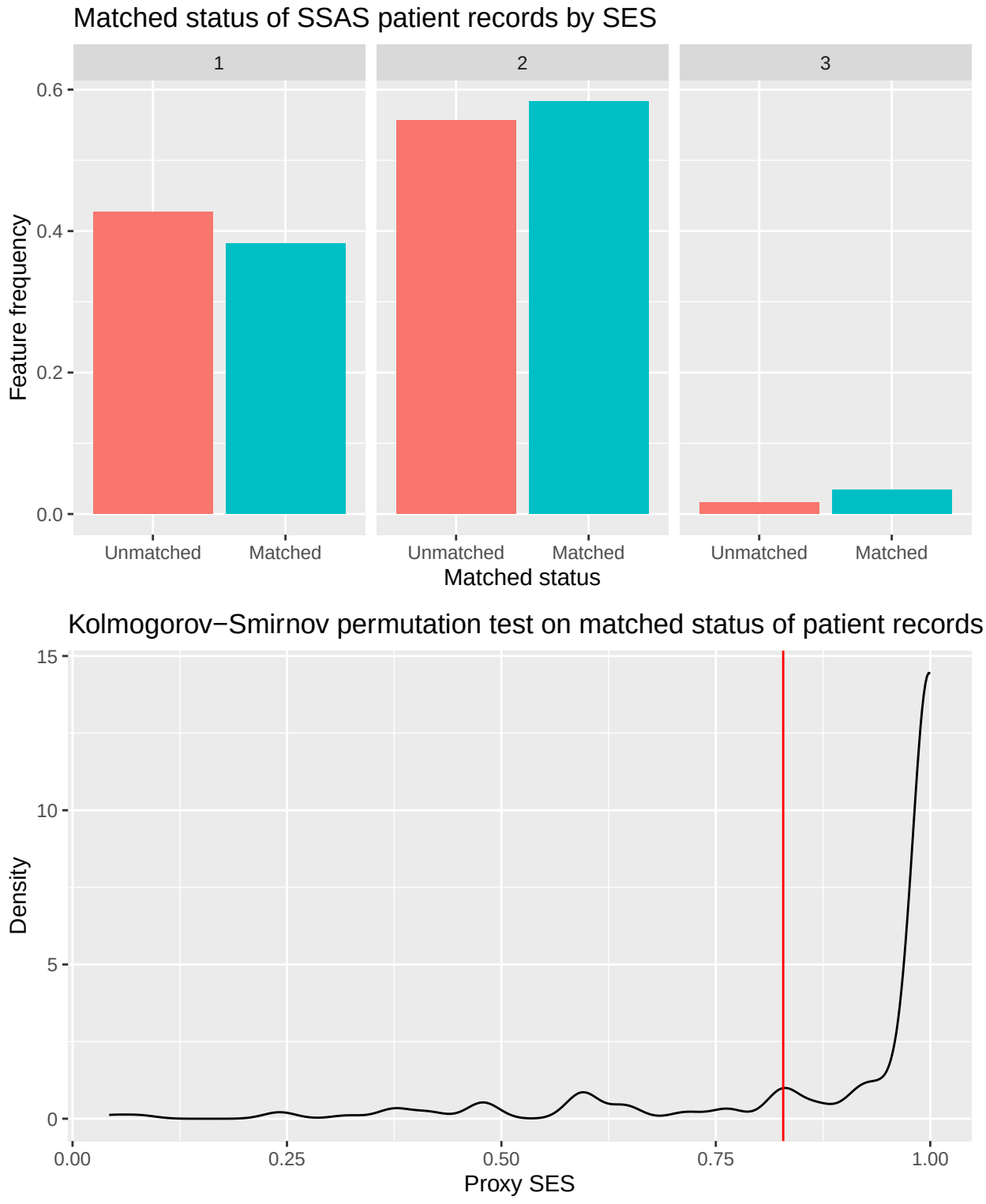

Figure 66: a) Difference in frequency of proxy SES among patient records that matched to school records and patient records that did not match. b) Density of Kolmogorov-Smirnov p-values for 2000 permutations of matched status for patient records by proxy SES with observed p-value shown in red.

10 Appendix A | R code

11 Appendix B | Research Protocol

See overleaf.