

Investigating the autism and ADHD prevalence in Chile through
Bayesian prevalence analysis and clinical record data linkage

Student 5526

2023-07-06

This dissertation is submitted for the degree of Master of Philosophy. The dissertation does not exceed the word limit for the respective Degree Committee. Word count: xxx TODO

1 Abstract

TODO

2 Declaration

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.

USN: xxxxx July, 2022

TODO - is this section needed?

3 Introduction

TODO - copy in intro text

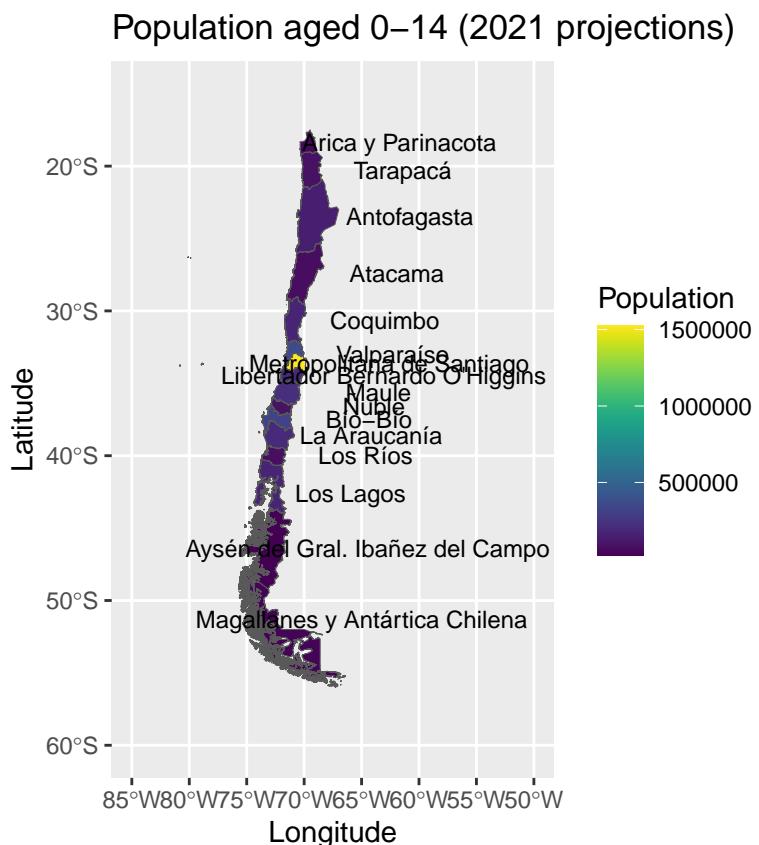


Figure 1: Population of 0-14 year olds in Chile in 2021 by region, from 2017 census projections.

4 Methods

TODO - copy in methods

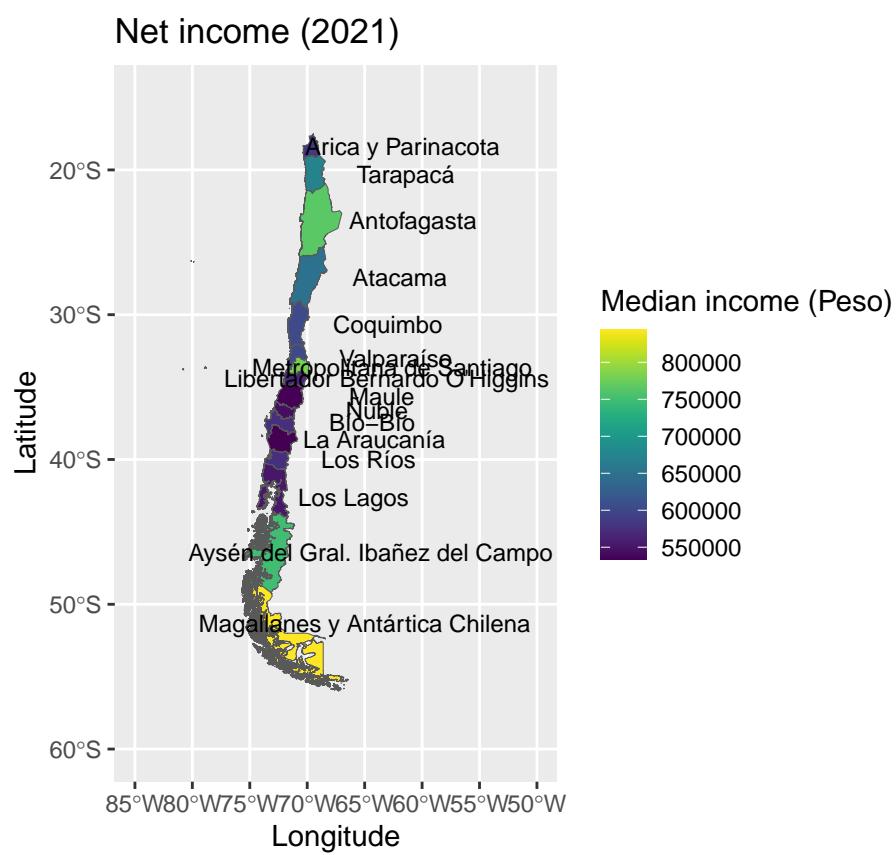


Figure 2: Net income from main job in Chile in 2021 by region, from the INE's Supplementary Income Survey.

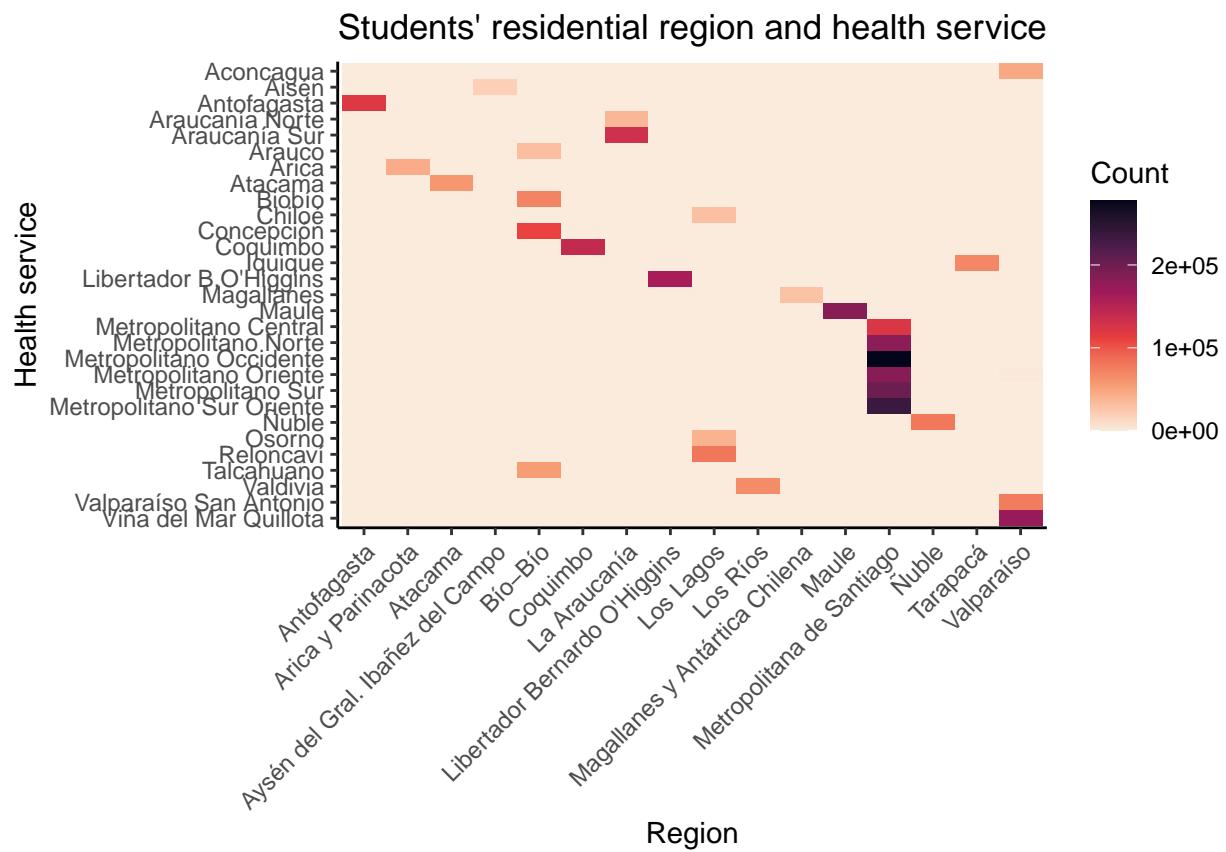


Figure 3: Residential communes aggregated to region level and the health services associated with the aggregated communes, with counts of the number of students resident in the communes in each health service's catchment area.

Communes in La Araucanía

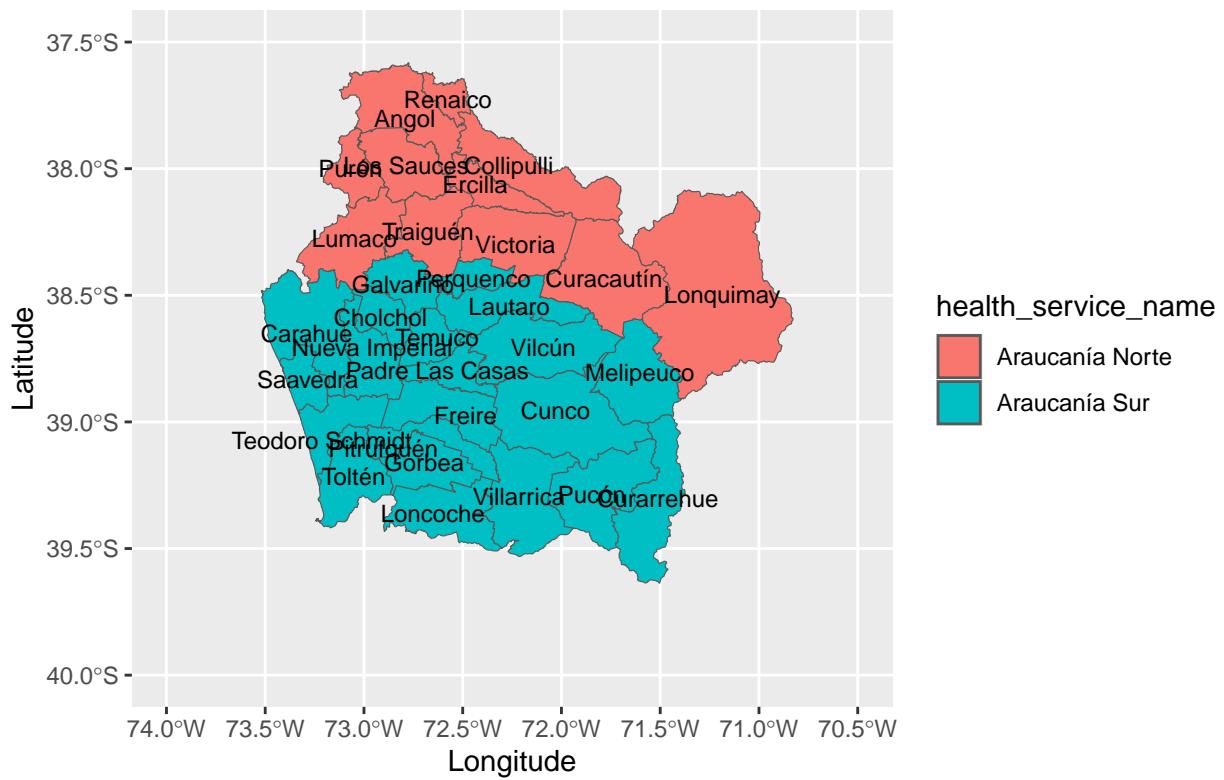


Figure 4: Communes in the Araucanía region, coloured red for the Araucanía Norte (north Araucanía) health services and blue for the Araucanía Sur (south Araucanía) health service.

5 Results

TODO - text explaining figures TODO - table formatting

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% and the global crude prevalence of ADHD was 1.51%, see table 1.

Table 1: Table 1a: Count and prevalence of autism cases by age in Chile school data with normal confidence intervals.

Age band	Autism cases	Prevalence % (95% CI)
6	1806	0.72 (0.69, 0.75)
7	1724	0.69 (0.66, 0.72)
8	1632	0.66 (0.63, 0.69)
9	1523	0.60 (0.57, 0.63)
10	1435	0.56 (0.53, 0.58)
11	1254	0.49 (0.46, 0.52)
12	1176	0.46 (0.43, 0.48)
13	1057	0.42 (0.40, 0.45)
14	805	0.33 (0.31, 0.36)
15	680	0.29 (0.27, 0.31)
16	665	0.28 (0.26, 0.30)
17	491	0.21 (0.19, 0.23)
18	301	0.34 (0.30, 0.38)

Table 2: Table 1b: Count and prevalence of ADHD cases by age in Chile school data with normal confidence intervals.

Age band	ADHD cases	Prevalence % (95% CI)
6	740	0.29 (0.27, 0.32)
7	1965	0.78 (0.75, 0.82)
8	3231	1.31 (1.27, 1.36)
9	4500	1.78 (1.73, 1.83)
10	5485	2.13 (2.07, 2.18)
11	5564	2.17 (2.11, 2.23)
12	5325	2.07 (2.01, 2.12)
13	4848	1.94 (1.88, 1.99)
14	3926	1.62 (1.57, 1.67)
15	3420	1.45 (1.40, 1.50)
16	3120	1.33 (1.28, 1.38)
17	2870	1.23 (1.19, 1.28)
18	1230	1.40 (1.33, 1.48)

```
## # A tibble: 1 x 6
##   crude_ci_lower crude_rate crude_ci_upper adjusted_ci_lower adjusted_~1 adjus~2
##   <dbl>        <dbl>        <dbl>        <dbl>        <dbl>        <dbl>
## 1     0.00468    0.00476     0.00484     0.00457    0.00465  0.00473
```

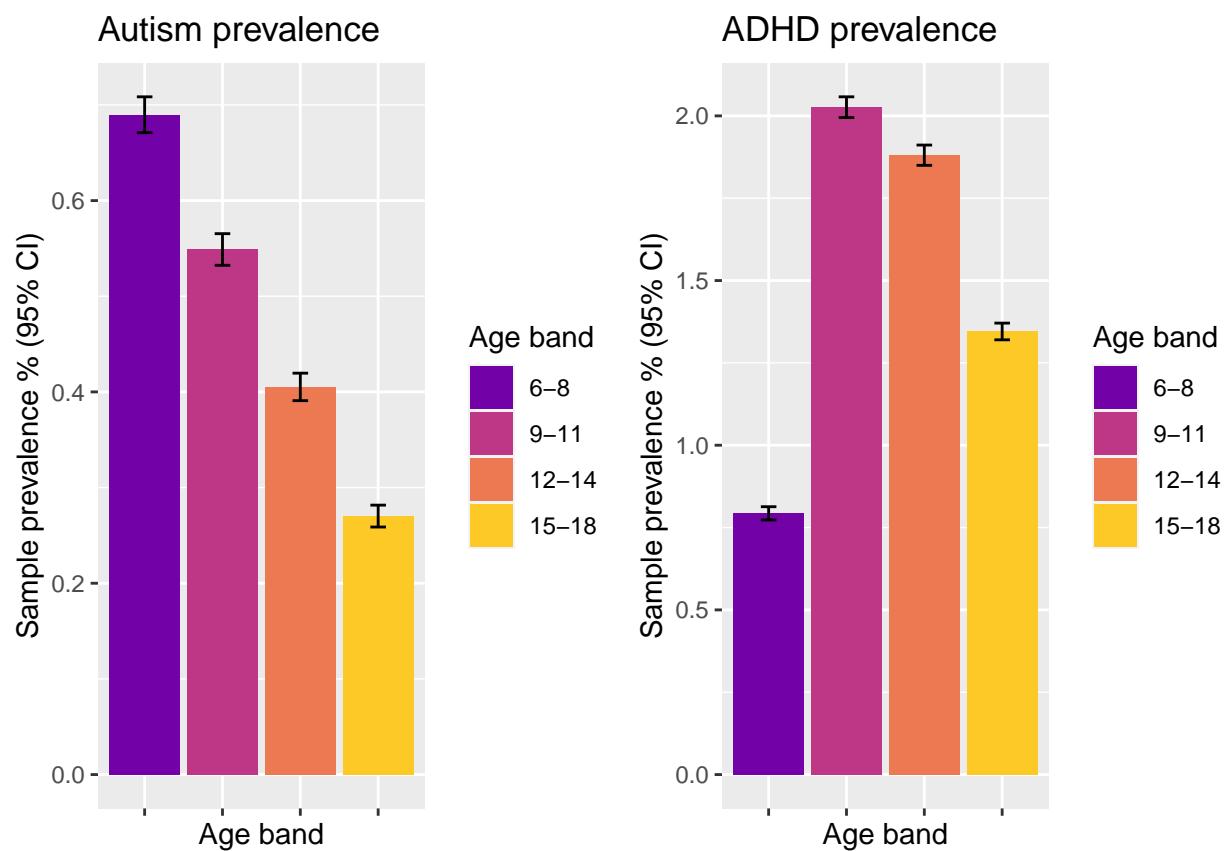


Figure 5: Sample prevalence of autism and adhd by age band. Bars show 95% normal confidence intervals.

```

## # ... with abbreviated variable names 1: adjusted_rate, 2: adjusted_ci_upper
## # A tibble: 1 x 6
##   crude_ci_lower crude_rate crude_ci_upper adjusted_ci_lower adjusted_~1 adjus~2
##       <dbl>        <dbl>        <dbl>        <dbl>        <dbl>        <dbl>
## 1     0.0150      0.0151      0.0153      0.0148      0.0150    0.0151
## # ... with abbreviated variable names 1: adjusted_rate, 2: adjusted_ci_upper

```

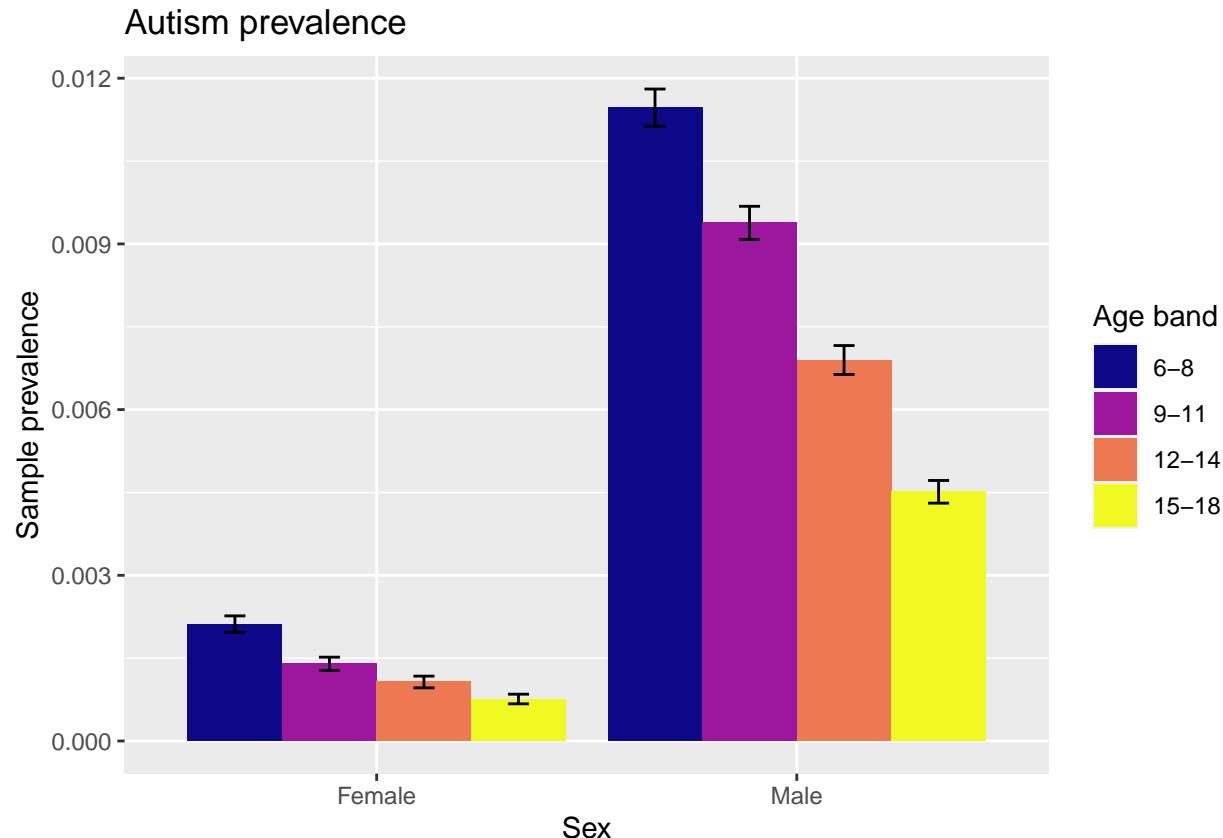


Figure 6: Sample prevalence of autism by age band and sex. Bars show 95% normal confidence intervals.

TODO - table summarising data content 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.

5.1.1 Autism and ADHD prevalence by sex

```

## # A tibble: 2 x 7
##   sex_desc crude_ci_lower crude_rate crude_ci_upper adjusted_c~1 adjus~2 adjus~3
##   <chr>        <dbl>        <dbl>        <dbl>        <dbl>        <dbl>        <dbl>
## 1 Male         0.00787     0.00801     0.00815     0.00773    0.00787    0.00801
## 2 Female       0.00127     0.00133     0.00139     0.00125    0.00131    0.00138
## # ... with abbreviated variable names 1: adjusted_ci_lower, 2: adjusted_rate,
## #   3: adjusted_ci_upper

## # A tibble: 2 x 7
##   sex_desc crude_ci_lower crude_rate crude_ci_upper adjusted_c~1 adjus~2 adjus~3
##   <chr>        <dbl>        <dbl>        <dbl>        <dbl>        <dbl>        <dbl>
## 1 Male         0.00787     0.00801     0.00815     0.00773    0.00787    0.00801
## 2 Female       0.00127     0.00133     0.00139     0.00125    0.00131    0.00138

```

```

## 1 Male          0.0196      0.0198      0.0201      0.0194      0.0197  0.0199
## 2 Female        0.00998     0.0101      0.0103      0.00997     0.0101  0.0103
## # ... with abbreviated variable names 1: adjusted_ci_lower, 2: adjusted_rate,
## #   3: adjusted_ci_upper

```

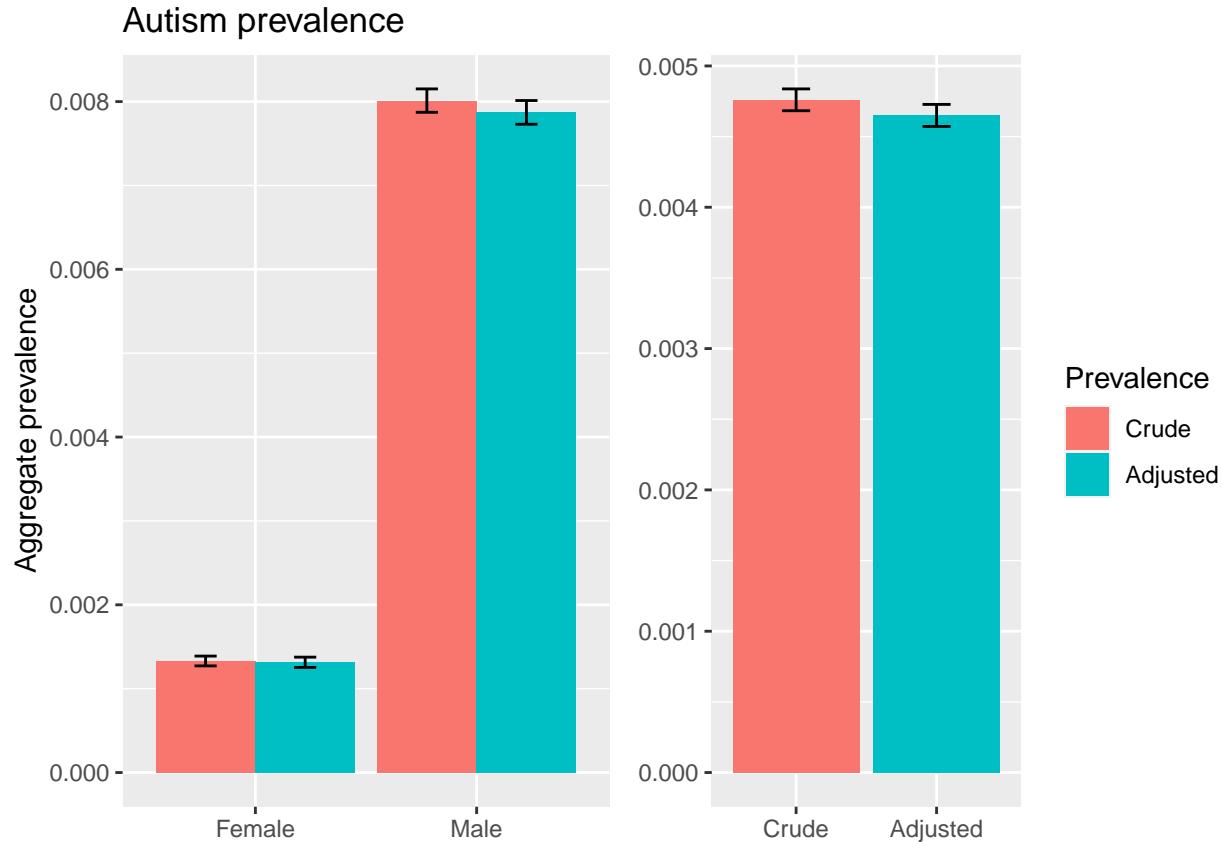


Figure 7: Crude and age- and sex-adjusted sample prevalences of autism, by sex and in aggregate. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

5.1.2 Autism and ADHD prevalence by health service

```

##           health_service_name crude_ci_lower crude_rate crude_ci_upper
## 1             Aconcagua      0.00378    0.00438      0.00497
## 2               Aisén       0.00629    0.00749      0.00869
## 3         Antofagasta      0.00789    0.00841      0.00893
## 4 Araucanía Norte      0.00244    0.00300      0.00356
## 5 Araucanía Sur       0.00336    0.00369      0.00402
## 6            Arauco      0.00637    0.00731      0.00826
## 7              Arica      0.00540    0.00612      0.00684
## 8            Atacama      0.00262    0.00306      0.00351
## 9            Biobío      0.00383    0.00431      0.00479
## 10           Chiloé       0.00375    0.00450      0.00524
## 11        Concepción      0.00732    0.00784      0.00837
## 12           Coquimbo      0.00378    0.00412      0.00445
## 13 Libertador B.O'Higgins 0.00402    0.00434      0.00467
## 14            Maule       0.00280    0.00305      0.00331

```

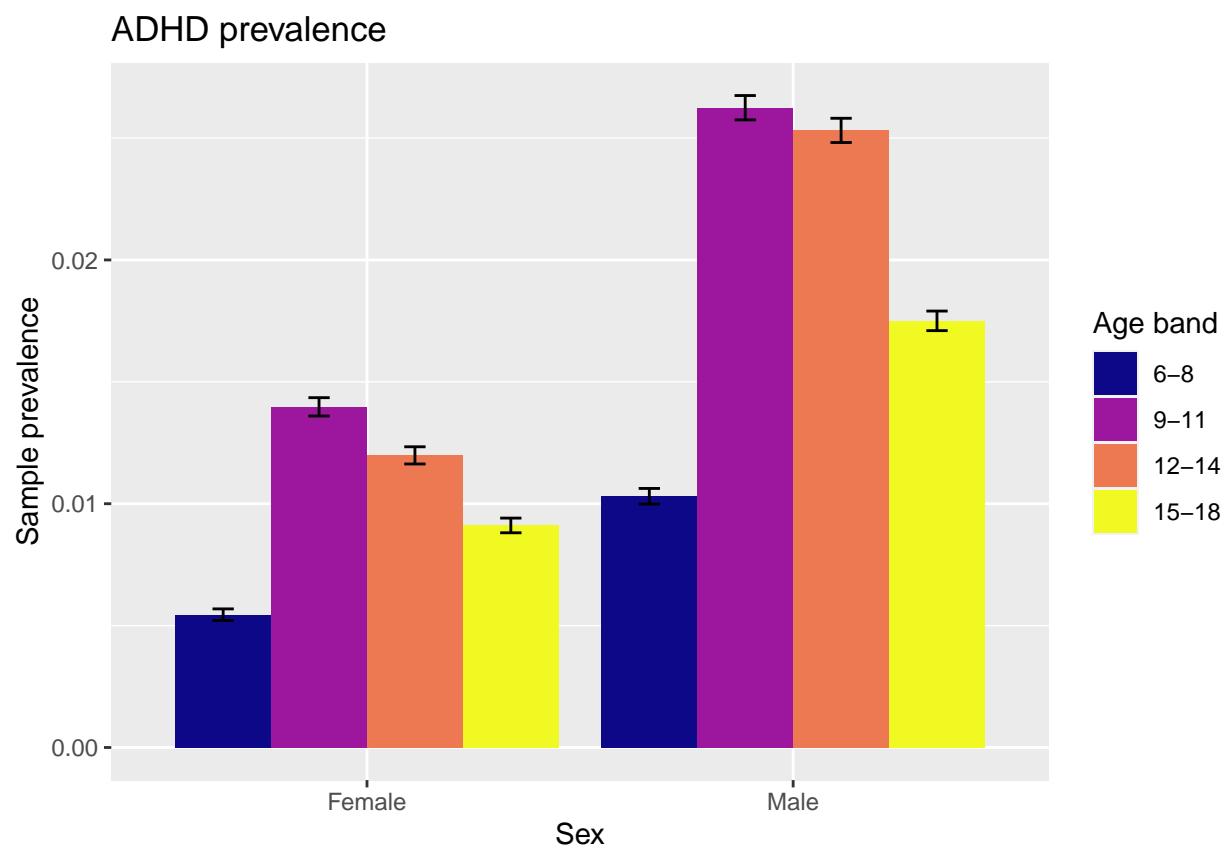


Figure 8: Sample prevalence of ADHD by age band and sex. Bars show 95% normal confidence intervals.

