Seroprevalence of SARS-CoV-2 in Chile

Antoine Brault[1]

[1] *Center for Mathematical Modeling, UMI CNRS 2807, Universidad de Chile. abrault@dim.uchile.cl*

February 7, 2021

On March 16, 2020, Chile records the first death of Covid-19. From that date until January 10, 2021, the epidemic caused between 17350 and 22967 deaths according to the Departamento de Estadisticas e Información de Salud (DEIS). The excess mortality recorded in the same period is 18557.5 deaths, which corresponds to 17.2% of the total number of deaths.

Despite this significant number of deaths, which indicates that SARS-CoV-2 has strongly affected the Chilean population, currently no seroprevalence study has been published to determine the proportion of the population that has been infected by the virus. To our knowledge, there is only one study that estimates from the number of deaths, the seroprevalence between 16.29% and 22.23% on September 1, 2020 [4]. However, in that article, data are not stratified by sex and poorly by age.

Here, we use the methodology developed in [4] with detailed data from DEIS until January 10, 2021 to determine the seroprevalence in regions of Chile and the infection fatality rate (IFR). In addition, we integrate hospitalization data by age and by sex at national level to improve the accuracy of our model, and measure the infection hospitalisation rate (IHR) at national level.

We find that on December 31, 2020, the seroprevalence is between 13.4% and 16% in Chile, and between 20.9% and 24.5% in the Metropolitana region (see Table 2). IFR is estimated about 0.6% and IHR between 2.10% and 2.40%. Moreover, there are many undetected Covid-19 cases. On September 2020, only between 20.3% and 24.2% of cases have been detected. This rate falls between 12.2% and 14.2% in the Metropolitana region (see Table 3).

These results, that will have to be confirmed by a serological study, indicate that the herd immunity which is estimated between 50% and 75