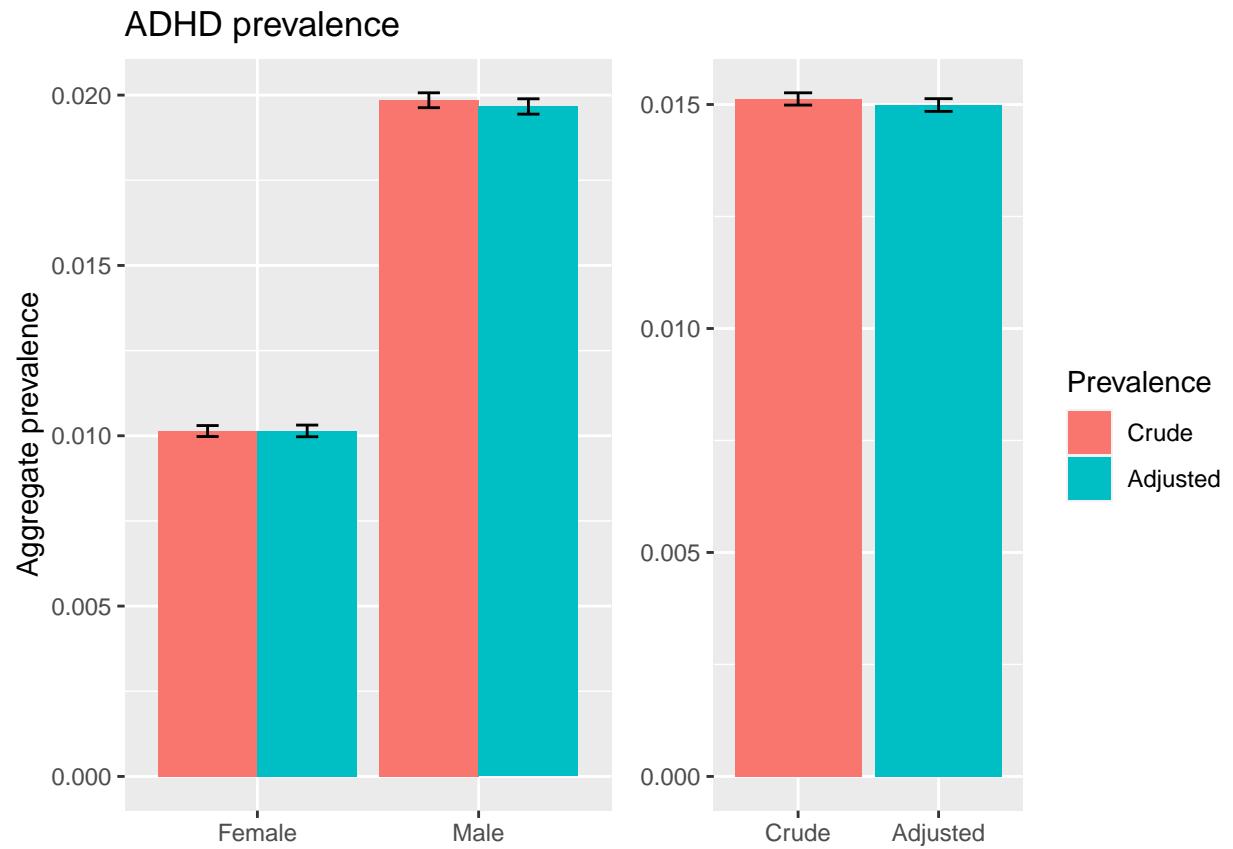


Figure 9: Crude and age- and sex-adjusted sample prevalences of ADHD, by sex and in aggregate. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

		Reloncaví	0.00379	0.00424	0.00469
## 15		Iquique	0.00403	0.00453	0.00503
## 16		Magallanes	0.00725	0.00831	0.00938
## 17		Metropolitano Central	0.00385	0.00421	0.00457
## 18		Metropolitano Norte	0.00269	0.00294	0.00318
## 19		Metropolitano Occidente	0.00336	0.00358	0.00381
## 20		Metropolitano Oriente	0.00279	0.00304	0.00329
## 21		Metropolitano Sur	0.00385	0.00413	0.00442
## 22		Metropolitano Sur Oriente	0.00342	0.00367	0.00391
## 23		Osorno	0.00380	0.00445	0.00510
## 24		Talcahuano	0.00763	0.00839	0.00916
## 25		Valdivia	0.00269	0.00311	0.00354
## 26		Valparaíso San Antonio	0.00635	0.00694	0.00752
## 27		Viña del Mar Quillota	0.00632	0.00670	0.00709
## 28		Ñuble	0.01238	0.01317	0.01396
## 29					
		adjusted_ci_lower	adjusted_rate	adjusted_ci_upper	
## 1		0.00369	0.00427	0.00499	
## 2		0.00634	0.00752	0.00895	
## 3		0.00773	0.00825	0.00882	
## 4		0.00245	0.00300	0.00376	
## 5		0.00338	0.00372	0.00410	
## 6		0.00624	0.00715	0.00825	
## 7		0.00539	0.00614	0.00704	
## 8		0.00266	0.00312	0.00368	
## 9		0.00373	0.00420	0.00475	
## 10		0.00364	0.00433	0.00518	
## 11		0.00718	0.00771	0.00828	
## 12		0.00364	0.00396	0.00433	
## 13		0.00393	0.00425	0.00460	
## 14		0.00277	0.00303	0.00332	
## 15		0.00373	0.00417	0.00467	
## 16		0.00384	0.00431	0.00486	
## 17		0.00724	0.00831	0.00958	
## 18		0.00380	0.00416	0.00456	
## 19		0.00262	0.00287	0.00314	
## 20		0.00316	0.00338	0.00361	
## 21		0.00274	0.00298	0.00325	
## 22		0.00371	0.00398	0.00427	
## 23		0.00339	0.00364	0.00390	
## 24		0.00371	0.00433	0.00507	
## 25		0.00739	0.00813	0.00898	
## 26		0.00260	0.00300	0.00349	
## 27		0.00625	0.00683	0.00749	
## 28		0.00620	0.00658	0.00699	
## 29		0.01208	0.01286	0.01371	
##		health_service_name	crude_ci_lower	crude_rate	crude_ci_upper
## 1		Aconcagua	0.01950	0.02079	0.02209
## 2		Aisén	0.02027	0.02232	0.02438
## 3		Antofagasta	0.00945	0.01001	0.01057
## 4		Araucanía Norte	0.01217	0.01334	0.01452
## 5		Araucanía Sur	0.01358	0.01422	0.01485
## 6		Arauco	0.01501	0.01641	0.01782
## 7		Arica	0.01040	0.01139	0.01237

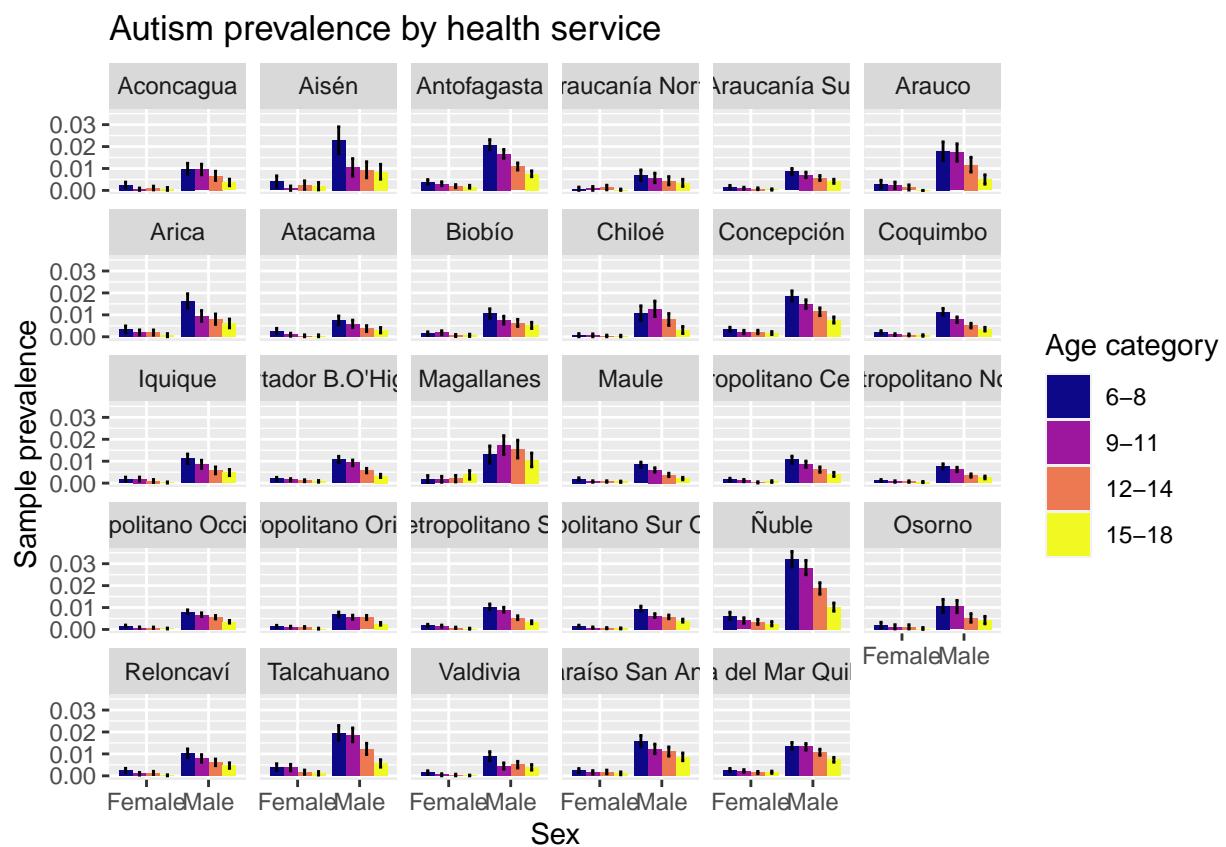


Figure 10: Sample prevalence of autism by health service, age band and sex. Bars show 95% normal confidence intervals.

Autism prevalence by health service



Figure 11: Crude and age- and sex-adjusted sample prevalences of autism by health service. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

```

## 8          Atacama    0.00435  0.00492  0.00549
## 9          Biobío    0.02159  0.02269  0.02378
## 10         Chiloé     0.02771  0.02960  0.03149
## 11         Concepción 0.02842  0.02942  0.03043
## 12         Coquimbo   0.01907  0.01980  0.02053
## 13  Libertador B.O'Higgins 0.01662  0.01726  0.01789
## 14          Maule     0.01143  0.01193  0.01243
## 15         Reloncaví  0.00952  0.01022  0.01092
## 16          Iquique   0.01403  0.01493  0.01583
## 17         Magallanes  0.02870  0.03072  0.03274
## 18  Metropolitano Central 0.01457  0.01526  0.01594
## 19  Metropolitano Norte   0.01361  0.01416  0.01471
## 20  Metropolitano Occidente 0.01054  0.01093  0.01131
## 21  Metropolitano Oriente  0.01163  0.01213  0.01263
## 22  Metropolitano Sur     0.01372  0.01424  0.01476
## 23 Metropolitano Sur Oriente 0.01512  0.01562  0.01611
## 24          Osorno     0.00946  0.01046  0.01145
## 25          Talcahuano 0.02926  0.03071  0.03215
## 26          Valdivia   0.00998  0.01077  0.01156
## 27  Valparaíso San Antonio 0.01132  0.01209  0.01286
## 28          Viña del Mar Quillota 0.01123  0.01174  0.01224
## 29          Ñuble      0.02019  0.02120  0.02220
##   adjusted_ci_lower adjusted_rate adjusted_ci_upper
## 1      0.01913    0.02044    0.02187
## 2      0.01970    0.02171    0.02398
## 3      0.00925    0.00983    0.01045
## 4      0.01179    0.01294    0.01426
## 5      0.01320    0.01384    0.01452
## 6      0.01497    0.01642    0.01806
## 7      0.01022    0.01122    0.01237
## 8      0.00431    0.00488    0.00556
## 9      0.02146    0.02260    0.02382
## 10     0.02680    0.02867    0.03069
## 11     0.02889    0.02998    0.03112
## 12     0.01922    0.02000    0.02082
## 13     0.01626    0.01691    0.01759
## 14     0.01105    0.01154    0.01207
## 15     0.00923    0.00991    0.01065
## 16     0.01403    0.01497    0.01599
## 17     0.02853    0.03064    0.03294
## 18     0.01426    0.01495    0.01568
## 19     0.01361    0.01417    0.01476
## 20     0.01082    0.01123    0.01166
## 21     0.01146    0.01196    0.01248
## 22     0.01354    0.01407    0.01462
## 23     0.01481    0.01531    0.01583
## 24     0.00922    0.01020    0.01129
## 25     0.02872    0.03022    0.03182
## 26     0.00983    0.01062    0.01150
## 27     0.01129    0.01208    0.01294
## 28     0.01097    0.01147    0.01201
## 29     0.02001    0.02105    0.02217
## # A tibble: 21 x 7

```

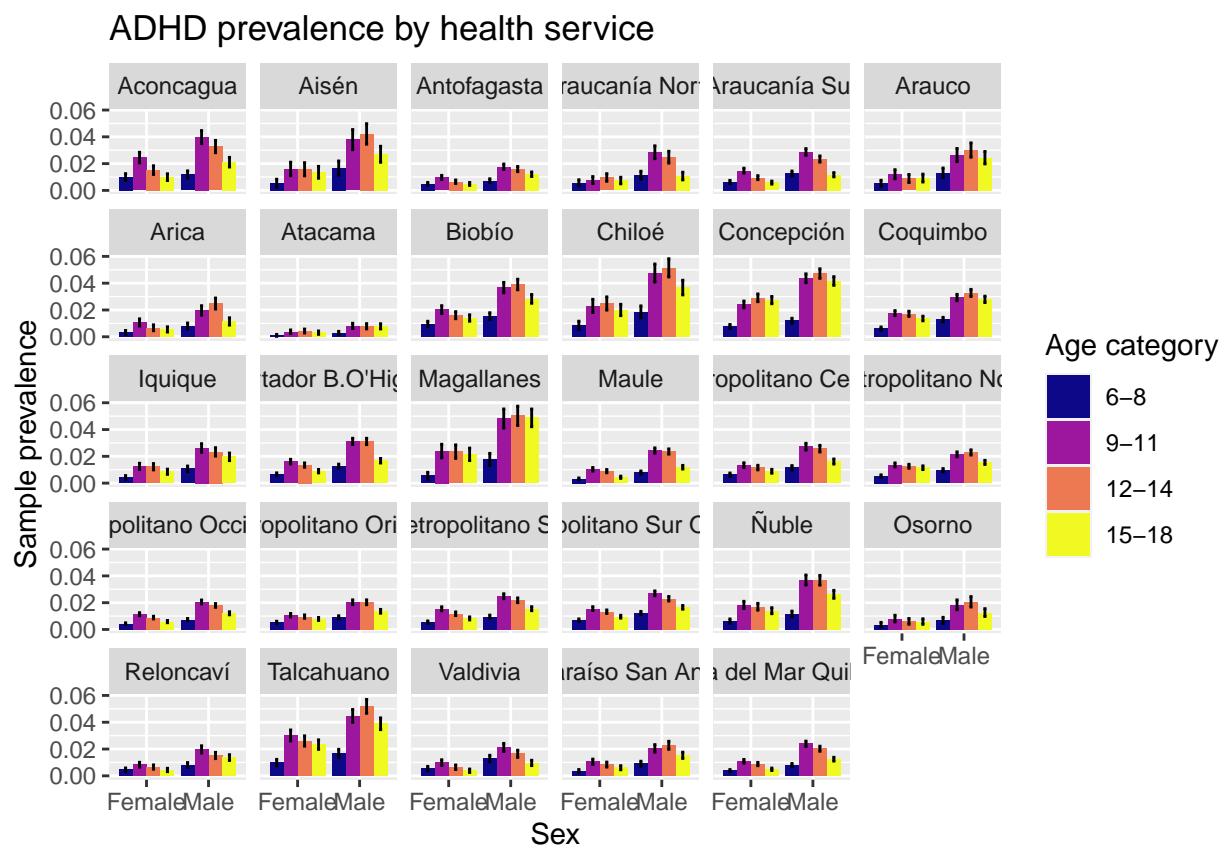


Figure 12: Sample prevalence of ADHD by health service, age band and sex. Bars show 95% normal confidence intervals.

ADHD prevalence by health service



Figure 13: Crude and age- and sex-adjusted sample prevalences of ADHD by health service. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

```

##   commune_name crude_ci_lower crude_rate crude_ci_upper adjus~1 adjus~2 adjus~3
##   <chr>          <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Carahue        0.00116    0.00266    0.00416  0.00129  0.0025   0.00588
## 2 Cholchol        0          0.00138    0.00295  0.00026  0.00128  0.00816
## 3 Cunco          0.00229    0.0048     0.00732  0.00258  0.00472  0.0100
## 4 Curarrehue     0          0.00147    0.00351  0.00017  0.00142  0.00804
## 5 Freire          0.00195    0.00381    0.00567  0.0021   0.00369  0.00672
## 6 Galvarino       0.00196    0.00478    0.0076   0.0023   0.00463  0.0101
## 7 Gorbea          0.00235    0.0054     0.00844  0.00258  0.0052   0.0118
## 8 Lautaro         0.00255    0.00401    0.00546  0.00259  0.00387  0.00647
## 9 Loncoche        0.00062    0.00202    0.00341  0.0009   0.00295  0.00845
## 10 Melipeuco      0          0.00187    0.00445  0.00022  0.00183  0.0144
## # ... with 11 more rows, and abbreviated variable names 1: adjusted_ci_lower,
## #   2: adjusted_rate, 3: adjusted_ci_upper

```

Autism prevalence by commune in Araucanía Sur

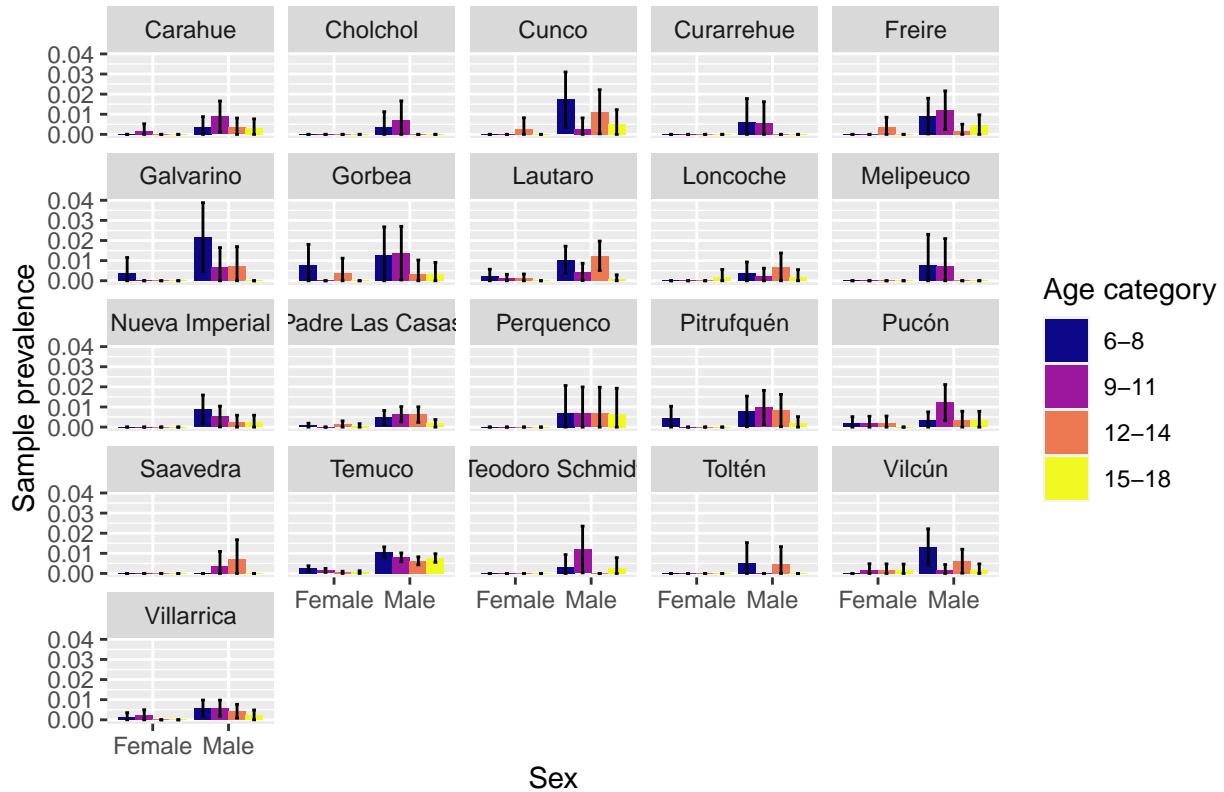


Figure 14: Sample prevalence of autism by commune in Araucanía Sur health service, age band and sex. Bars show 95% normal confidence intervals.

5.1.3 Autism and ADHD prevalence by socio-economic status

```

## # A tibble: 7 x 7
##   school_fee      crude_ci_lower crude_rate crude_ci_~1 adjus~2 adjus~3 adjus~4
##   <fct>          <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Free            0.00556    0.00566    0.00576  0.00537  0.00547  0.00557
## 2 $1,000-$10,000 0.00221    0.00714    0.0121   0.00293  0.00686  0.0325
## 3 $10,001-$25,000 0.00154    0.002      0.00246  0.00157  0.00203  0.00269

```

Autism prevalence by commune in Araucanía Sur

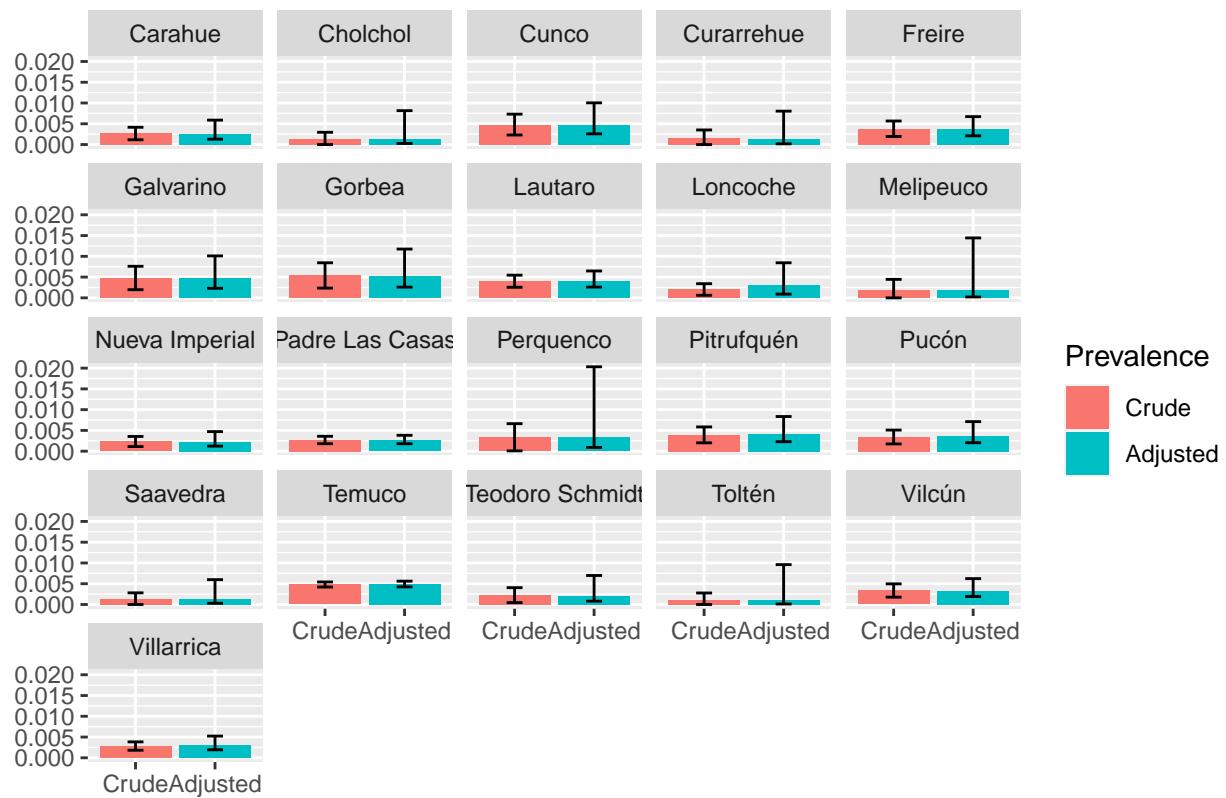


Figure 15: Crude and age- and sex-adjusted sample prevalences of autism by commune in Araucanía Sur health service. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

```

## 4 $25,001-$50,000      0.0028      0.00303     0.00327 0.0029 0.00316 0.00345
## 5 $50,001-$100,000    0.00365     0.00388     0.00411 0.00373 0.00399 0.00427
## 6 $100,001+            0.00039     0.00047     0.00055 0.00039 0.00046 0.00056
## 7 No information       0.00442     0.00504     0.00566 0.00403 0.00458 0.00524
## # ... with abbreviated variable names 1: crude_ci_upper, 2: adjusted_ci_lower,
## #   3: adjusted_rate, 4: adjusted_ci_upper

```

Autism prevalence by SES status



Figure 16: Sample prevalence of autism by socio-economic (SES) status of student's family, age band and sex. Bars show 95% normal confidence intervals.

```

## # A tibble: 7 x 7
##   school_fee      crude_ci_lower crude_rate crude_ci_~1 adjus~2 adjus~3 adjus~4
##   <fct>          <dbl>        <dbl>        <dbl>        <dbl>        <dbl>        <dbl>
## 1 Free             0.0163       0.0165       0.0167       0.0160       0.0162       0.0164
## 2 $1,000-$10,000    0           0.00179      0.00426      0.00013      0.00108      0.0274
## 3 $10,001-$25,000   0.00953     0.0106       0.0116       0.0094       0.0105       0.0117
## 4 $25,001-$50,000   0.0154       0.0159       0.0165       0.0159       0.0165       0.0172
## 5 $50,001-$100,000  0.0184       0.0190       0.0195       0.0184       0.0190       0.0196
## 6 $100,001+          0.00208     0.00225      0.00242      0.00212      0.00231      0.00251
## 7 No information     0.0111       0.0121       0.0131       0.0113       0.0122       0.0133
## # ... with abbreviated variable names 1: crude_ci_upper, 2: adjusted_ci_lower,
## #   3: adjusted_rate, 4: adjusted_ci_upper

```

5.1.4 Autism and ADHD prevalence by ethnicity

```

## # A tibble: 3 x 7
##   ethnic_2_group      crude_ci_lower crude_rate crude_ci_~1 adjus~2 adjus~3 adjus~4
##   <fct>          <dbl>        <dbl>        <dbl>        <dbl>        <dbl>        <dbl>
## 1 White              0.0163       0.0165       0.0167       0.0160       0.0162       0.0164
## 2 Black              0.0111       0.0121       0.0131       0.0113       0.0122       0.0133
## 3 Other              0.0163       0.0165       0.0167       0.0160       0.0162       0.0164

```

Autism prevalence by SES status



Figure 17: Crude and age- and sex-adjusted sample prevalences of autism by socio-economic (SES) status of student's family. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

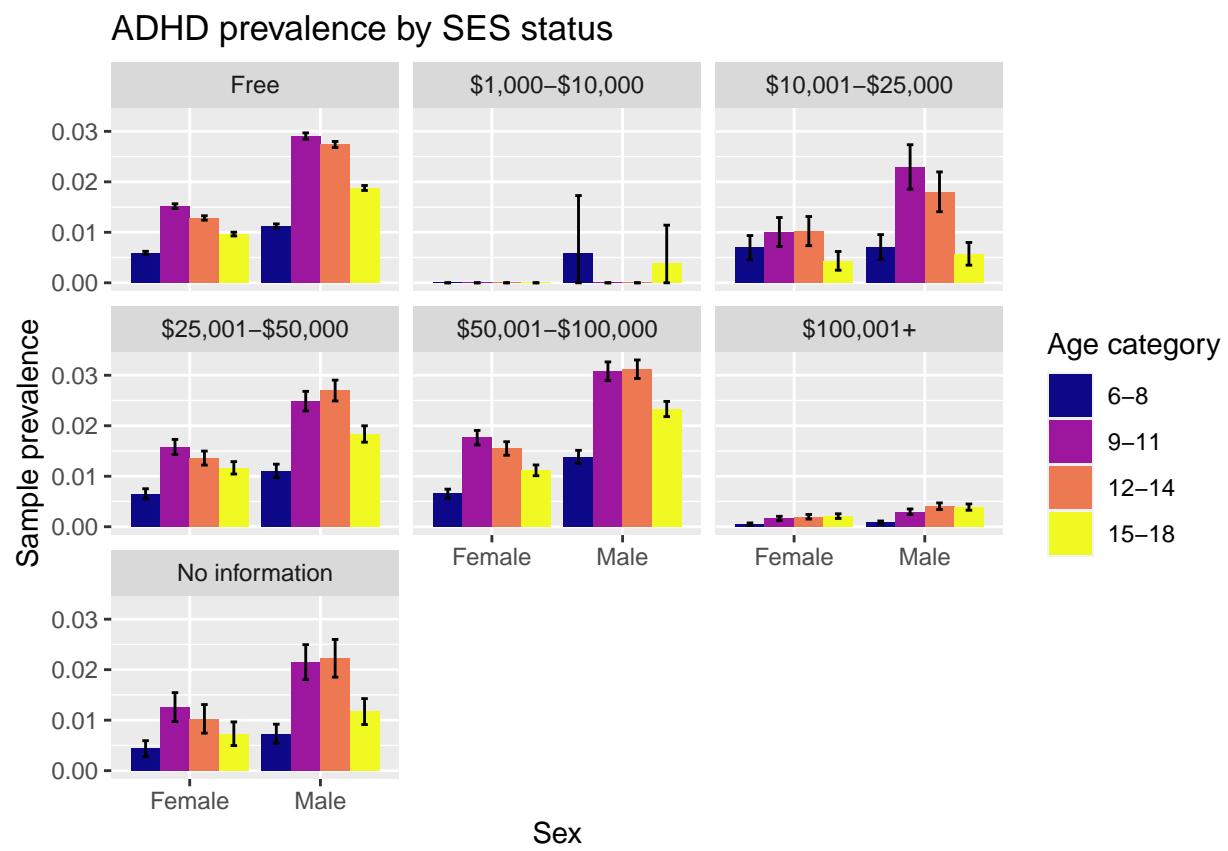


Figure 18: Sample prevalence of ADHD by socio-economic (SES) status of student's family, age band and sex. Bars show 95% normal confidence intervals.

ADHD prevalence by SES status



Figure 19: Crude and age- and sex-adjusted sample prevalences of ADHD by socio-economic (SES) status of student's family. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

```

##   <chr>          <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Mapuche        0.00335    0.00363    0.00391   0.0032    0.00347  0.00376
## 2 No Indigenous group 0.00473    0.00481    0.00489   0.00462   0.0047   0.00479
## 3 Other ethnic group 0.00557    0.0064     0.00723   0.00523   0.00604  0.00702
## # ... with abbreviated variable names 1: crude_ci_upper, 2: adjusted_ci_lower,
## #   3: adjusted_rate, 4: adjusted_ci_upper

```

Autism prevalence by ethnicity

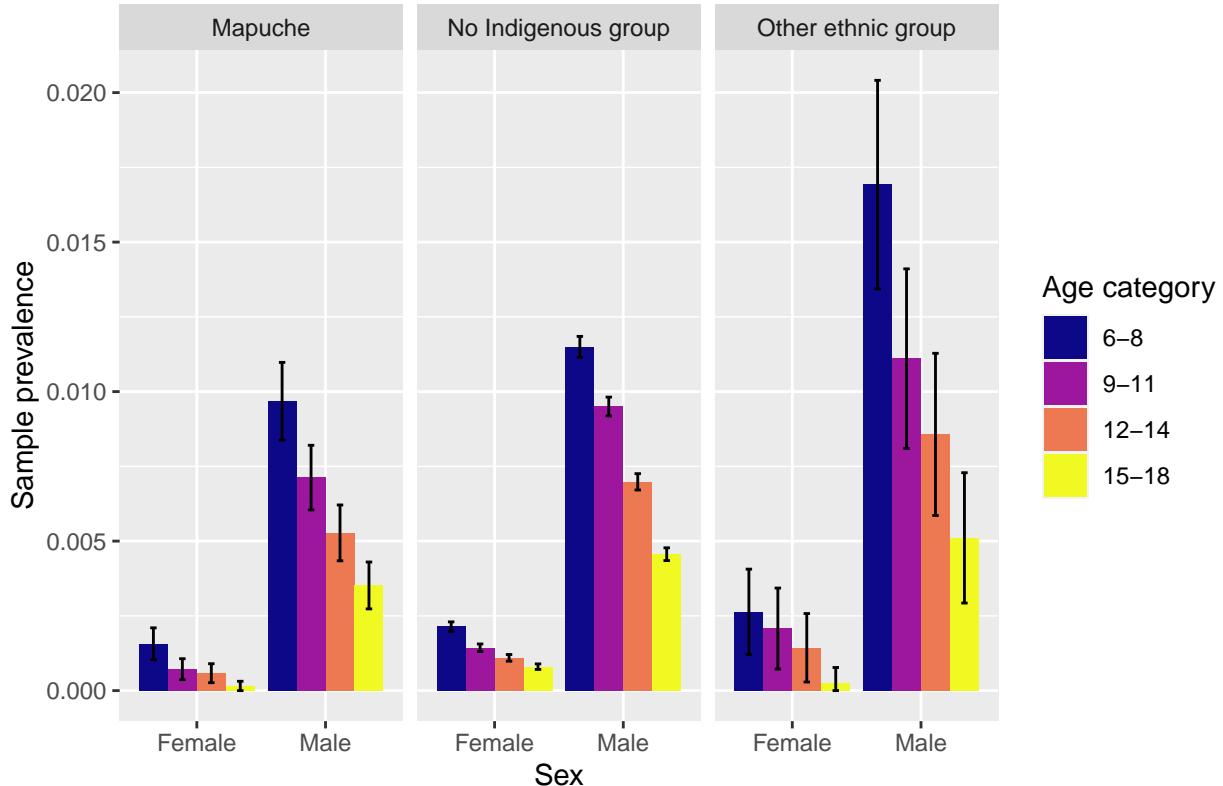


Figure 20: Sample prevalence of autism by ethnicity, age band and sex. Bars show 95% normal confidence intervals.

```

## # A tibble: 3 x 7
##   ethnic_2_group      crude_ci_lower crude_rate crude_~1 adjus~2 adjus~3 adjus~4
##   <chr>                  <dbl>       <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Mapuche            0.0132      0.0138     0.0143     0.0128     0.0133     0.0139
## 2 No Indigenous group 0.0151      0.0152     0.0154     0.0150     0.0151     0.0153
## 3 Other ethnic group  0.0104      0.0115     0.0126     0.0106     0.0118     0.0132
## # ... with abbreviated variable names 1: crude_ci_upper, 2: adjusted_ci_lower,
## #   3: adjusted_rate, 4: adjusted_ci_upper

```

5.1.5 Autism and ADHD prevalence by rurality

```

## # A tibble: 2 x 7
##   school_rurality crude_ci_lower crude_rate crude_ci_u~1 adjus~2 adjus~3 adjus~4
##   <chr>                  <dbl>       <dbl>      <dbl>      <dbl>      <dbl>
## 1 Rural                0.00629    0.00662    0.00694   0.00541   0.00573  0.0061
## 2 Urban                0.00452    0.0046     0.00468   0.00447   0.00455  0.00463
## # ... with abbreviated variable names 1: crude_ci_upper, 2: adjusted_ci_lower,

```

Autism prevalence by ethnicity

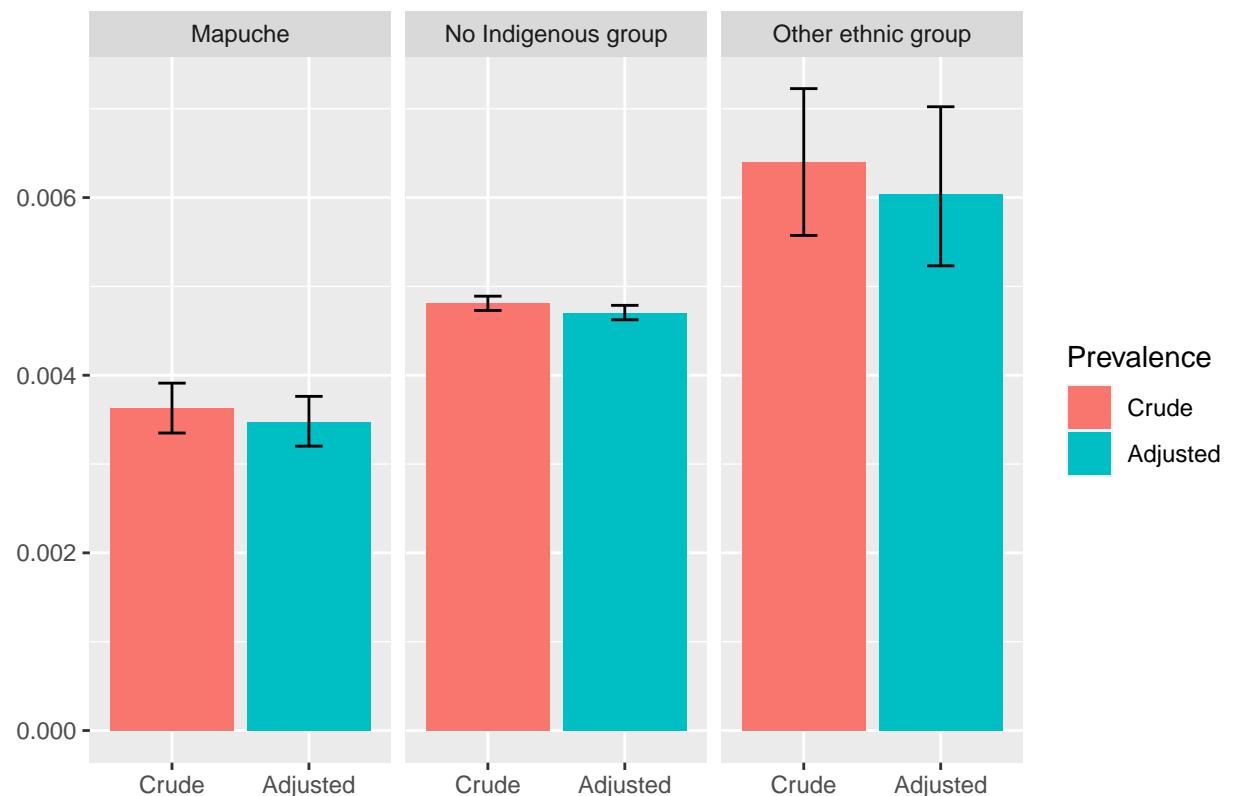


Figure 21: Crude and age- and sex-adjusted sample prevalences of autism by ethnicity. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

ADHD prevalence by ethnicity

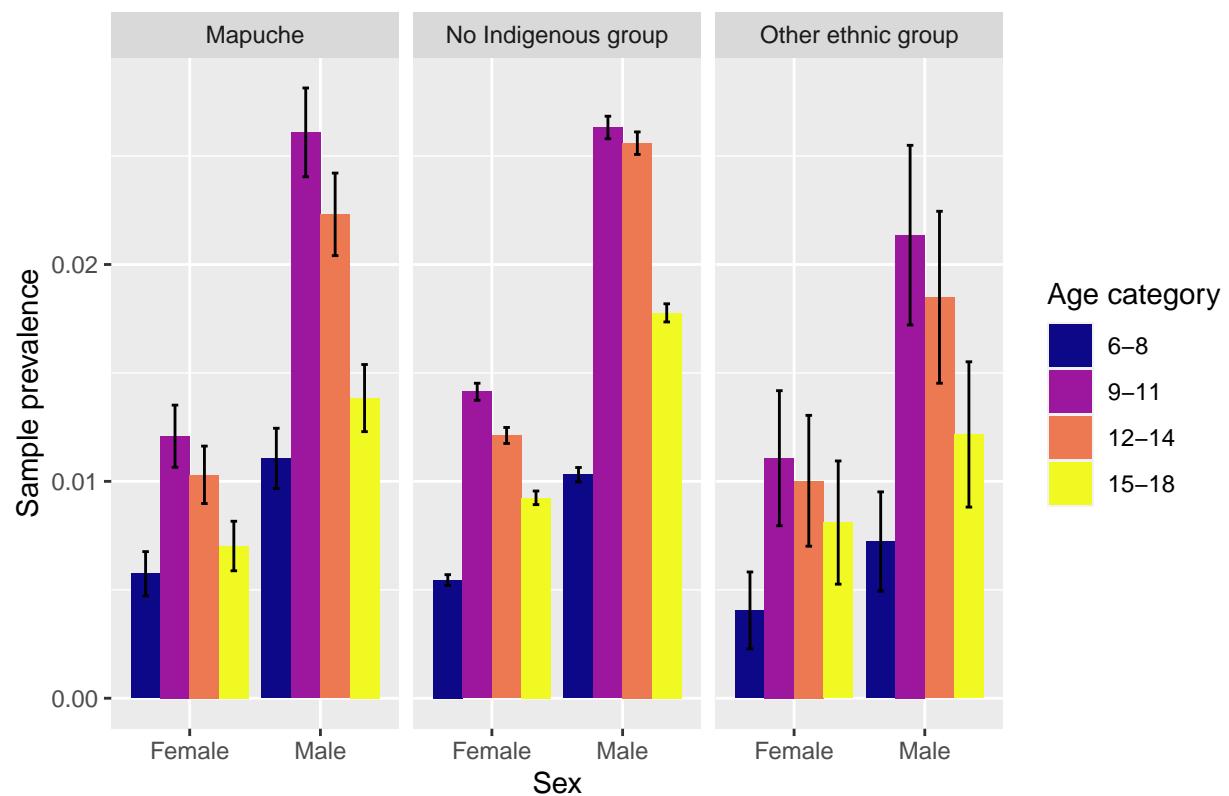


Figure 22: Sample prevalence of ADHD by ethnicity, age band and sex. Bars show 95% normal confidence intervals.

ADHD prevalence by ethnicity

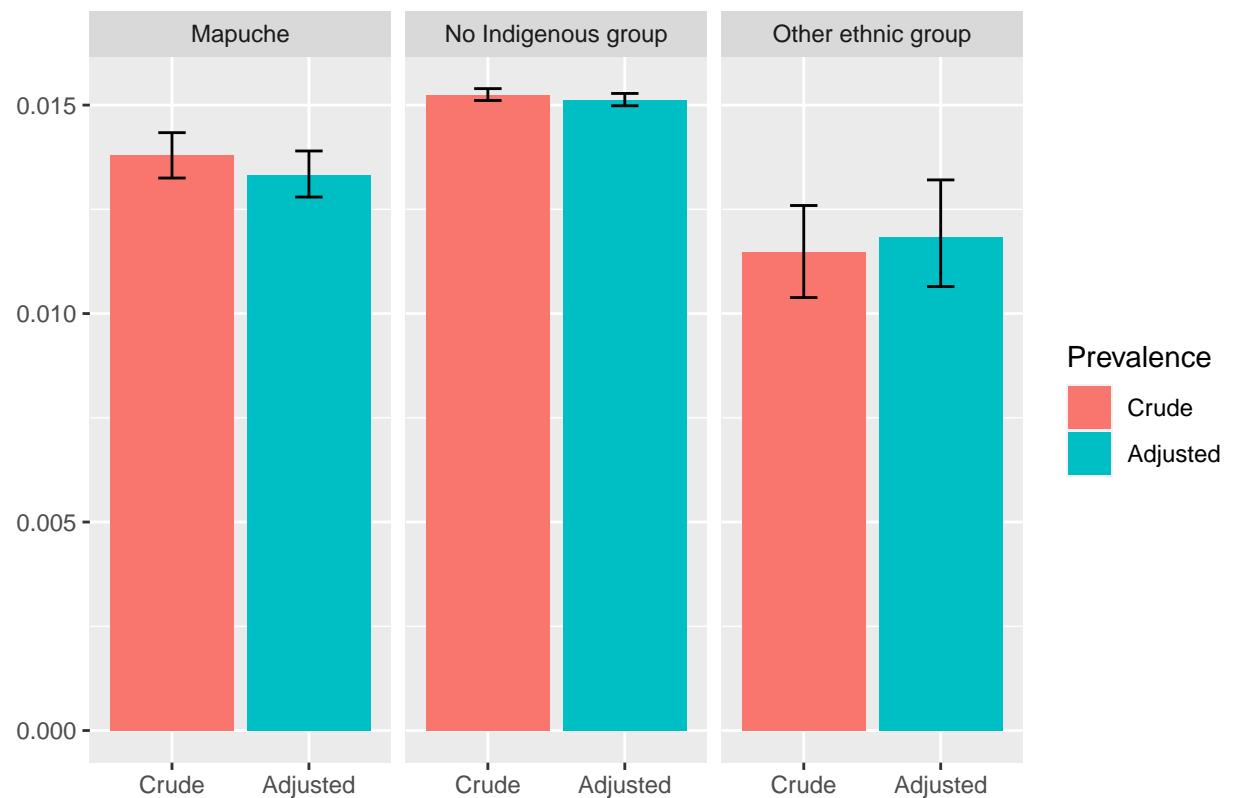


Figure 23: Crude and age- and sex-adjusted sample prevalences of ADHD by ethnicity. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

```
## # 3: adjusted_rate, 4: adjusted_ci_upper
```

Autism prevalence by school's rurality

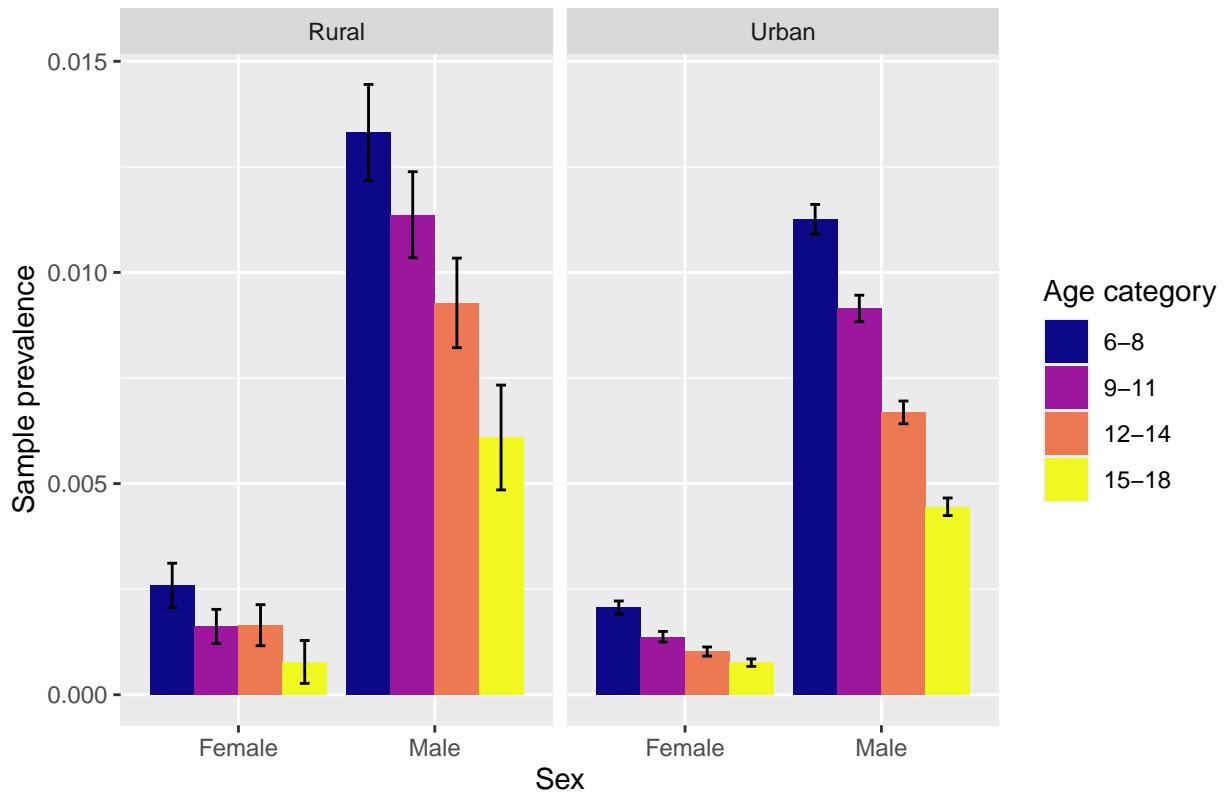


Figure 24: Sample prevalence of autism by school's rurality, age band and sex. Bars show 95% normal confidence intervals.

```
## # A tibble: 2 x 7
##   school_rurality crude_ci_lower crude_rate crude_ci_u~1 adjus~2 adjus~3 adjus~4
##   <chr>              <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Rural               0.0172     0.0177     0.0182     0.0161     0.0167     0.0174
## 2 Urban                0.0148     0.0149     0.0150     0.0147     0.0148     0.0150
## # ... with abbreviated variable names 1: crude_ci_upper, 2: adjusted_ci_lower,
## #   3: adjusted_rate, 4: adjusted_ci_upper
```

5.2 Frequentist prevalence estimation

5.3 Clinical data

TODO - table summarising data content. Number of unique patients

Autism prevalence by school's rurality

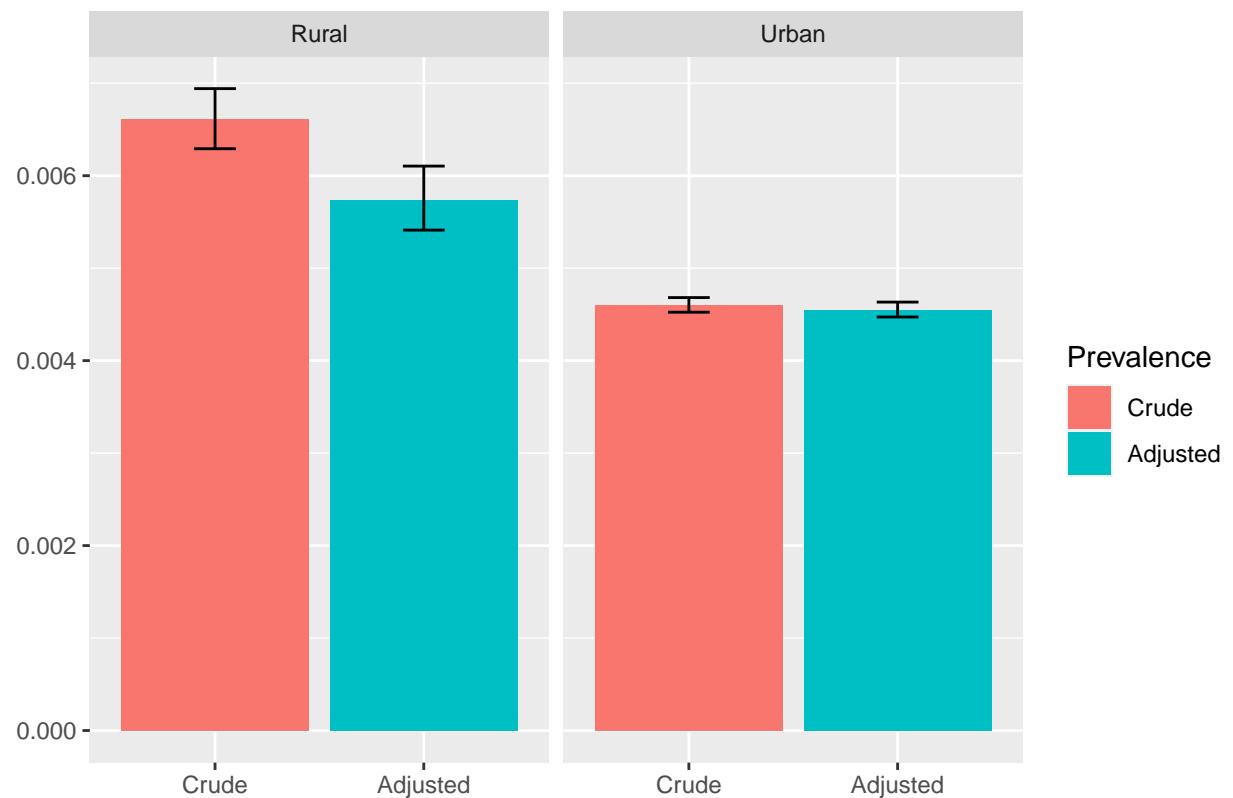


Figure 25: Crude and age- and sex-adjusted sample prevalences of autism by school's rurality. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

ADHD prevalence by school's rurality

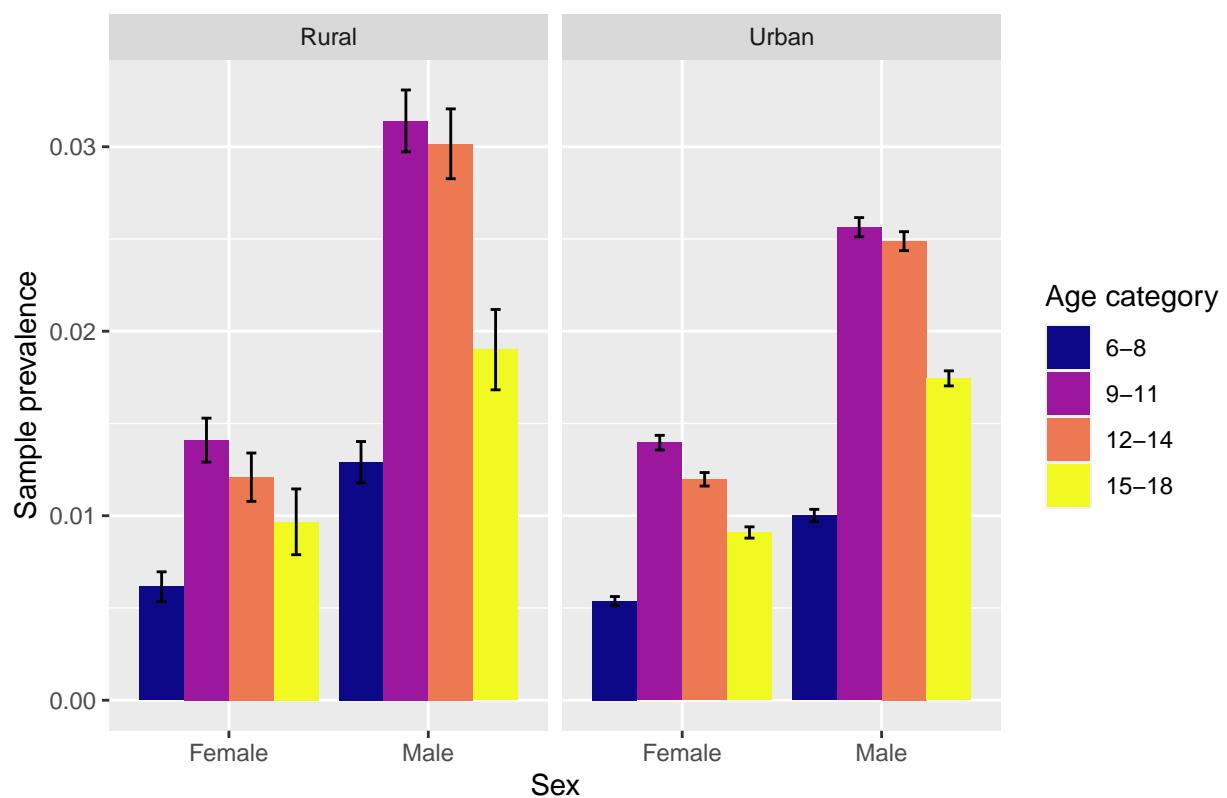


Figure 26: Sample prevalence of ADHD by school's rurality, age band and sex. Bars show 95% normal confidence intervals.

ADHD prevalence by rurality

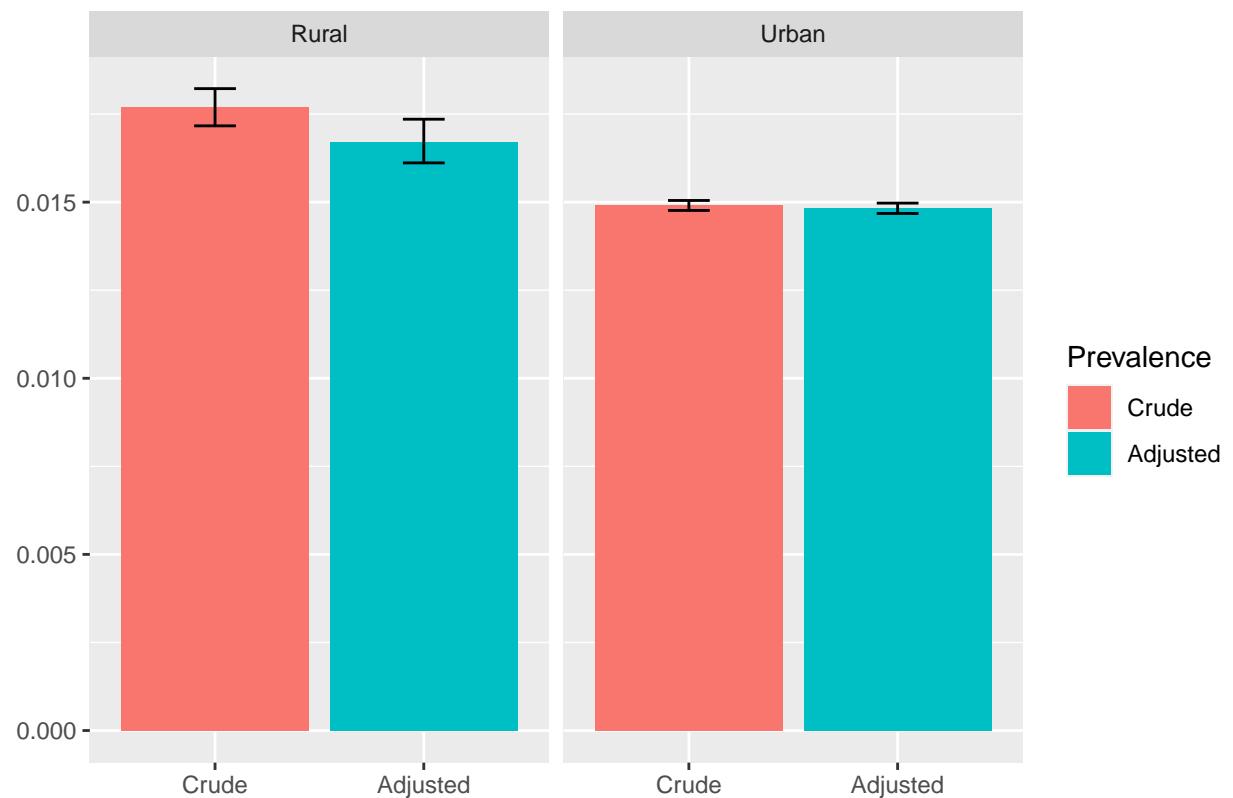
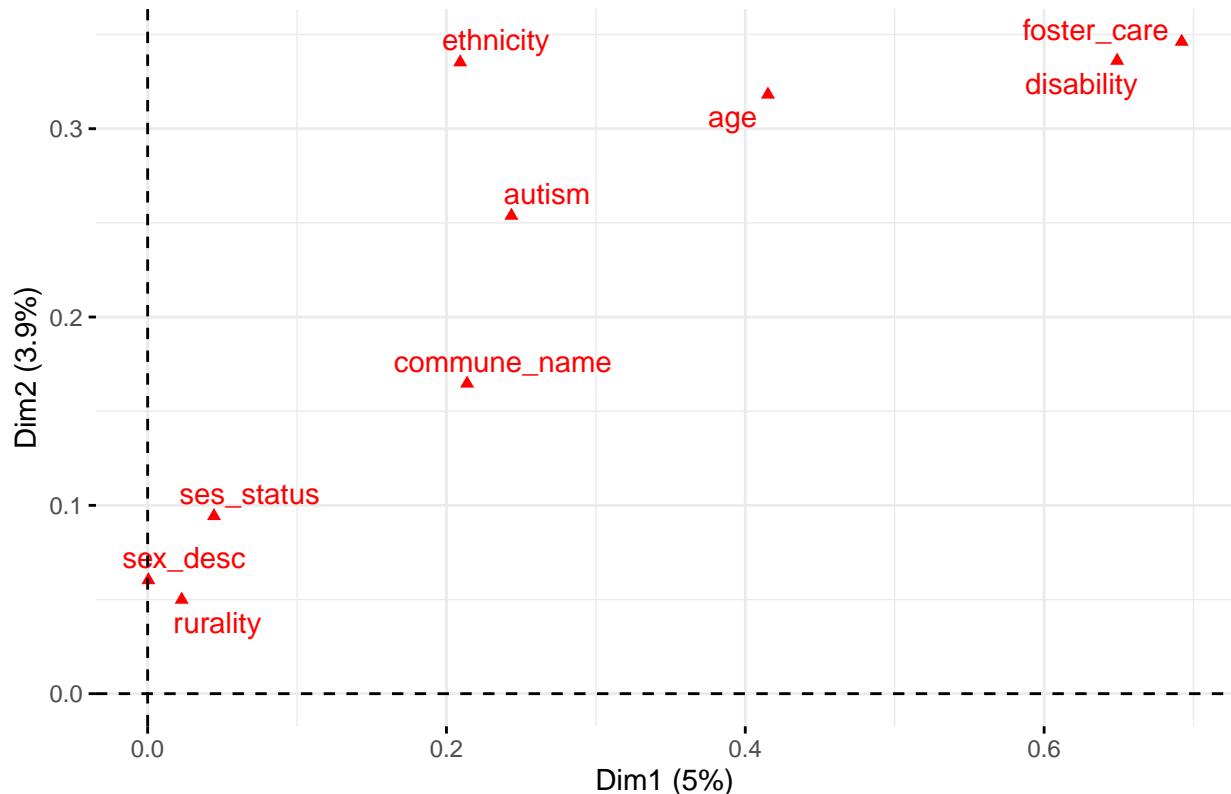


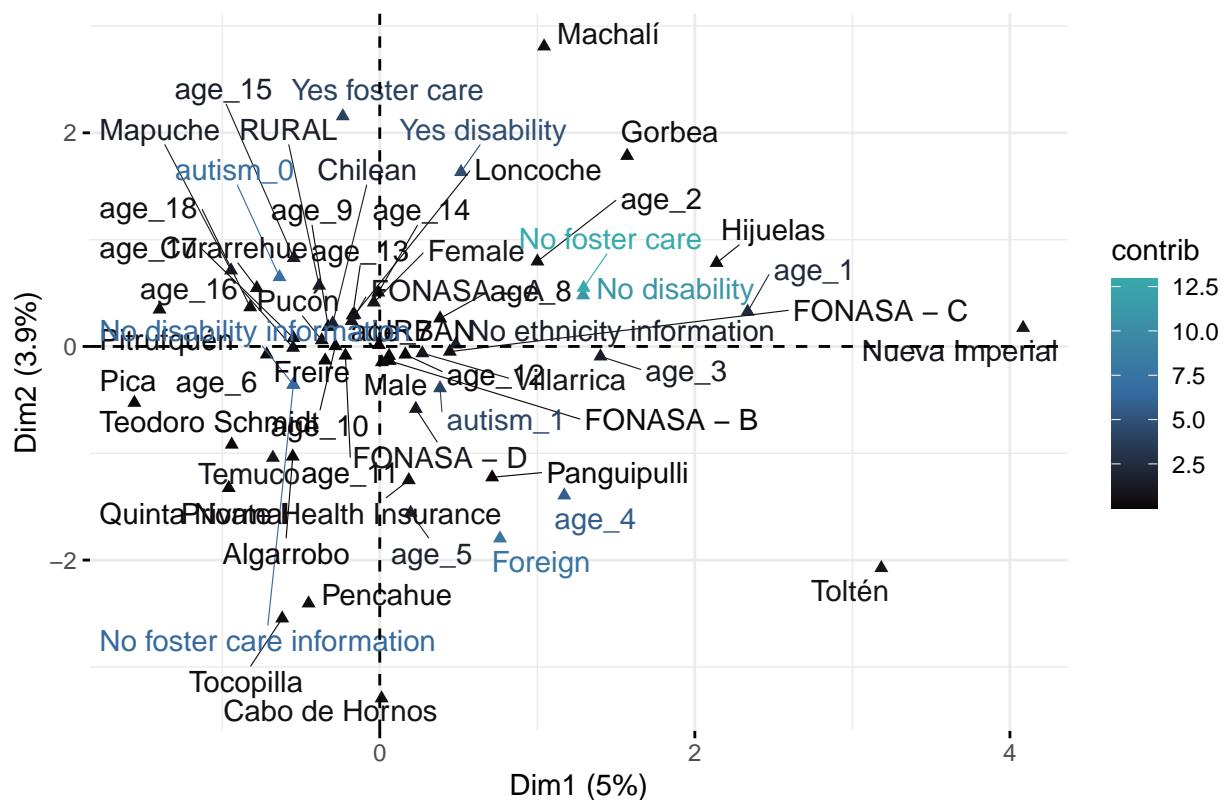
Figure 27: Crude and age- and sex-adjusted sample prevalences of ADHD by school's rurality. Bars for crude prevalence show 95% normal confidence intervals and bars for adjusted prevalence show 95% gamma confidence intervals.

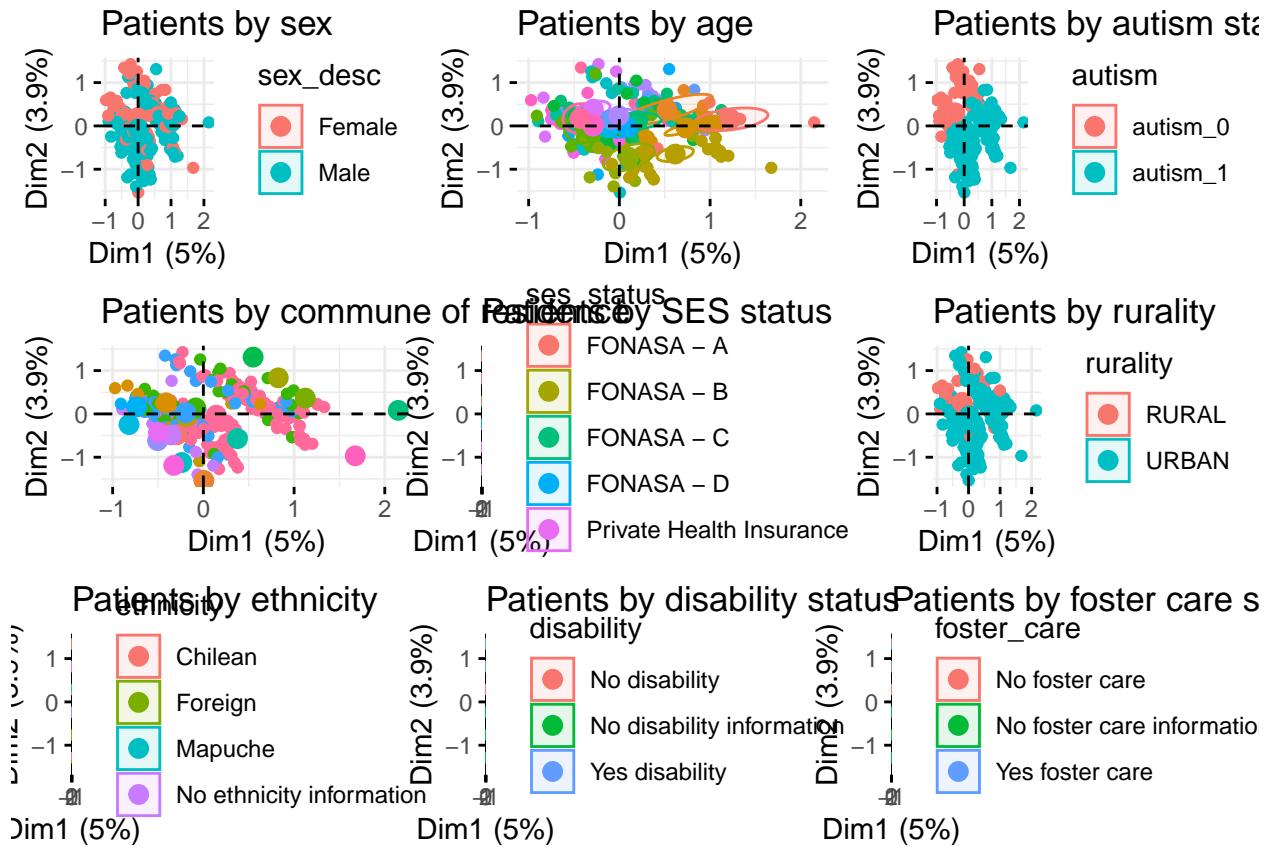
5.4 MCA

Variables – MCA

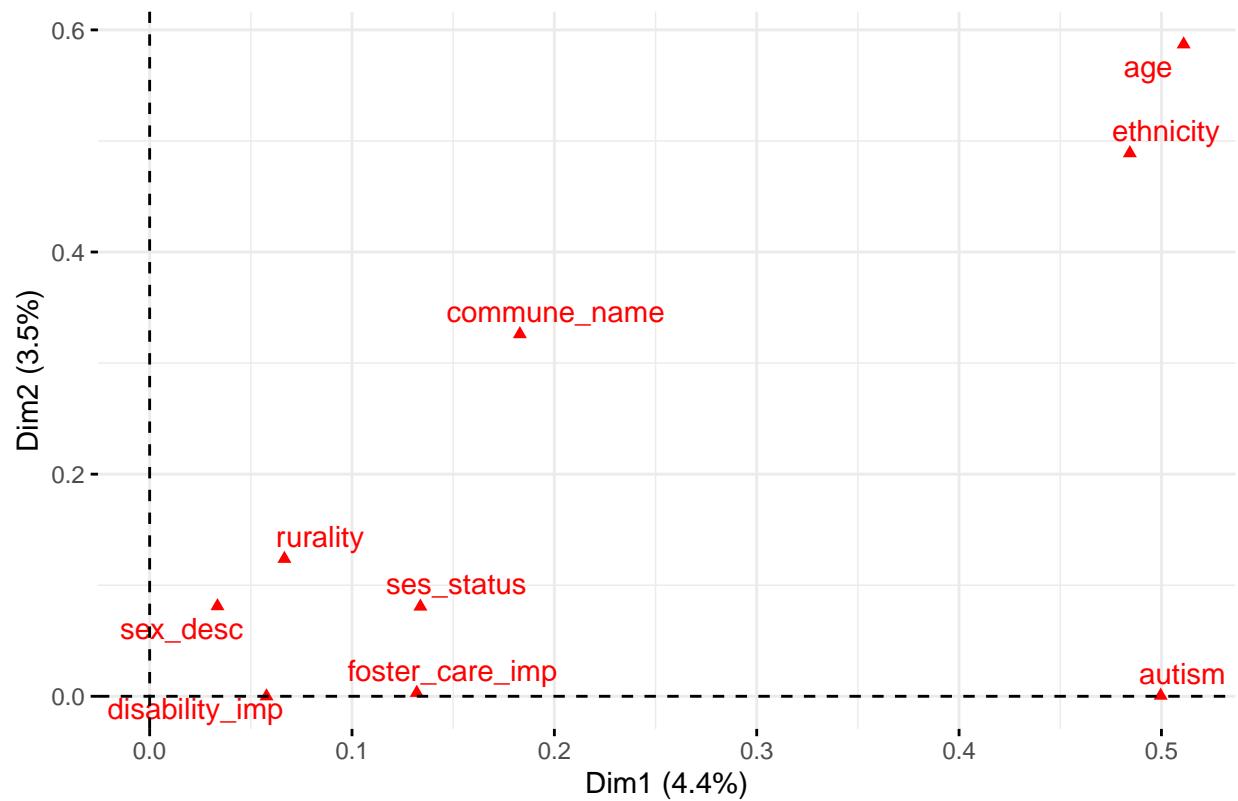


Variable categories – MCA

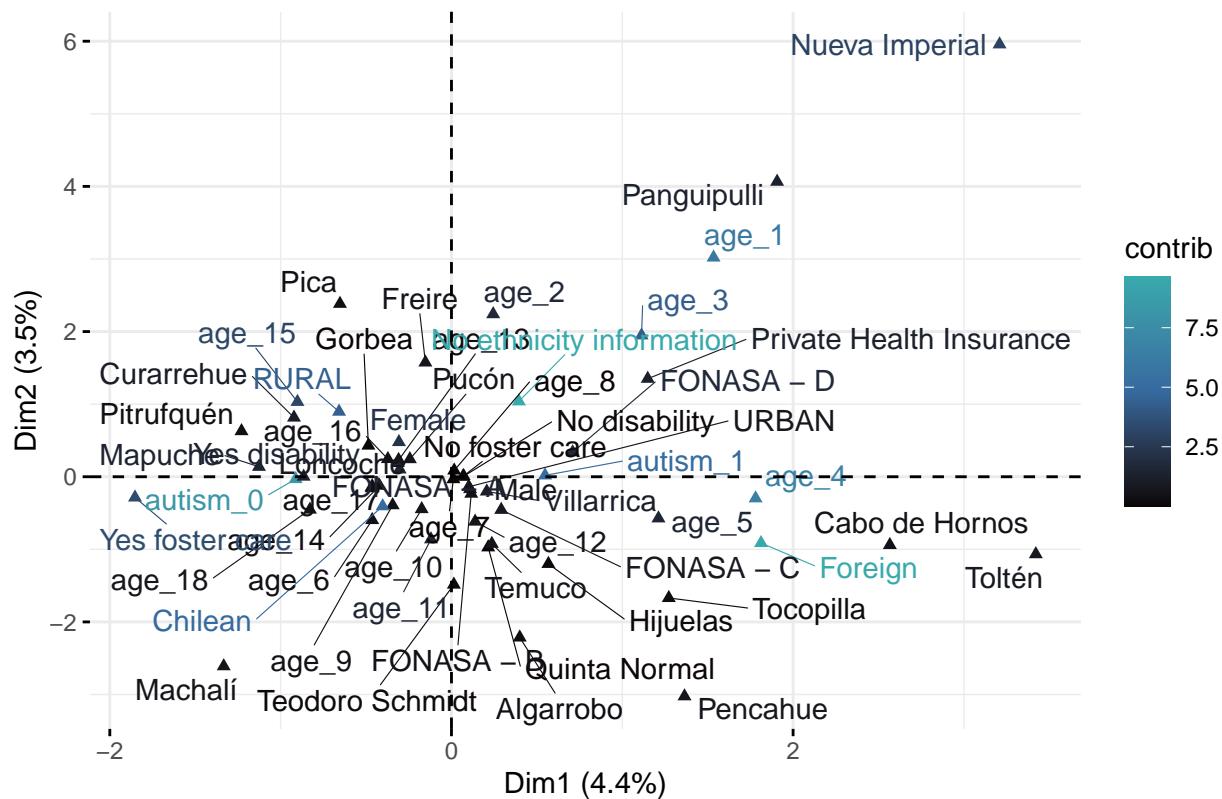


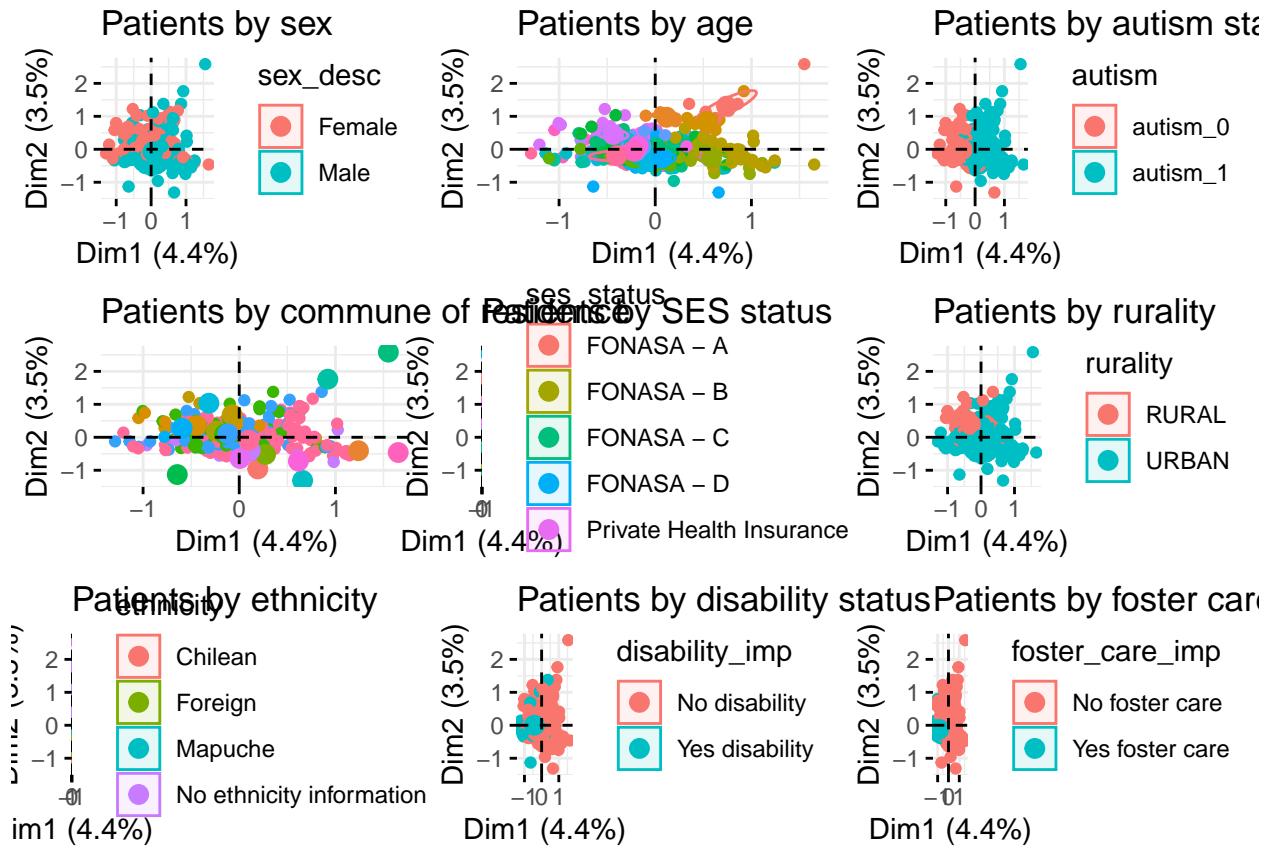


Variables – MCA



Variable categories – MCA





5.5 Probabilistic data linkage

NB: there are 1688 unique ID's in patients and it's 1702 rows long because some people are represented in 2 communes.

5.5.1 Number matched

Using perfect match on sex, date of birth and commune of residence, 197 matches can be found between the school and patient records. 187 unique school records can be perfectly matched to clinical records. 193 patients can be linked to school records.

There are no patients that lived in different communes therefore were in the patient dataset twice that are matched to multiple school records.

5.5.2 Differences between un/matched

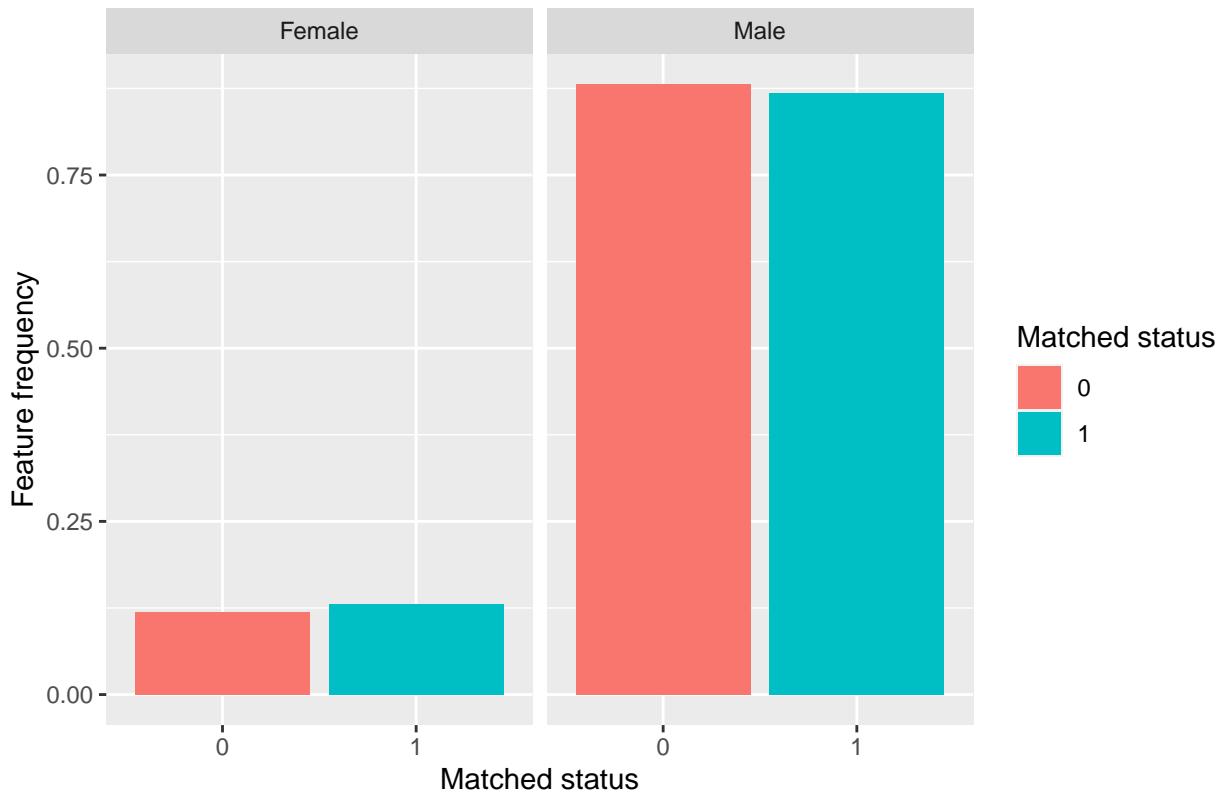
```
##  
## Two-sample Kolmogorov-Smirnov test  
##  
## data: na.omit(school_yes$sex.school) and na.omit(school_no$sex.school)  
## D = 0.012308, p-value = 1  
## alternative hypothesis: two-sided  
  
##  
## Two-sample Kolmogorov-Smirnov test  
##  
## data: as.numeric(na.omit(school_yes$ses_status.school)) and as.numeric(na.omit(school_no$ses_status.school))
```

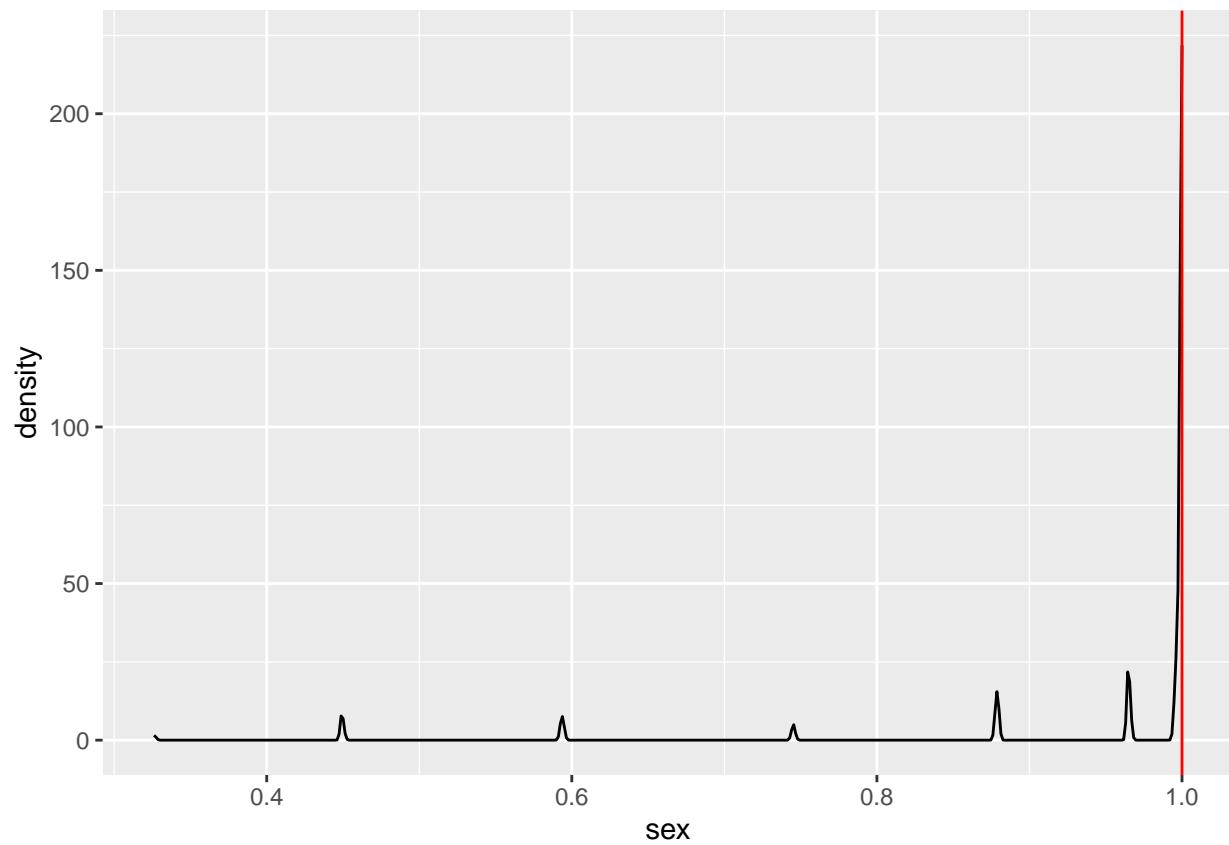
```

## D = 0.094886, p-value = 0.2286
## alternative hypothesis: two-sided
##
## Two-sample Kolmogorov-Smirnov test
##
## data: as.numeric(na.omit(school_yes$commune_code)) and as.numeric(na.omit(school_no$commune_code))
## D = 0.20245, p-value = 9.174e-05
## alternative hypothesis: two-sided

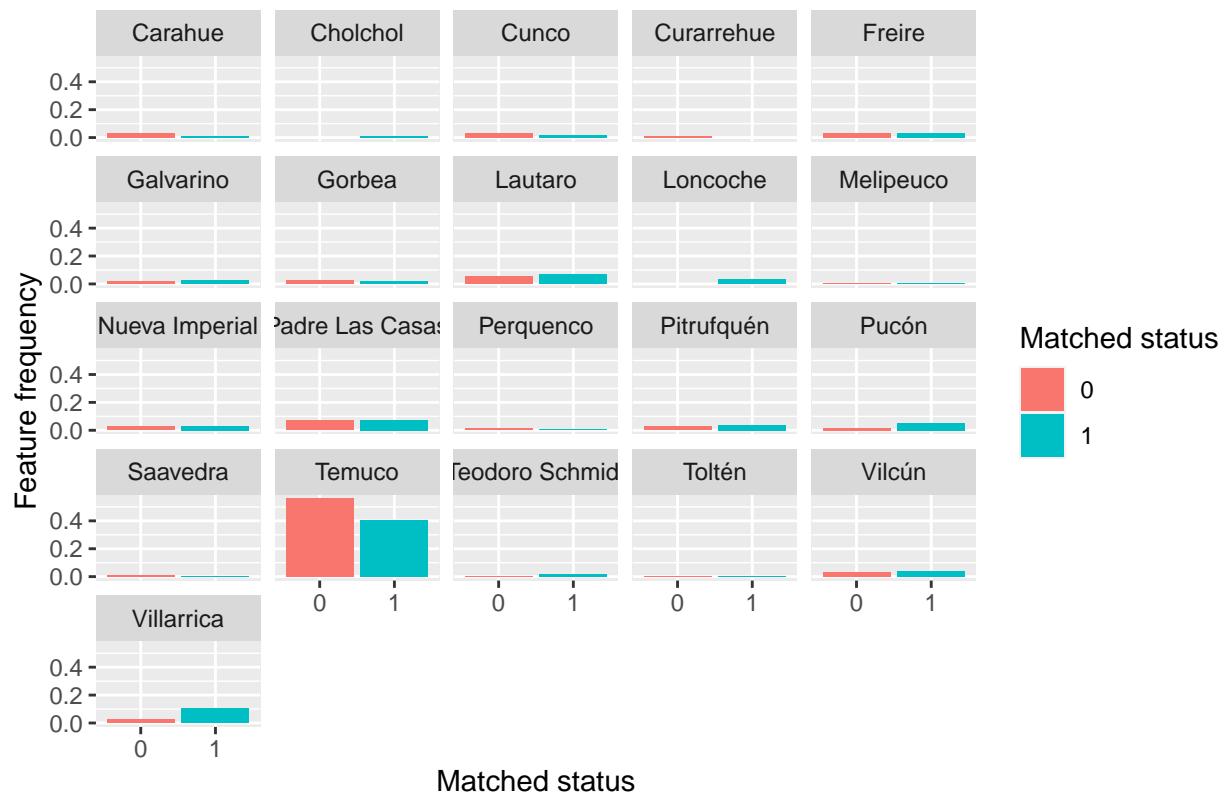
```

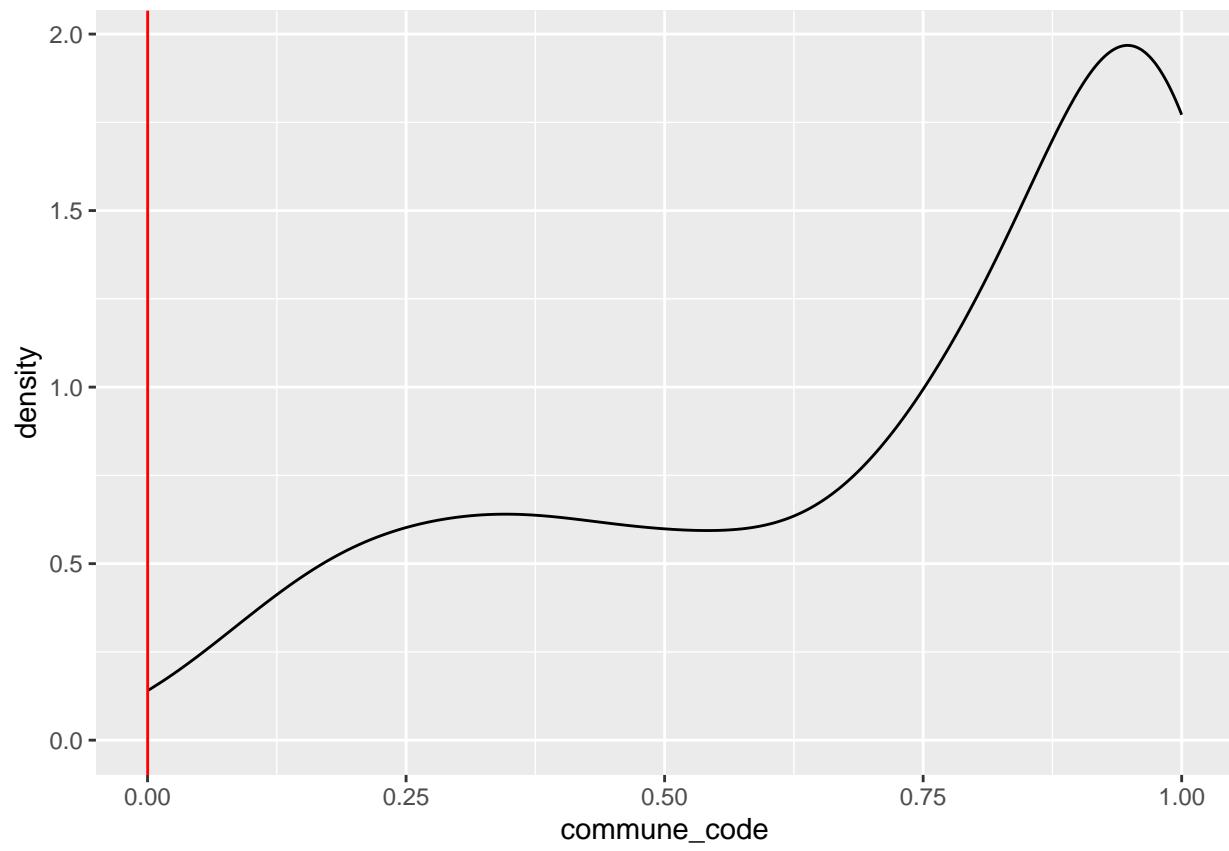
Matching of school record to clinical record by sex



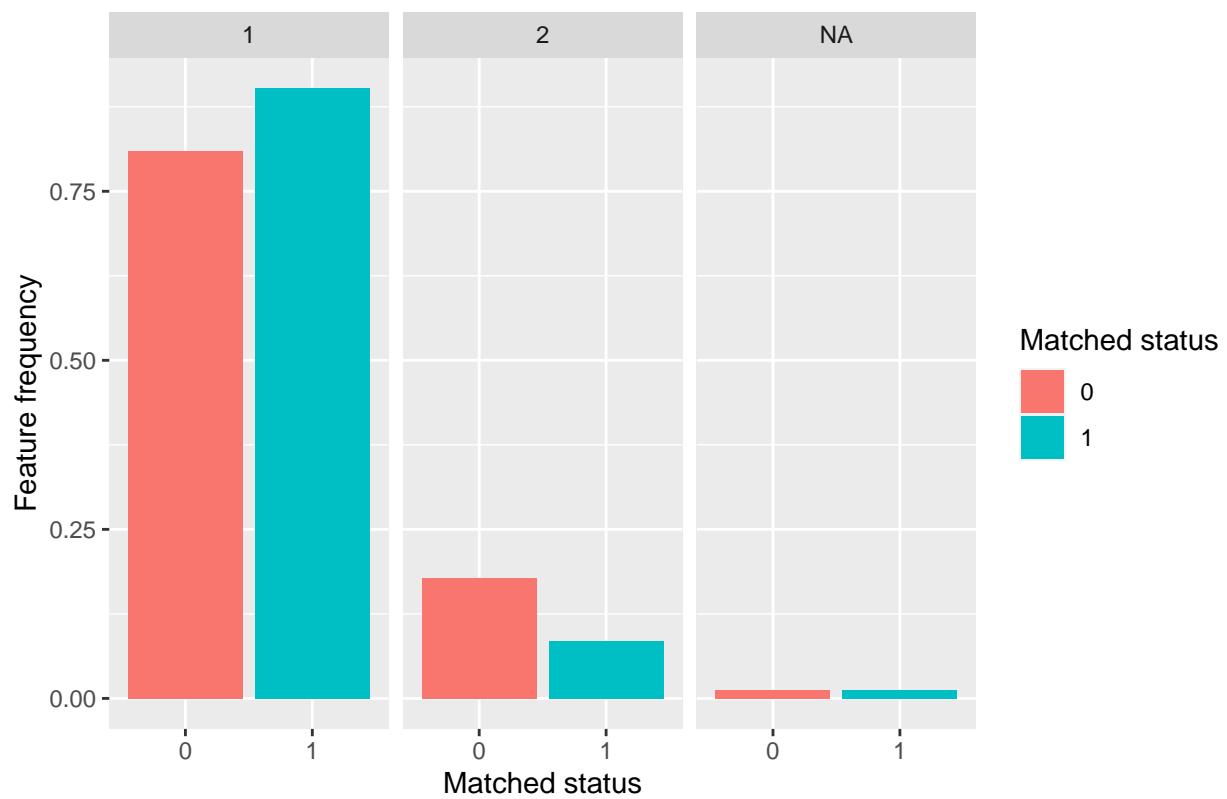


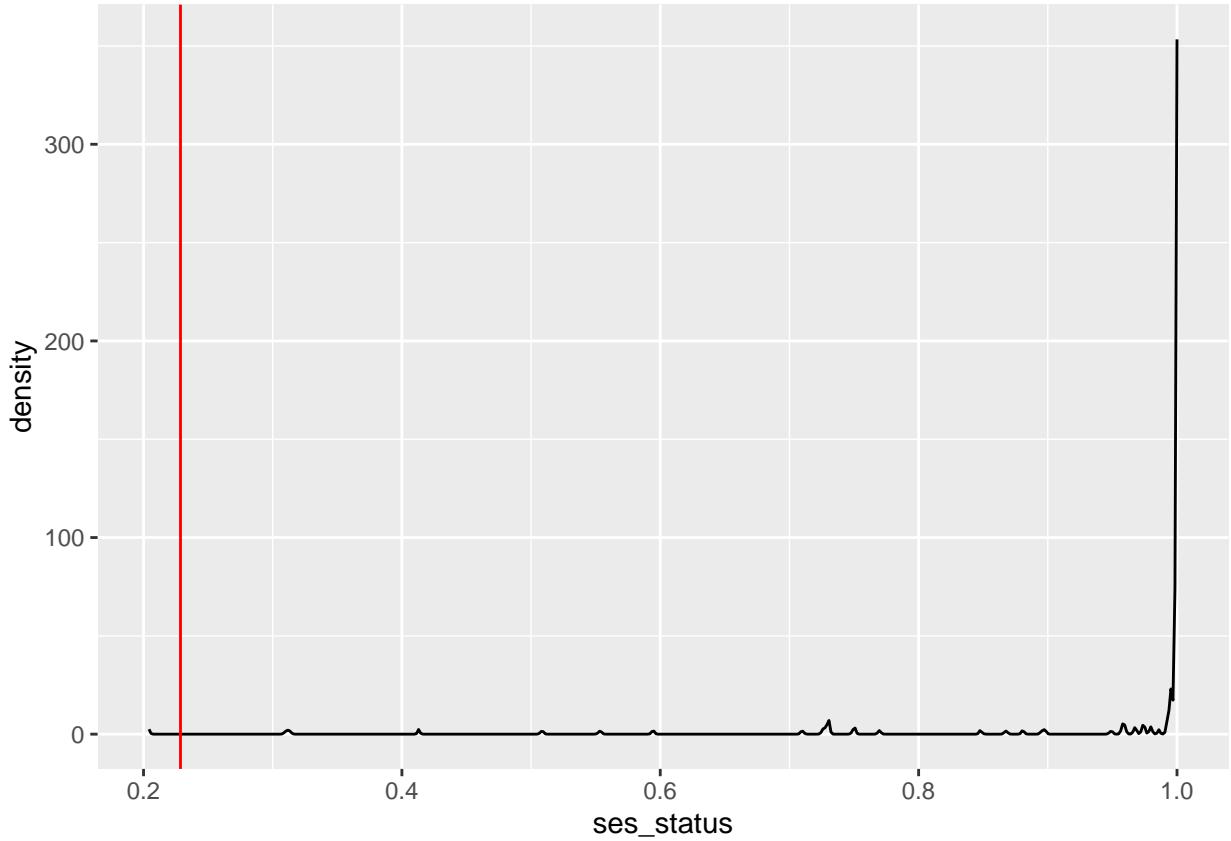
Matching of school record to clinical record by commune





Matching of school record to clinical record by SES status





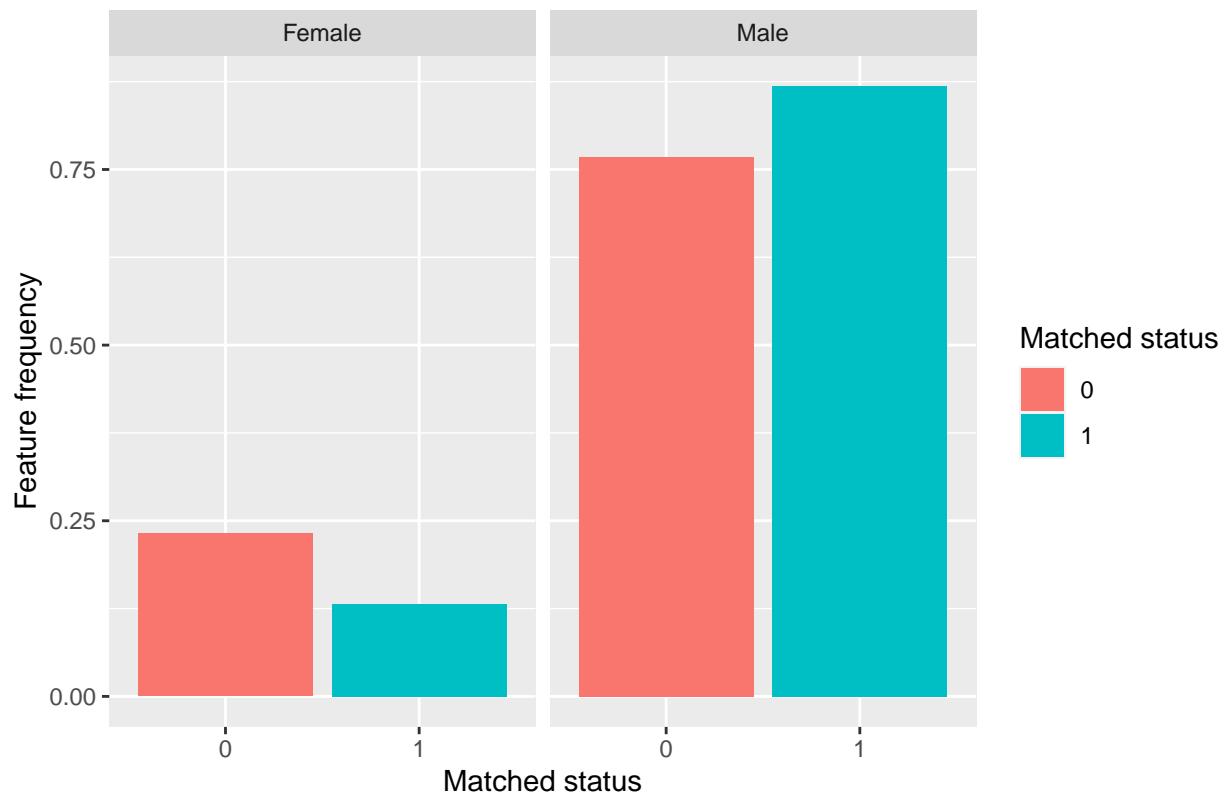
Bit easier to match SES status of 1 (probably more common)

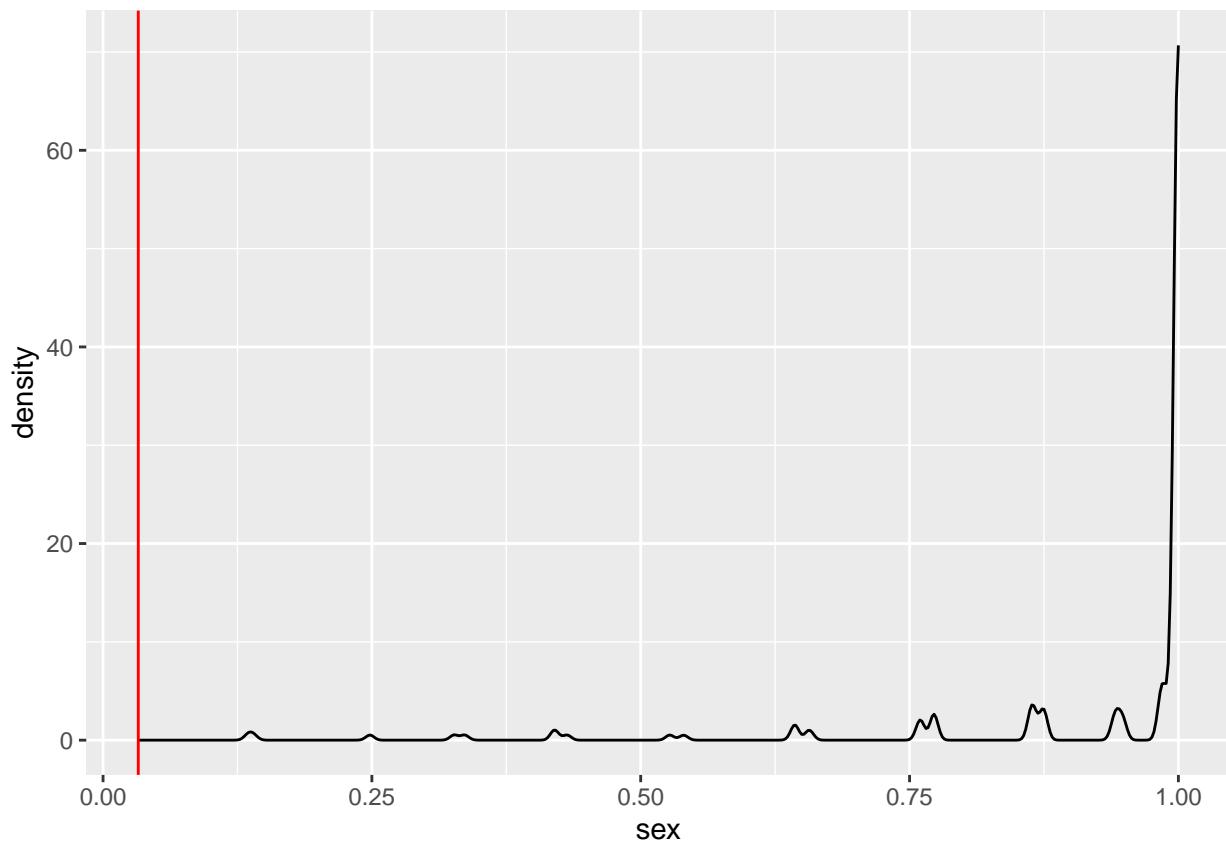
Our matched/non-matched are not different by sex (p-value in Kolmog is same as most of distribution of permuted pvals) but are different by commune and ses status. Cohen's D test isn't suitable to compare the matched and un-matched because the data don't have standard deviations.

??Add commune maps here with size of sample for school and clinical?? Also size of other features.

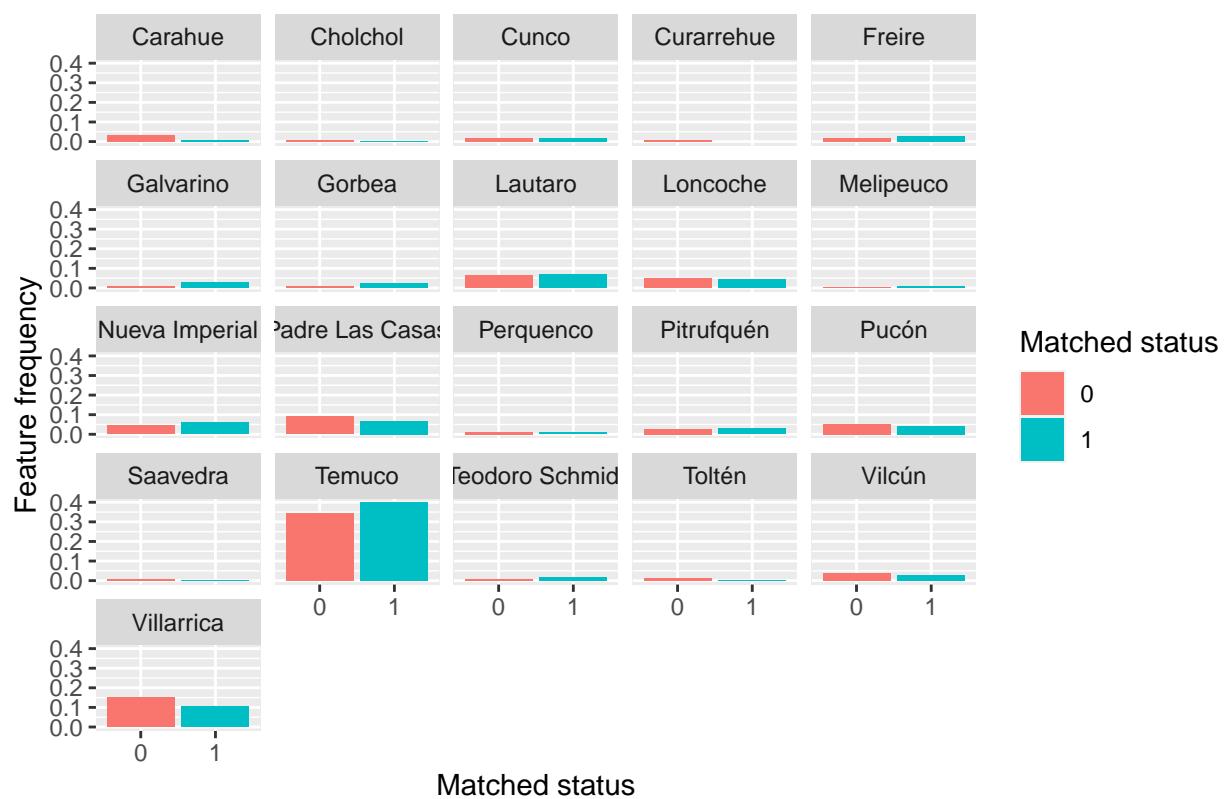
```
##  
## Two-sample Kolmogorov-Smirnov test  
##  
## data: na.omit(patients_yes$sex.patient) and na.omit(patients_no$sex.patient)  
## D = 0.10057, p-value = 0.03276  
## alternative hypothesis: two-sided  
  
##  
## Two-sample Kolmogorov-Smirnov test  
##  
## data: as.numeric(na.omit(patients_yes$ses_status.patient)) and as.numeric(na.omit(patients_no$ses_s...  
## D = 0.067733, p-value = 0.3086  
## alternative hypothesis: two-sided  
  
##  
## Two-sample Kolmogorov-Smirnov test  
##  
## data: as.numeric(na.omit(patients_yes$commune_code)) and as.numeric(na.omit(patients_no$commune_code))  
## D = 0.084804, p-value = 0.1074  
## alternative hypothesis: two-sided
```

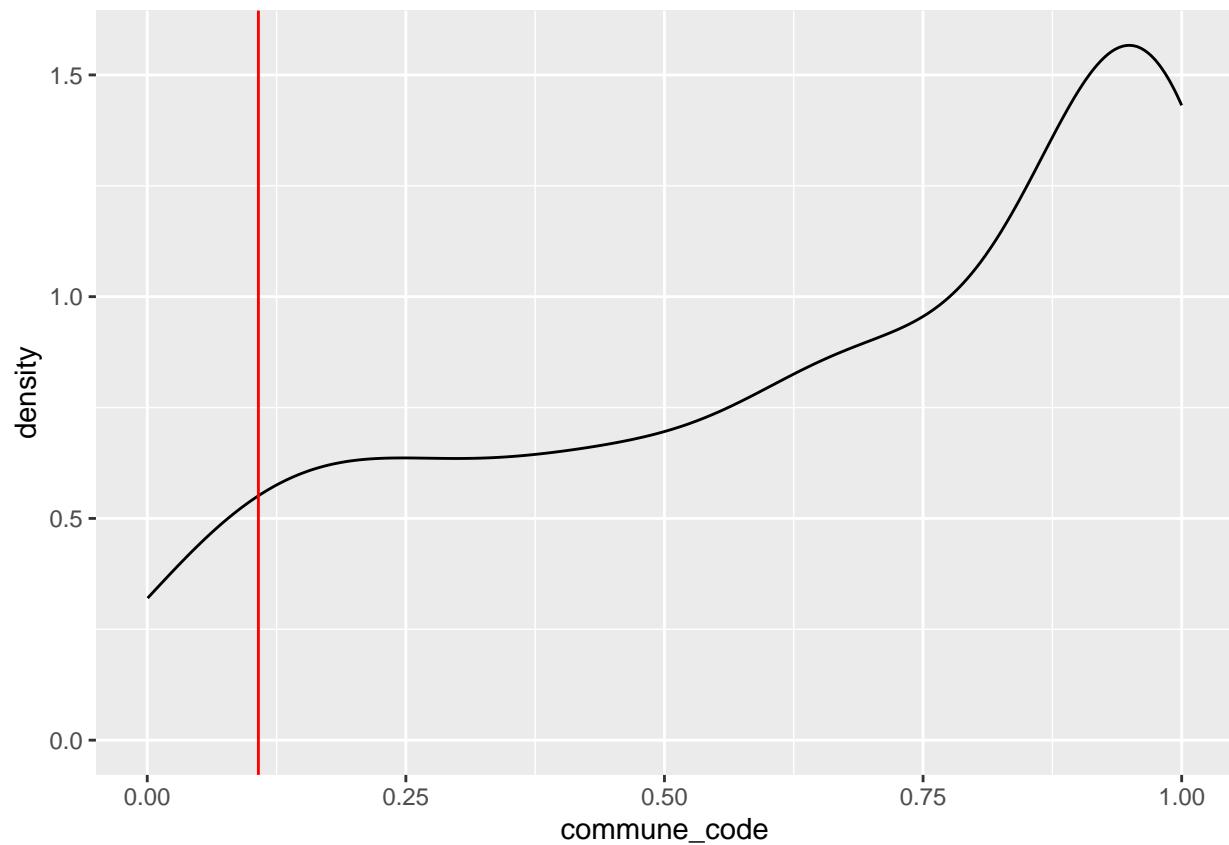
Matching of clinical record to school record by sex



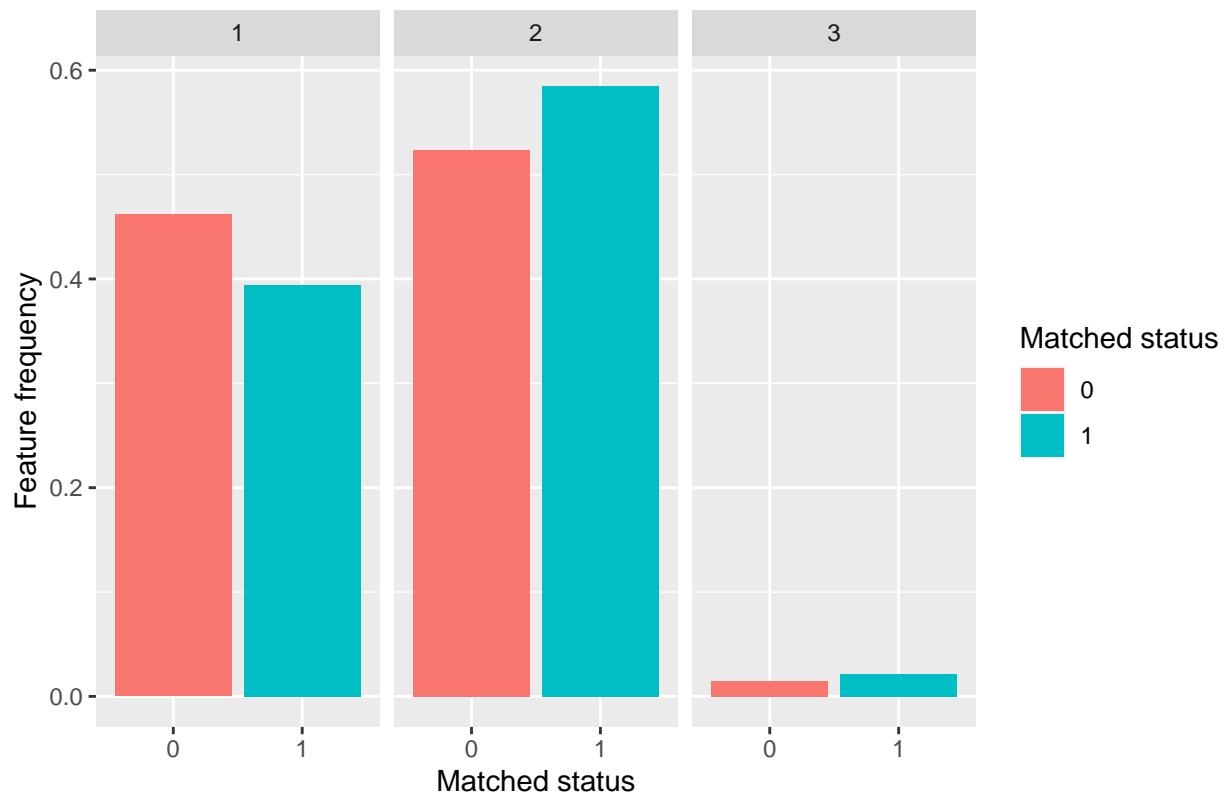


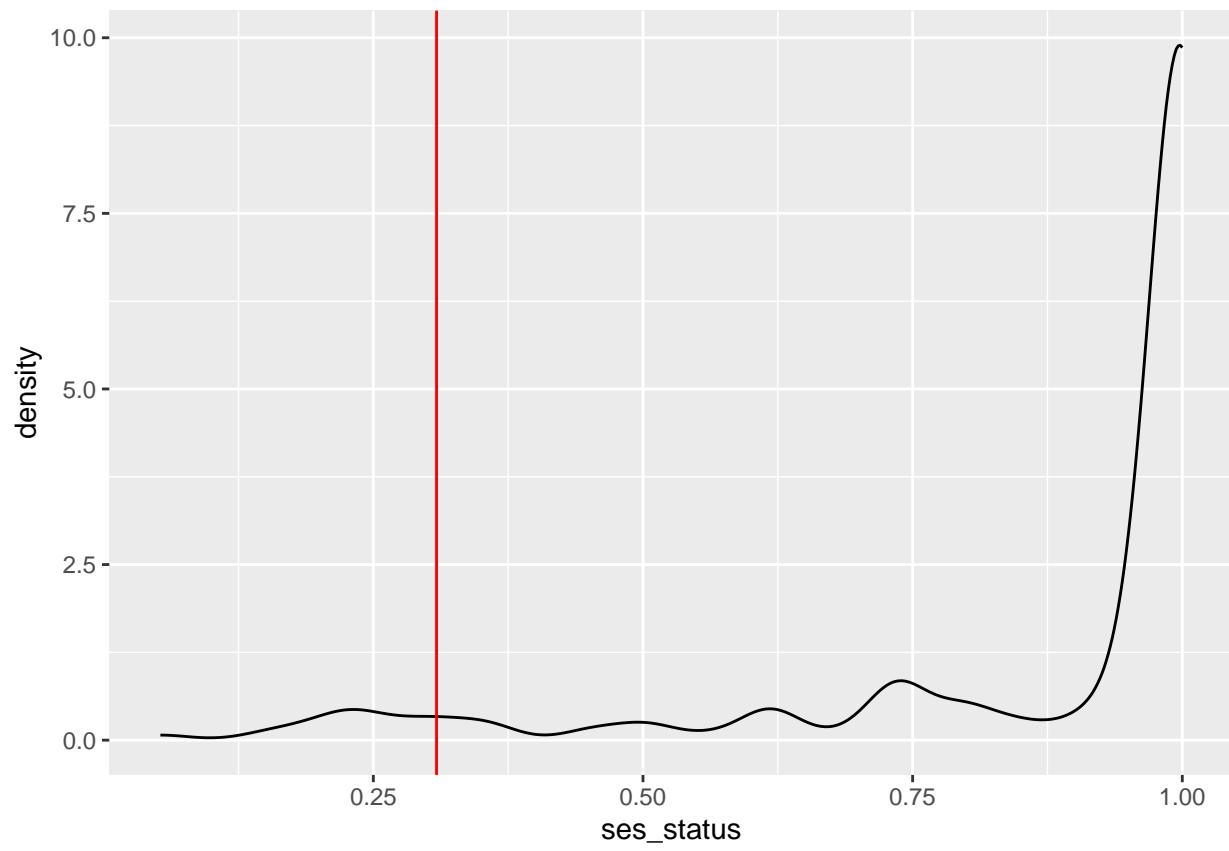
Matching of clinical record to school record by commune





Matching of clinical record to school record by SES status



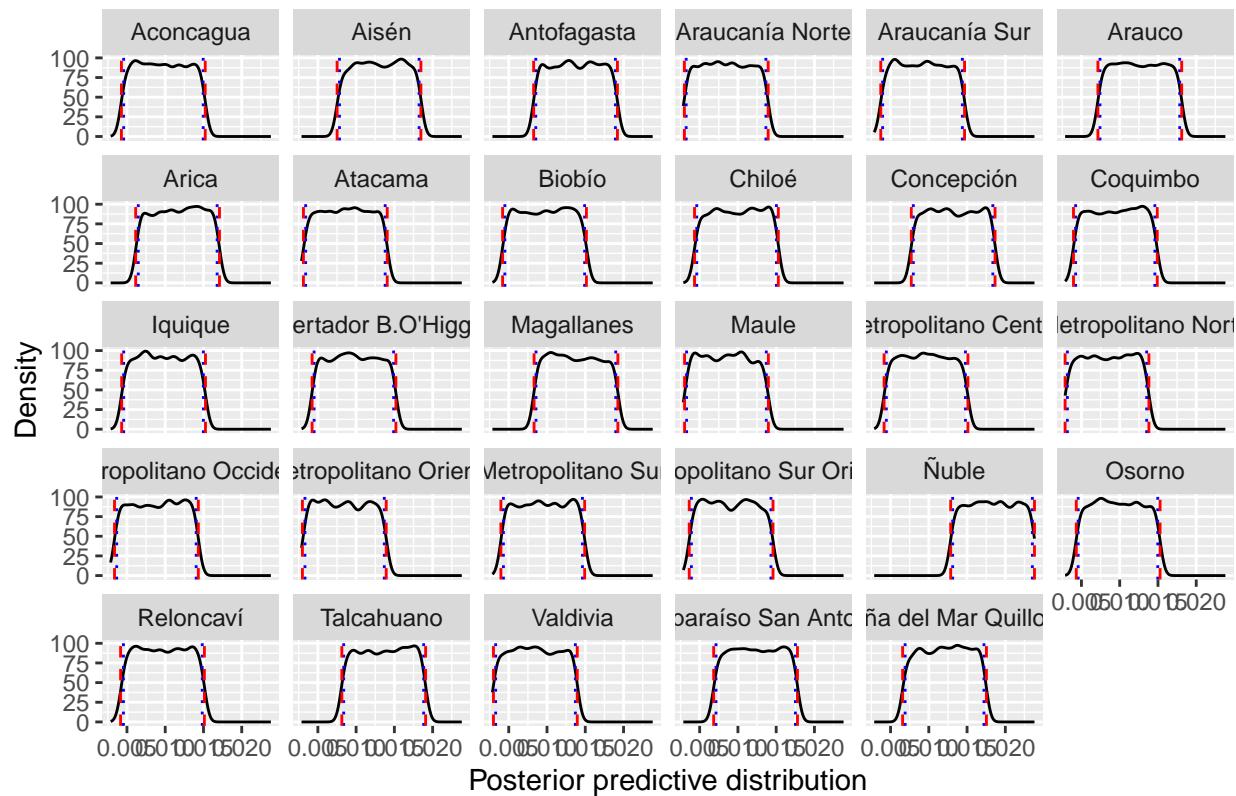


TODO - comment on whether any of the matched patients had only intellectual disability and not autism.

5.5.3 Prev delta

5.6 Bayesian prevalence projection

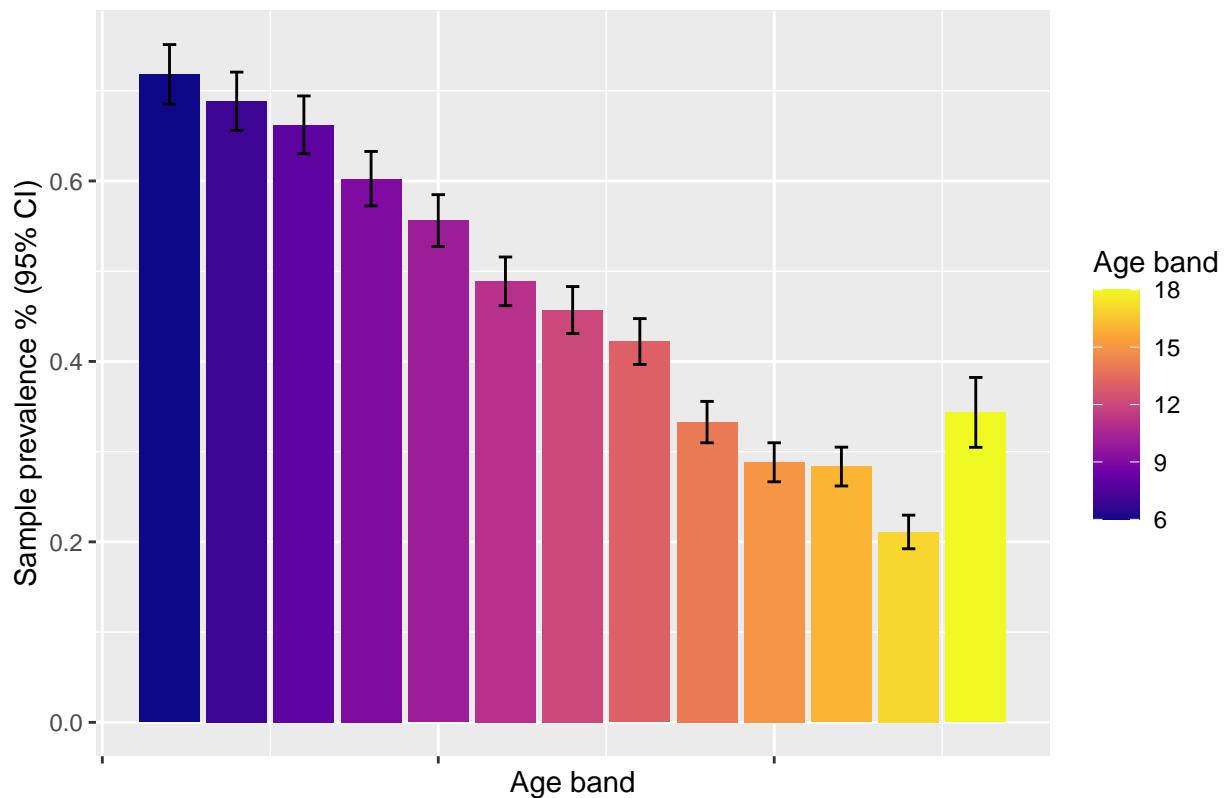
Region specific priors informed by clinical data from Araucanía Sur



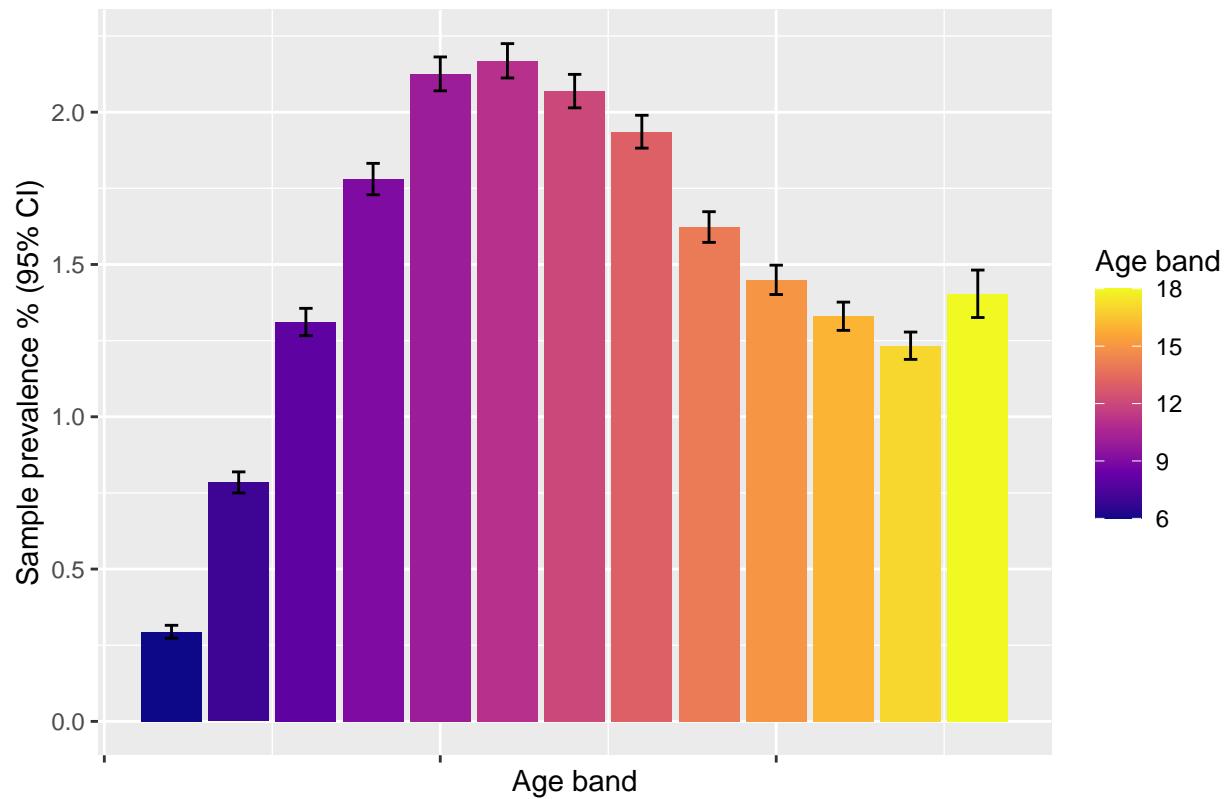
Red is the prior 95% CI (adjusted sample rate and it + prev_delta), blue is posterior 95% CrI

6 Supplementary materials

Autism prevalence



ADHD prevalence



6.1 Bayesian prevalence estimation

- 6.1.1 Random effect on sex
- 6.1.2 Random effect on health service
- 6.1.3 Random effect on commune in Araucanía Sur health service
- 6.1.4 Random effect on socio-economic status
- 6.1.5 Random effect on ethnicity
- 6.1.6 Random effect on school's rurality

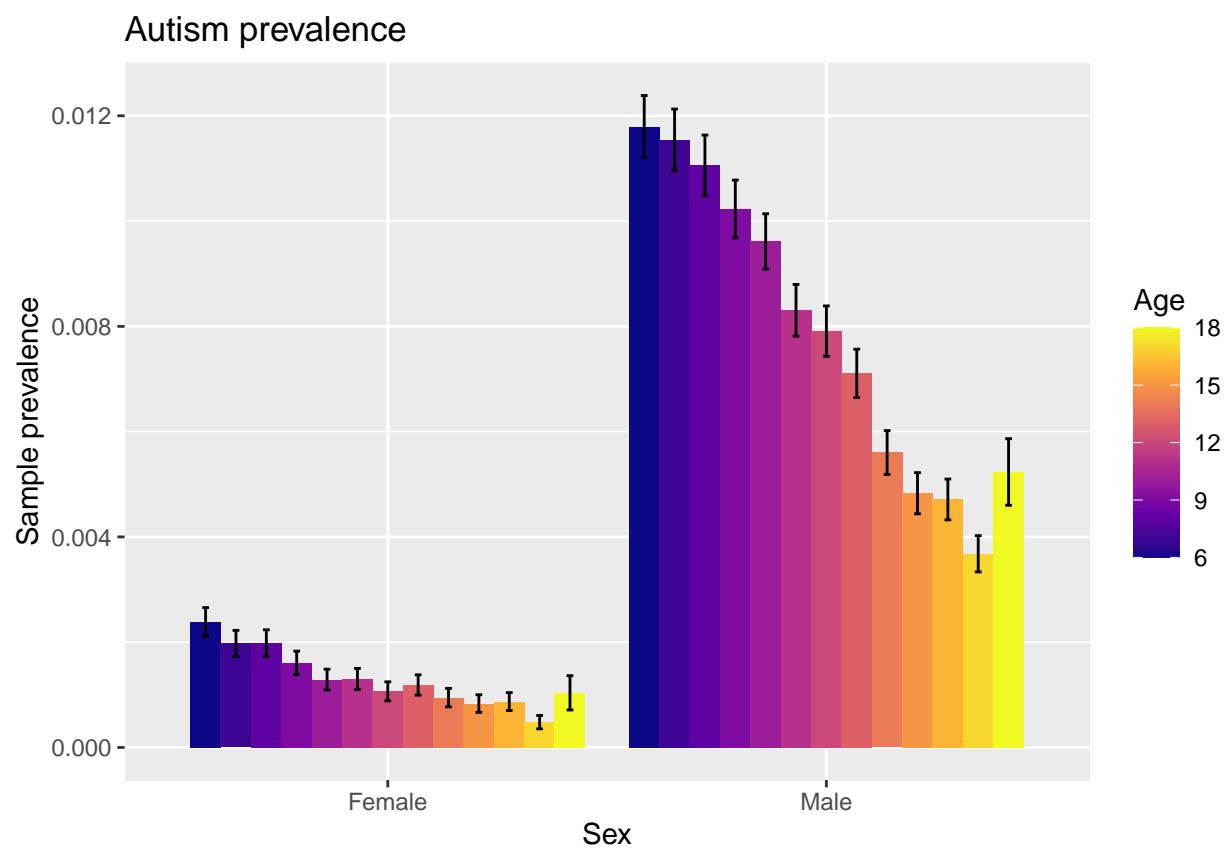


Figure 28: Sample prevalence of autism by age and sex. Bars show 95% normal confidence intervals.

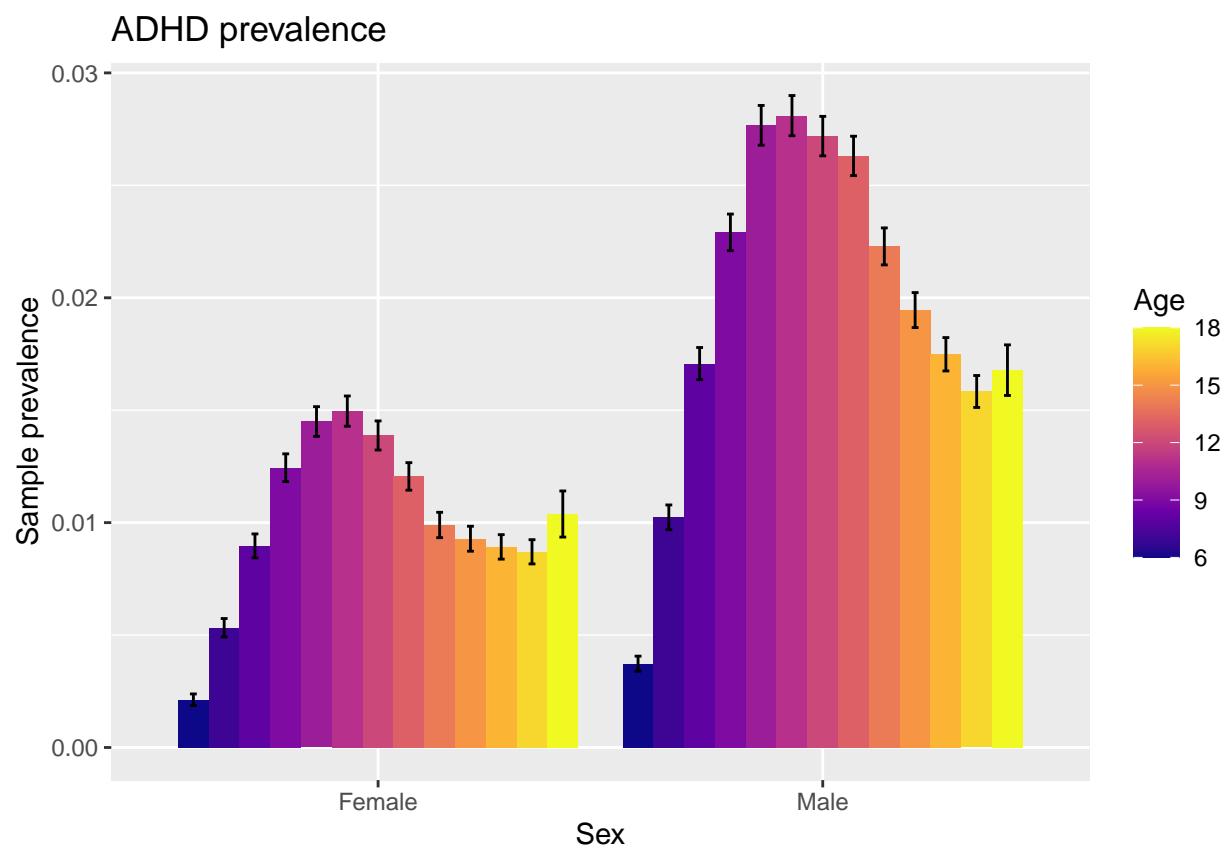


Figure 29: Sample prevalence of ADHD by age and sex. Bars show 95% normal confidence intervals.

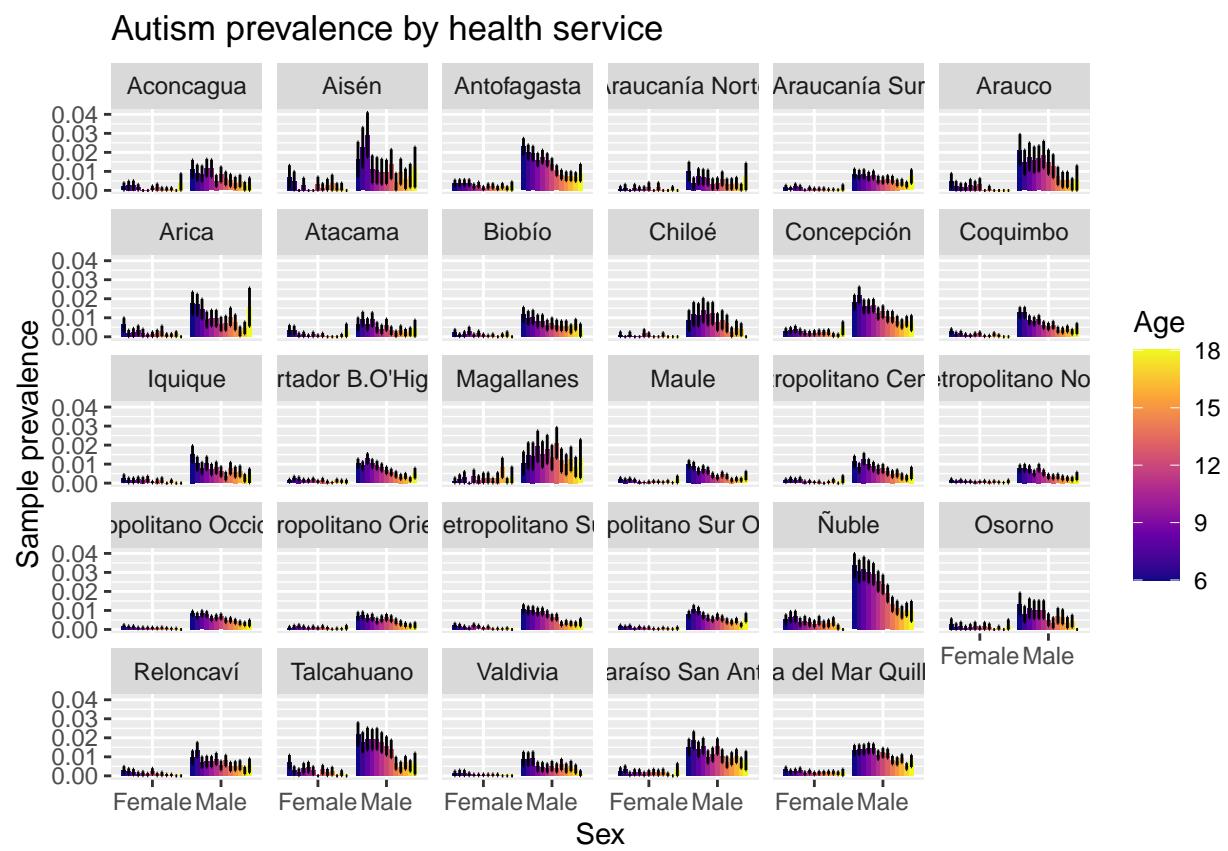


Figure 30: Sample prevalence of autism by health service, age and sex. Bars show 95% normal confidence intervals.

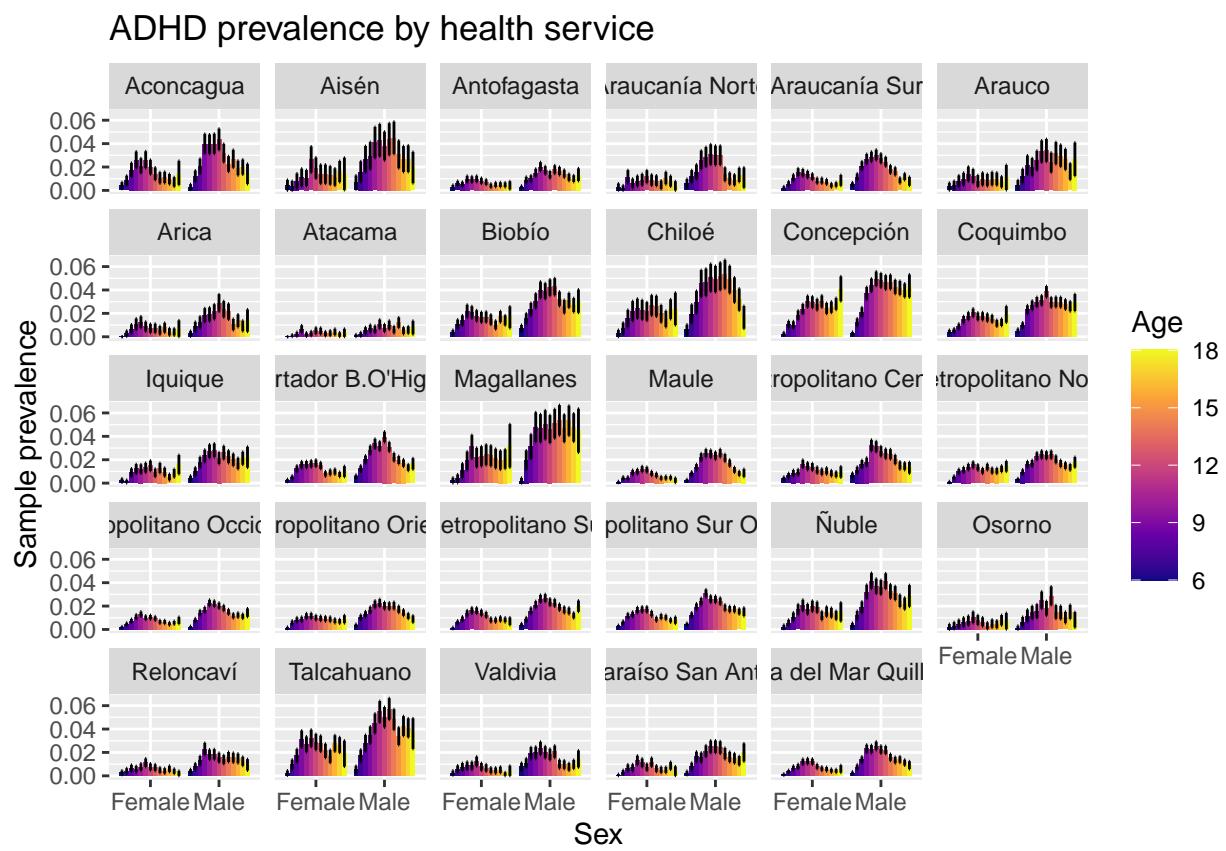


Figure 31: Sample prevalence of ADHD by health service, age and sex. Bars show 95% normal confidence intervals.

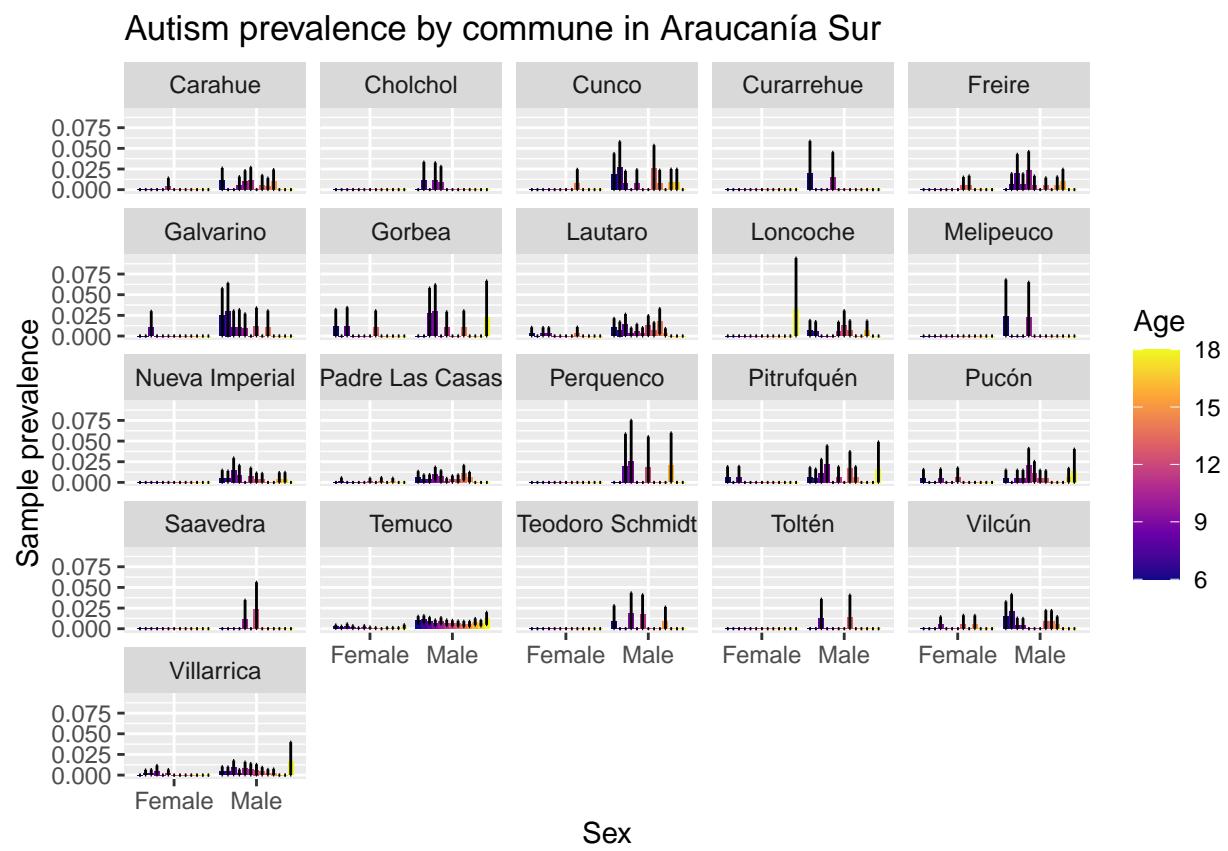


Figure 32: Sample prevalence of autism by commune in Araucanía Sur health service, age and sex. Bars show 95% normal confidence intervals.

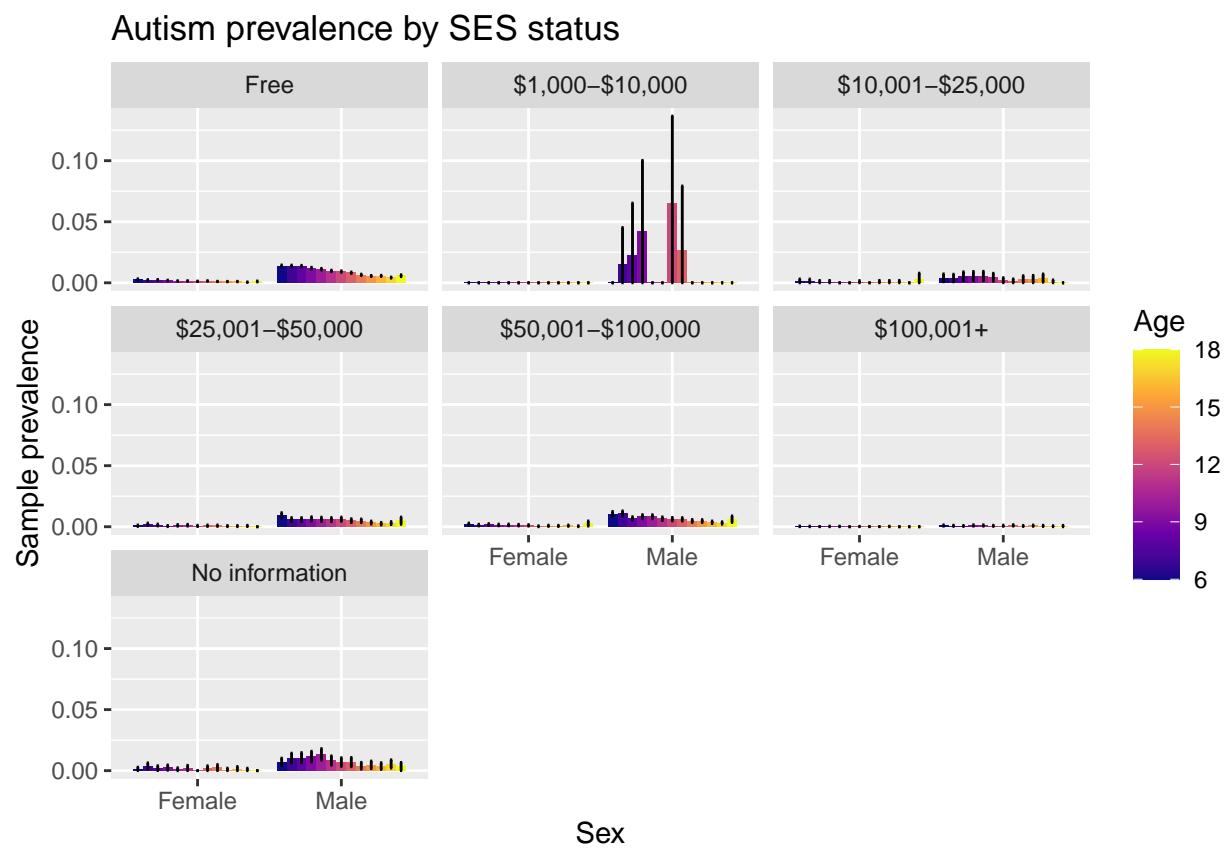


Figure 33: Sample prevalence of autism by socio-economic (SES) status of student's family, age and sex. Bars show 95% normal confidence intervals.

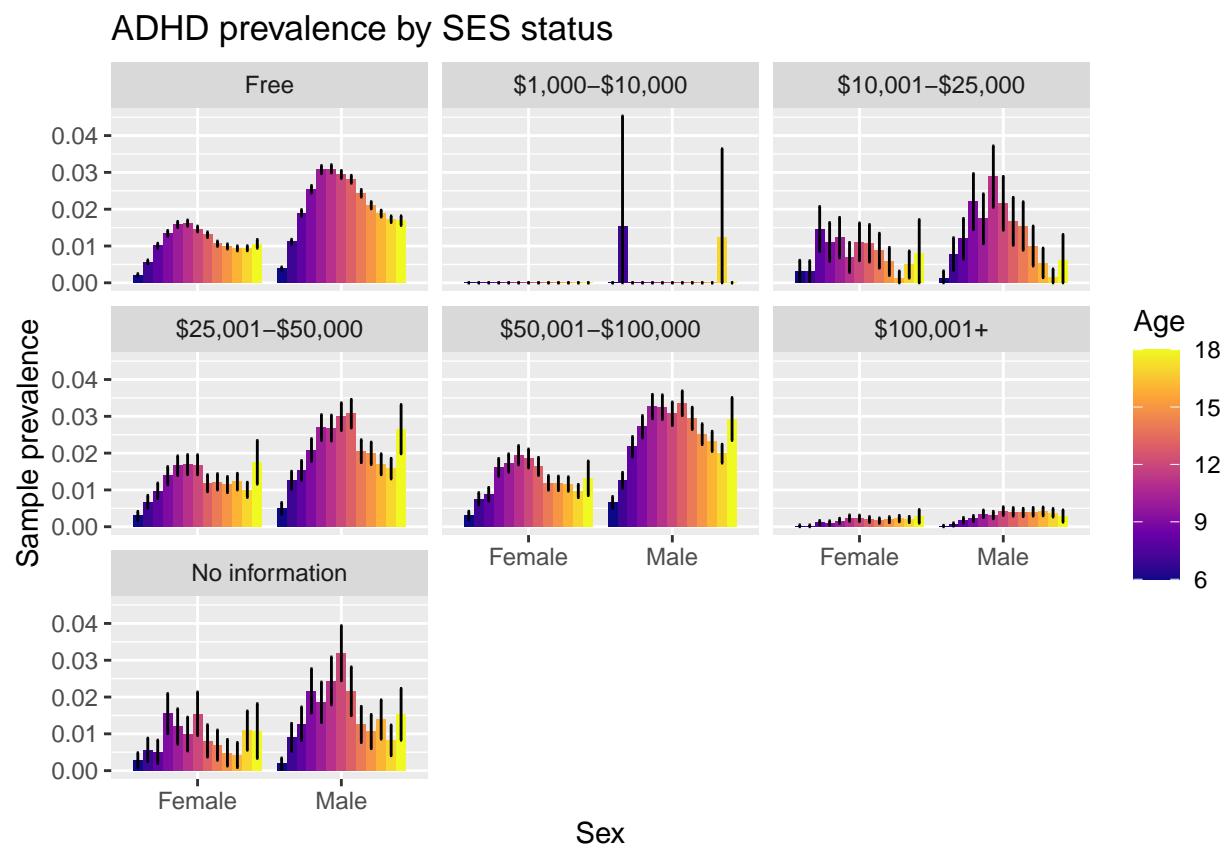


Figure 34: Sample prevalence of ADHD by socio-economic (SES) status of student's family, age and sex. Bars show 95% normal confidence intervals.

Autism prevalence by ethnicity

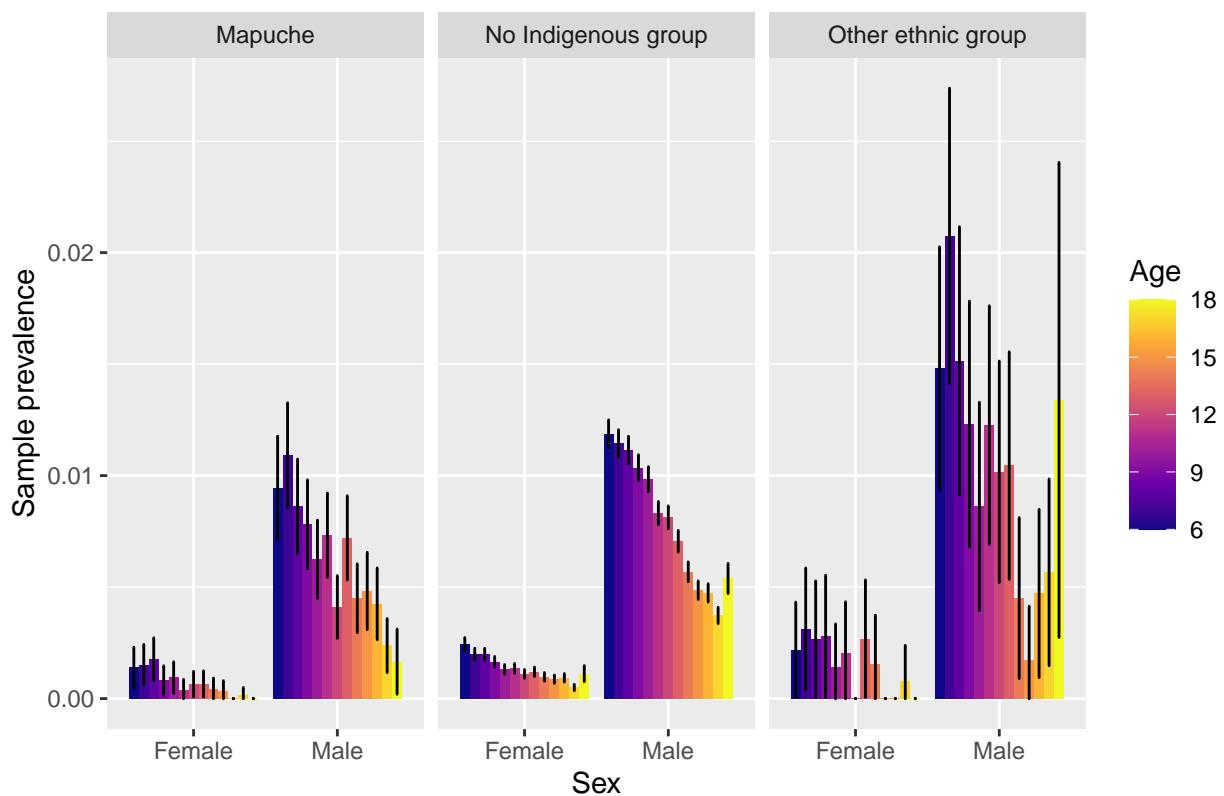


Figure 35: Sample prevalence of autism by ethnicity, age and sex. Bars show 95% normal confidence intervals.

ADHD prevalence by ethnicity

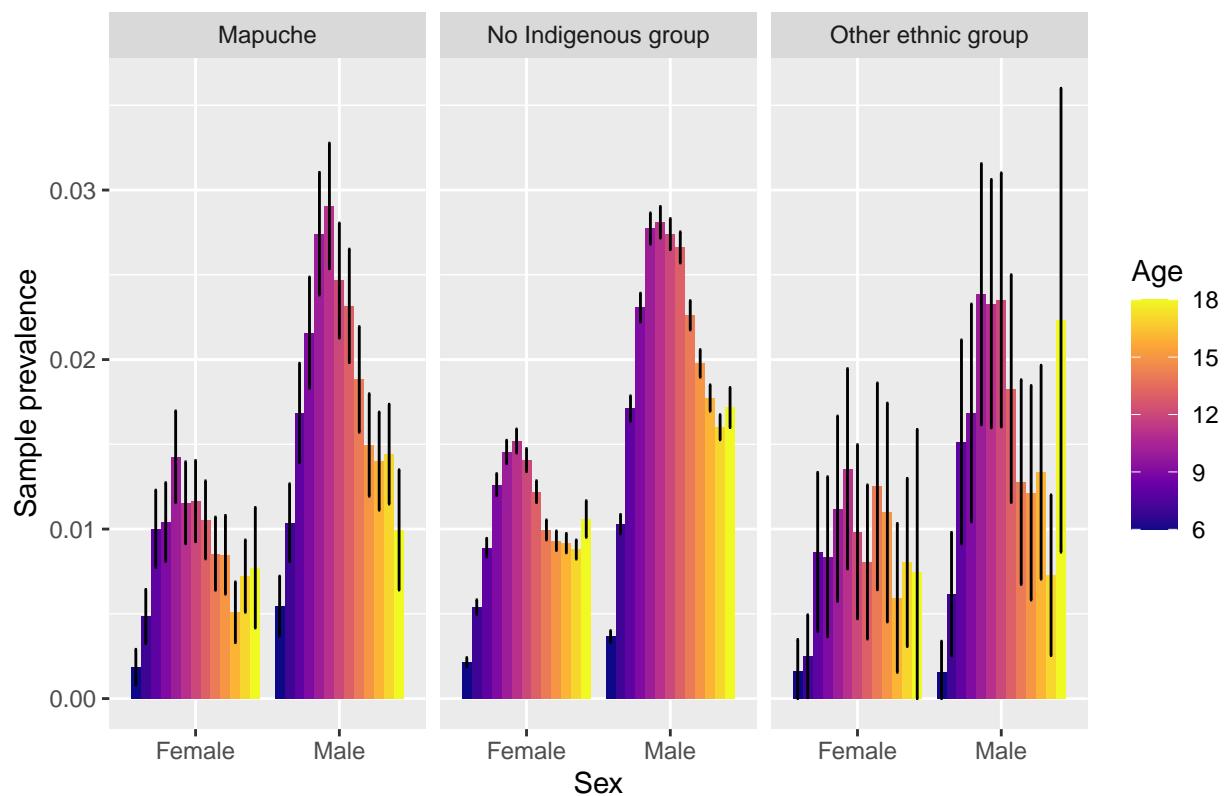


Figure 36: Sample prevalence of ADHD by ethnicity, age and sex. Bars show 95% normal confidence intervals.

Autism prevalence by school's rurality

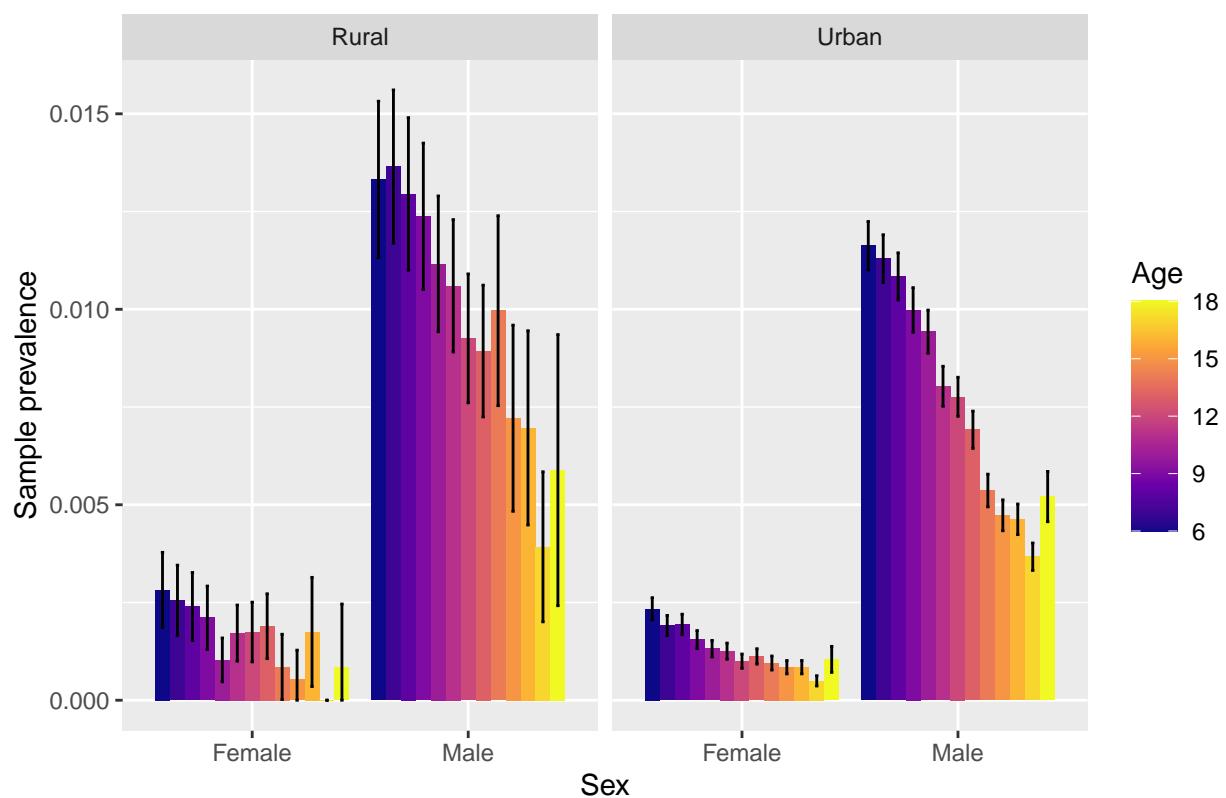


Figure 37: Sample prevalence of autism by school's rurality, age and sex. Bars show 95% normal confidence intervals.

ADHD prevalence by school's rurality

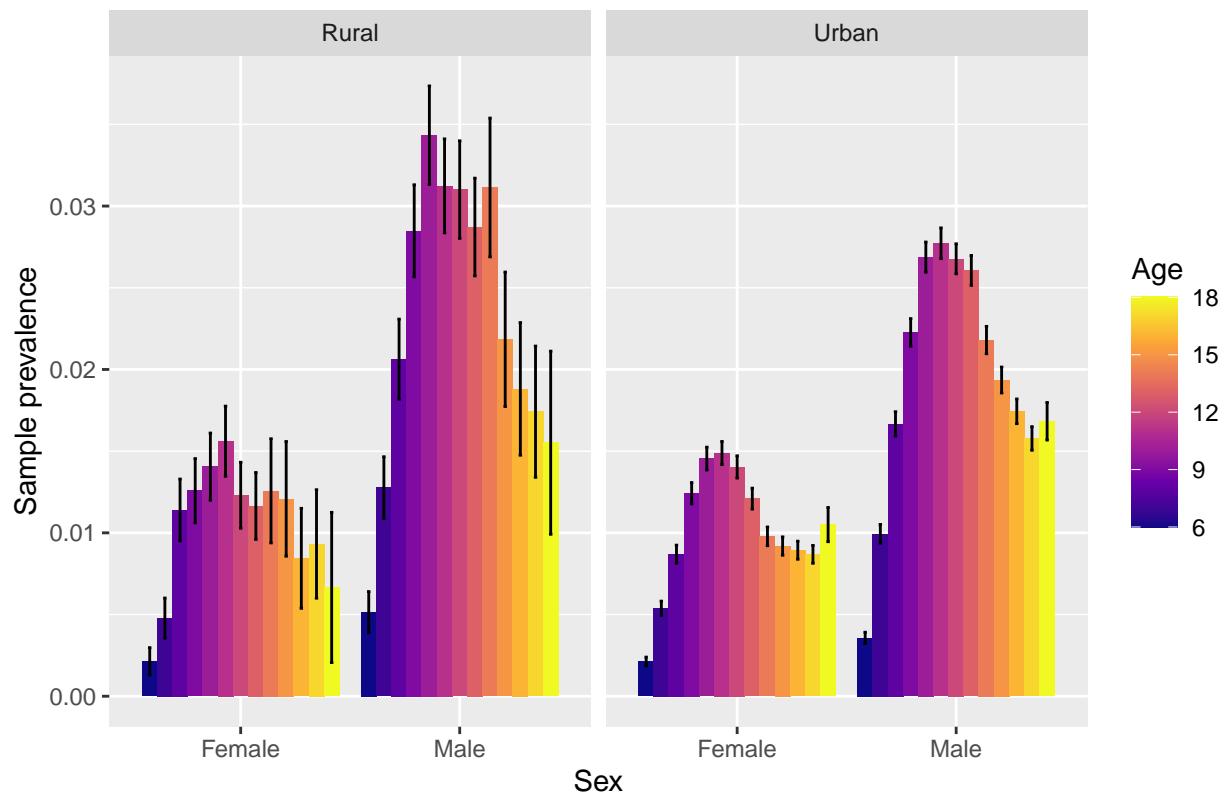


Figure 38: Sample prevalence of ADHD by school's rurality, age and sex. Bars show 95% normal confidence intervals.

Autism prevalence, prior mean = 0.00465, prior sd = 3.98e-05

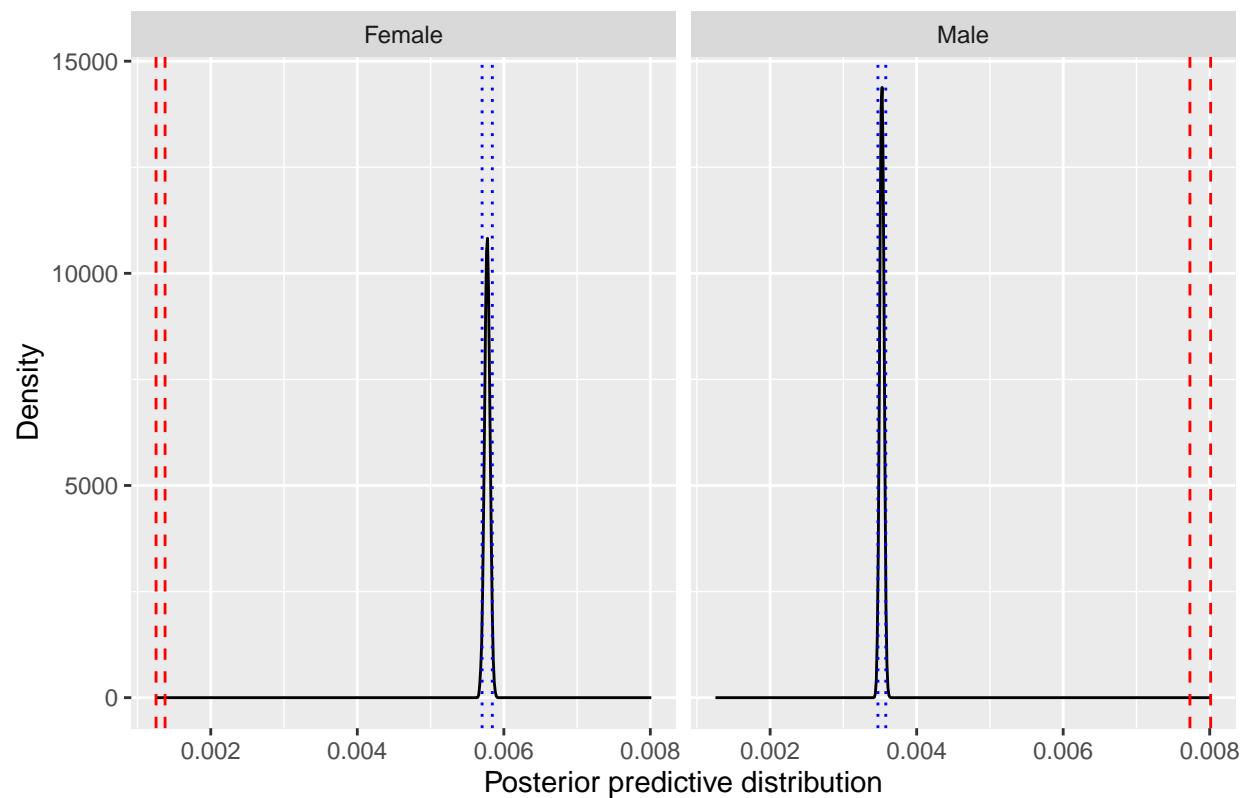


Figure 39: Posterior predictive distribution for autism with a random effect on sex, and with age- and sex-adjusted global prevalence prior.

ADHD prevalence, prior mean = 0.015, prior sd = 7.25e-05

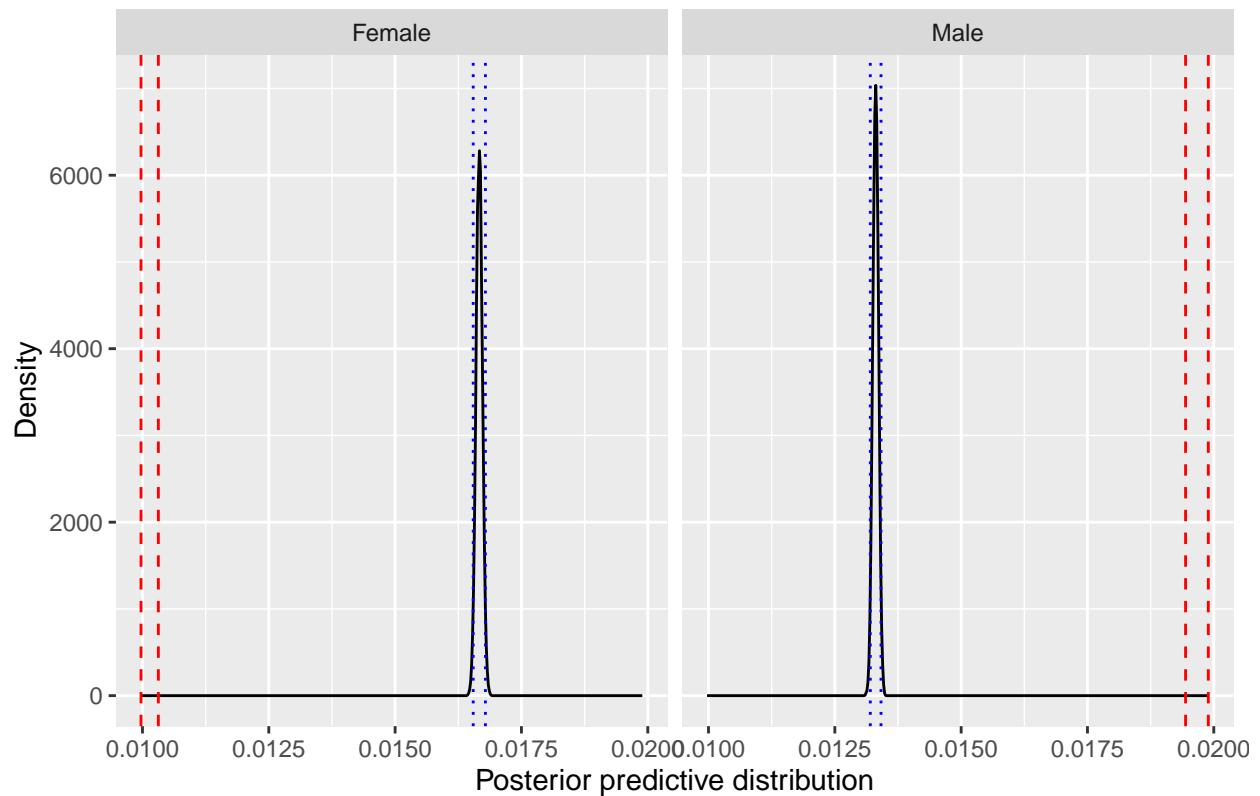


Figure 40: Posterior predictive distribution for ADHD with a random effect on sex, and with age- and sex-adjusted global prevalence prior.

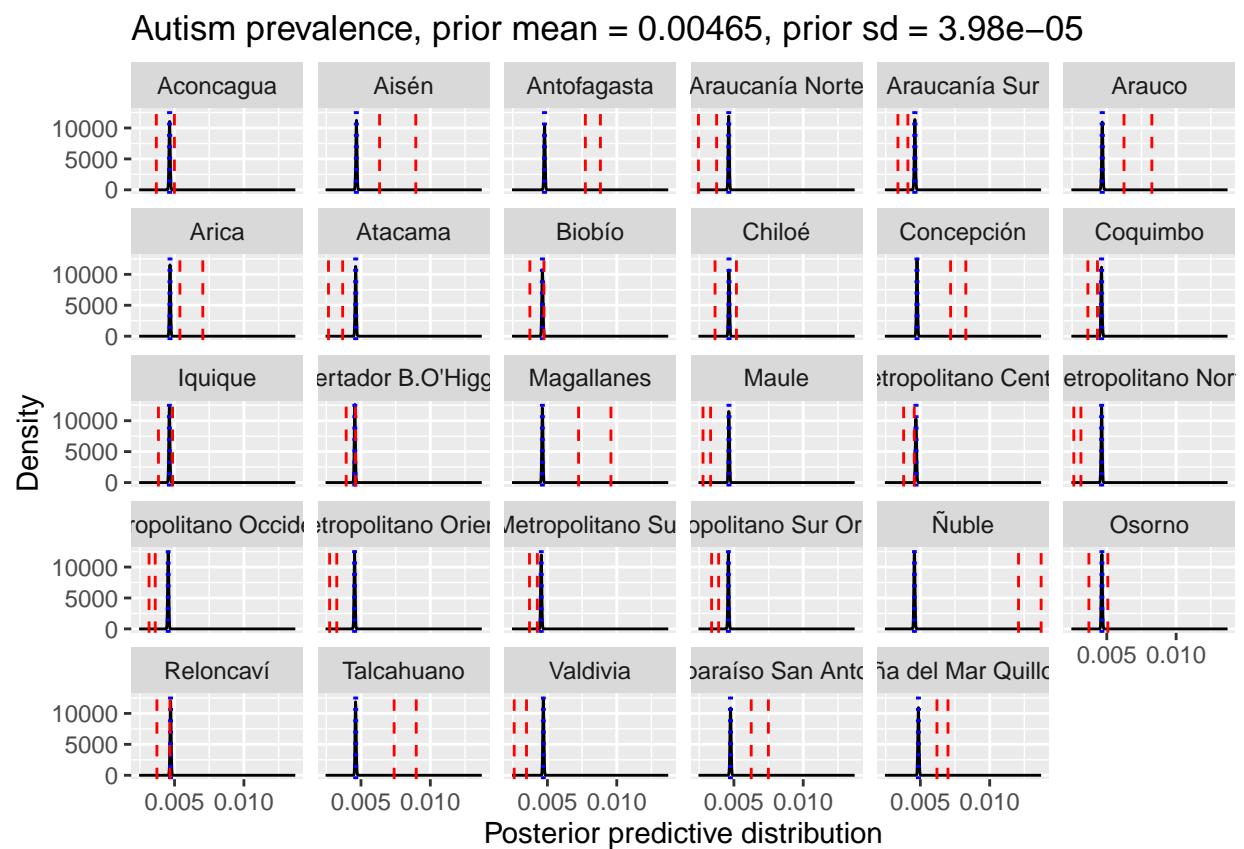


Figure 41: Posterior predictive distribution for autism with a random effect on student's health service, and with age- and sex-adjusted global prevalence prior.

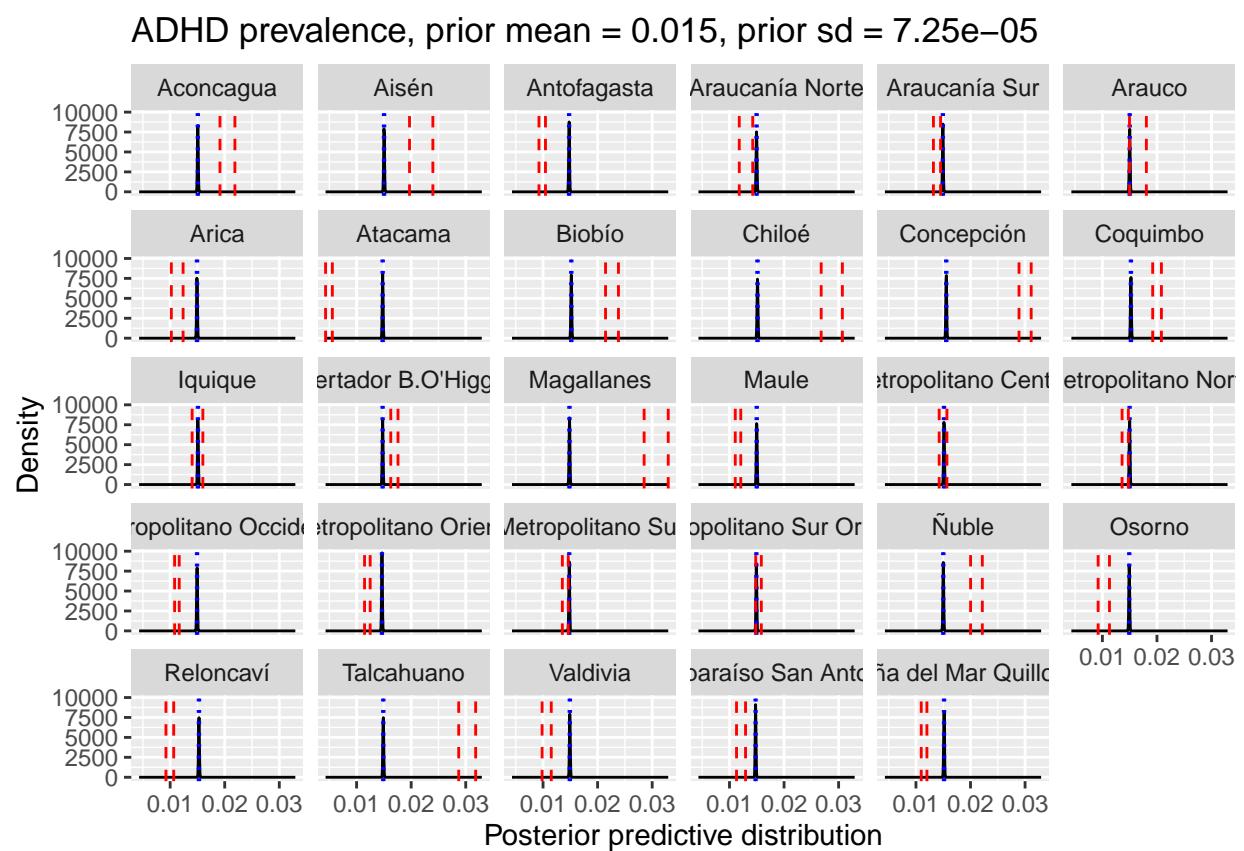


Figure 42: Posterior predictive distribution for ADHD with a random effect on health service, and with age- and sex-adjusted global prevalence prior.

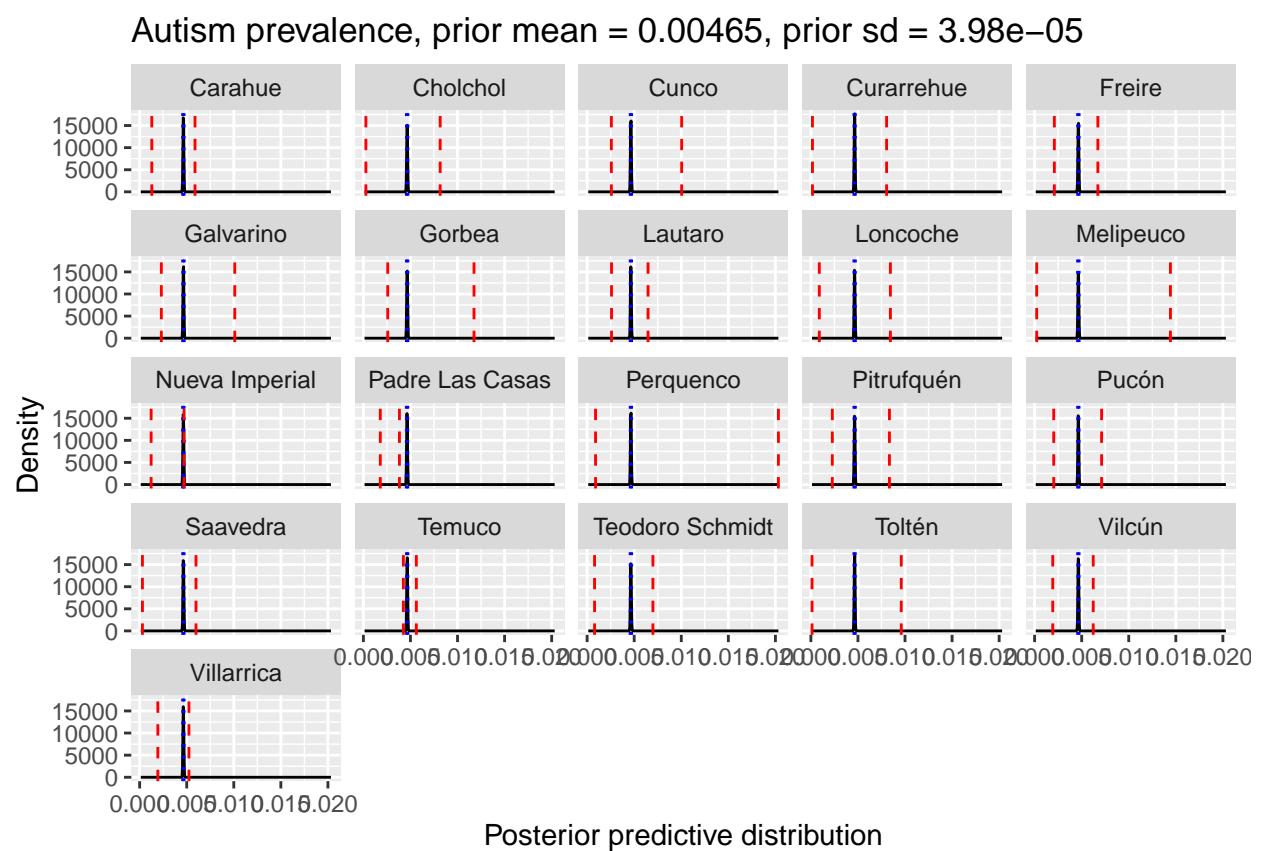


Figure 43: Posterior predictive distribution for autism with a random effect on commune within Araucanía Sur health service, and with age- and sex-adjusted global prevalence prior.

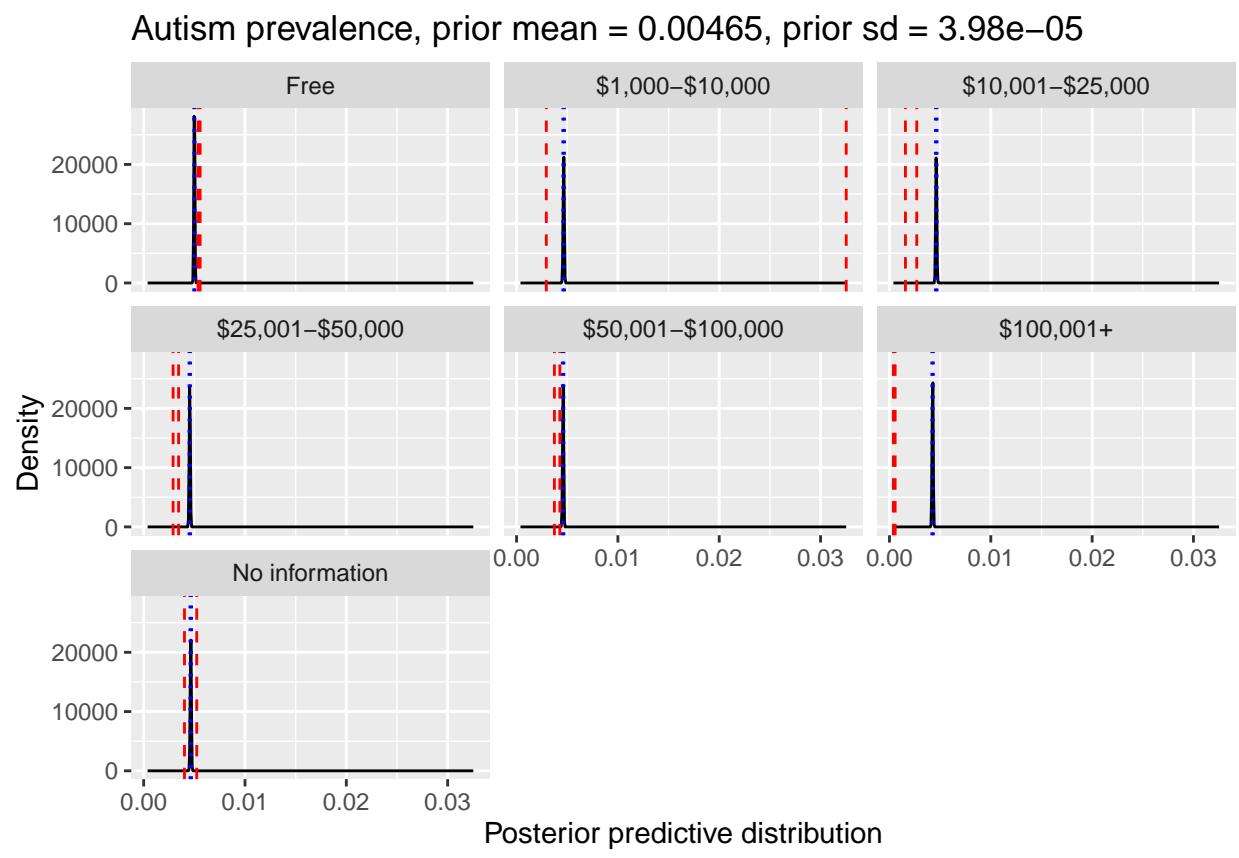


Figure 44: Posterior predictive distribution for autism with a random effect on socio-economic status of student's family, and with age- and sex-adjusted global prevalence prior.

ADHD prevalence, prior mean = 0.015, prior sd = 7.25e-05

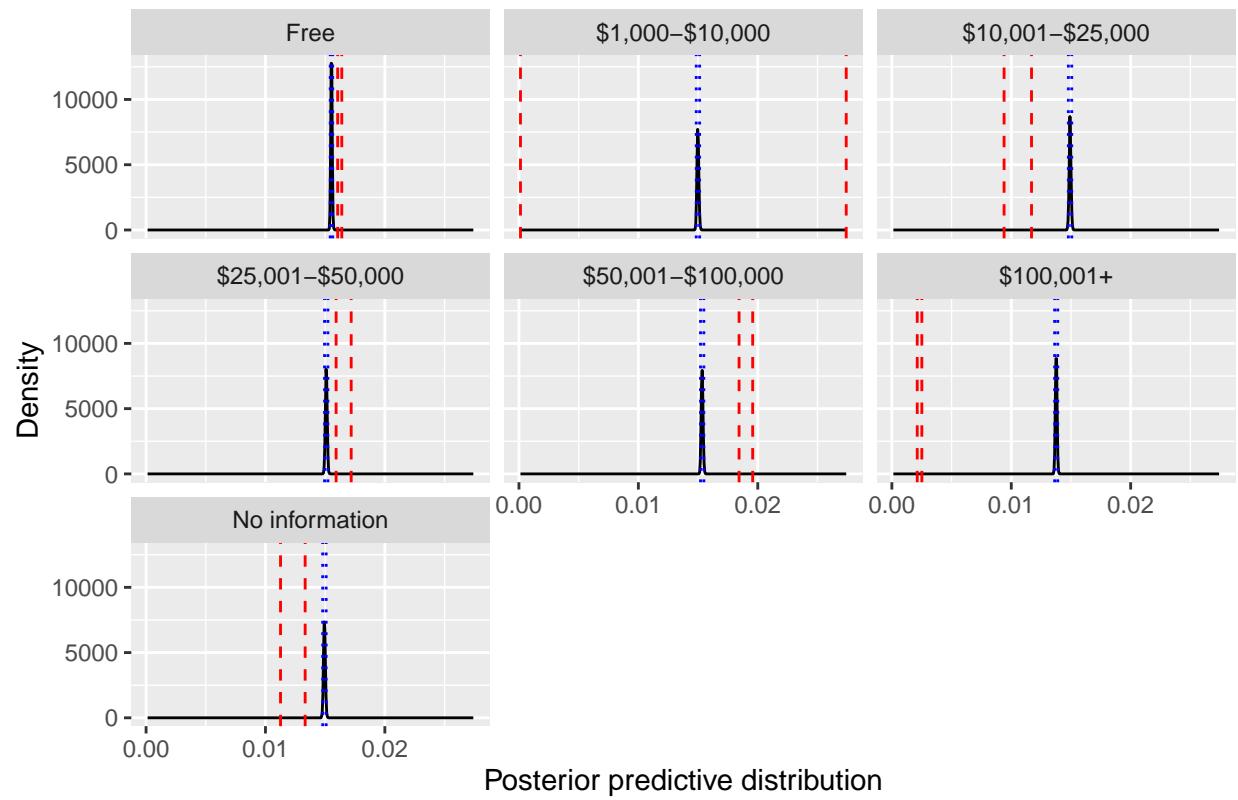


Figure 45: Posterior predictive distribution for ADHD with a random effect on socio-economic status of student's family, and with age- and sex-adjusted global prevalence prior.

Autism prevalence, prior mean = 0.00465, prior sd = 3.98e-05

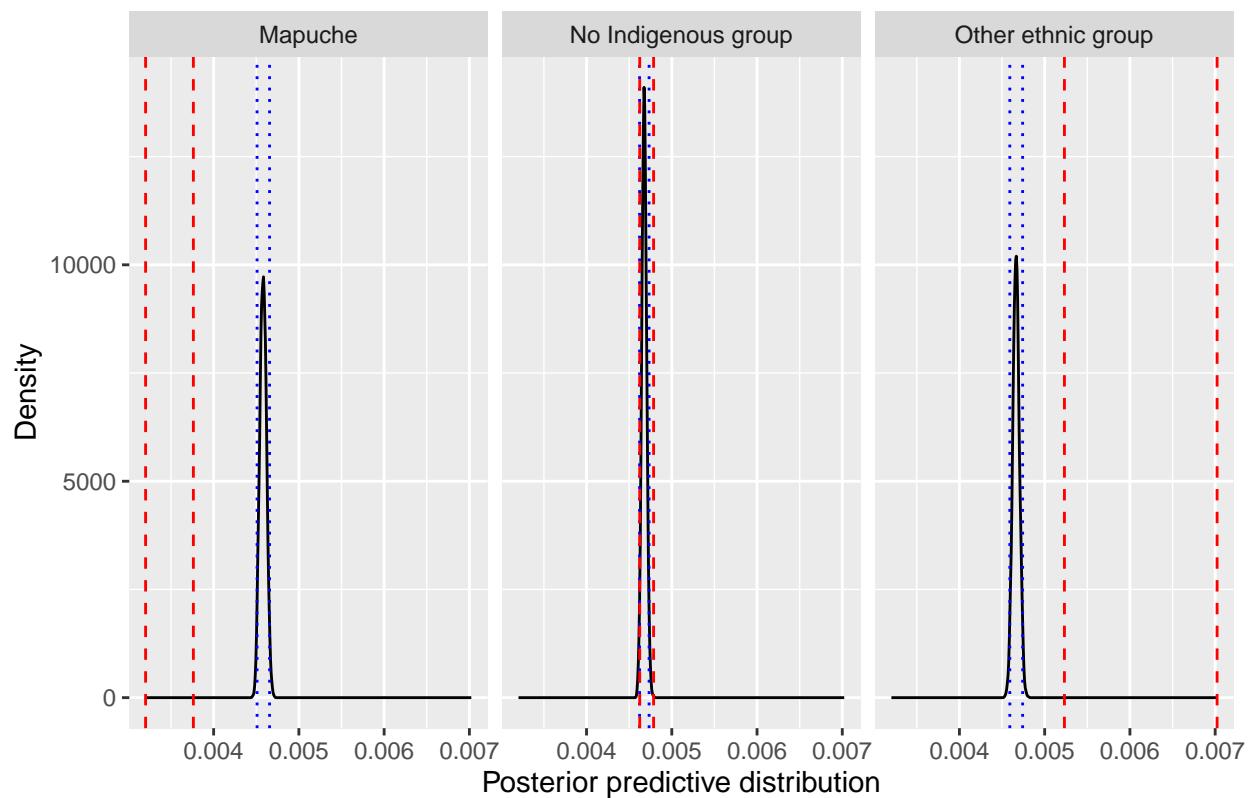


Figure 46: Posterior predictive distribution for autism with a random effect on ethnicity, and with age- and sex-adjusted global prevalence prior.

ADHD prevalence, prior mean = 0.015, prior sd = 7.25e-05

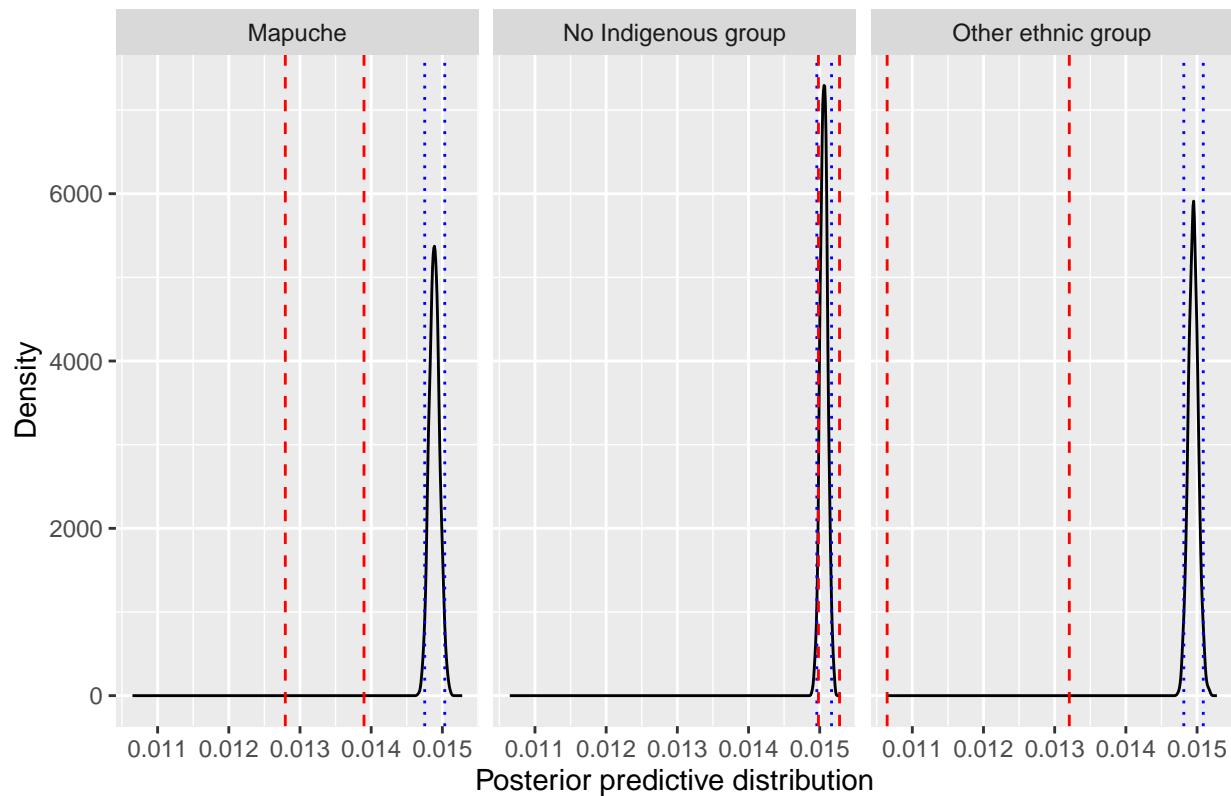


Figure 47: Posterior predictive distribution for ADHD with a random effect on ethnicity, and with age- and sex-adjusted global prevalence prior.

Autism prevalence, prior mean = 0.00465, prior sd = 3.98e-05

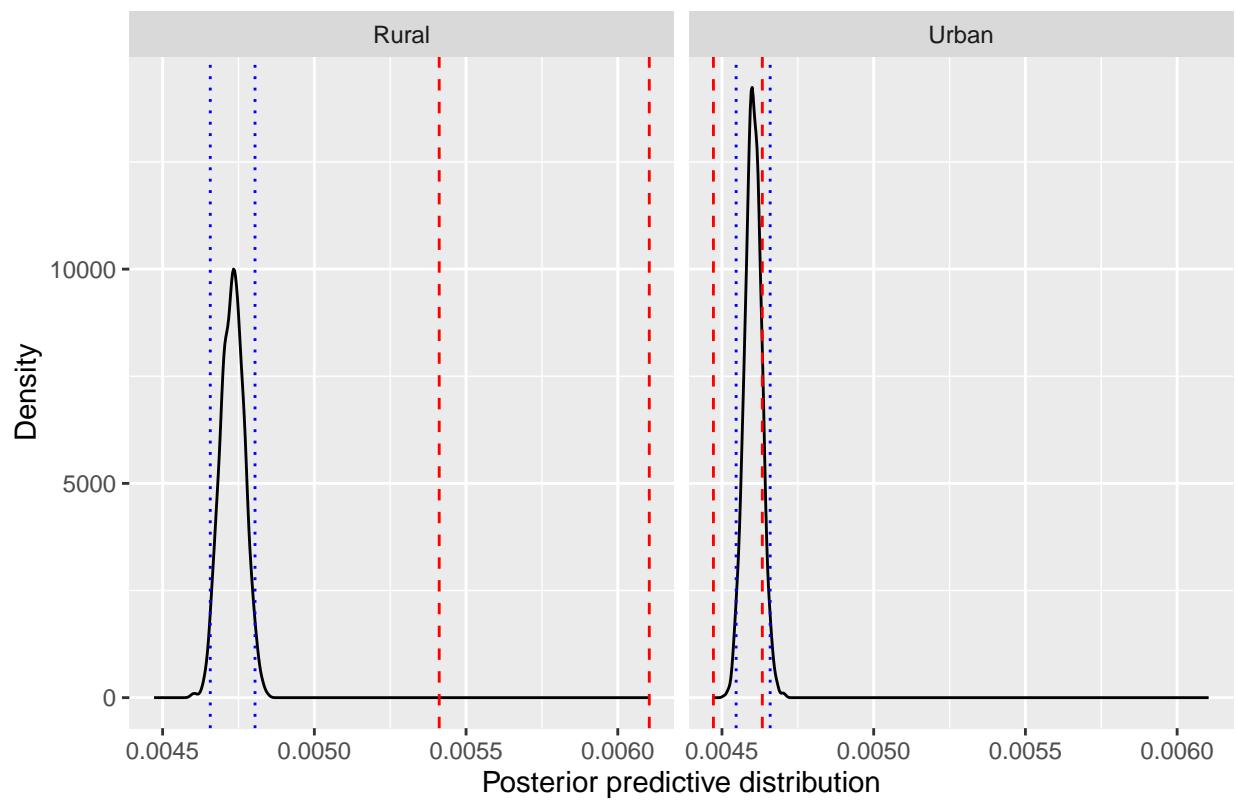


Figure 48: Posterior predictive distribution for autism with a random effect on school's rurality, and with age- and sex-adjusted global prevalence prior.

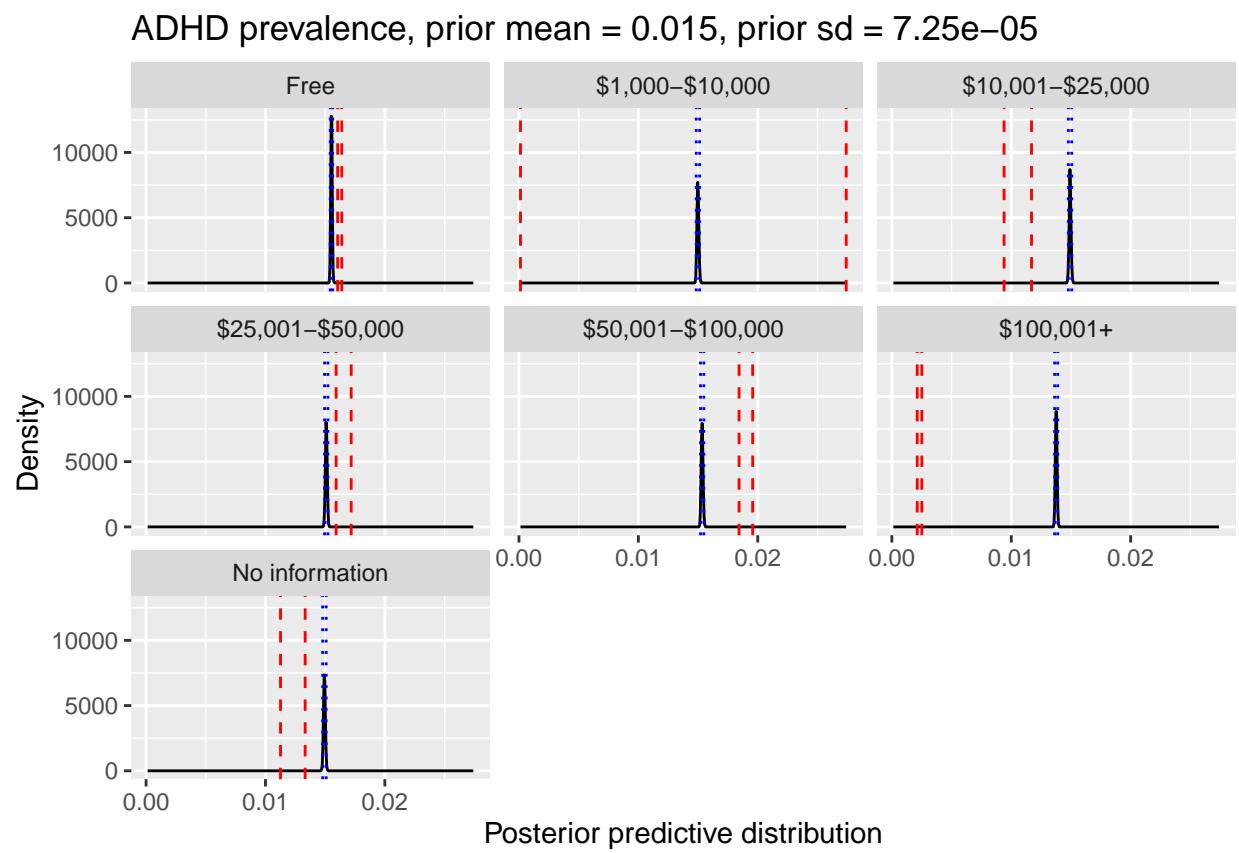


Figure 49: Posterior predictive distribution for ADHD with a random effect on school's rurality, and with age- and sex-adjusted global prevalence prior.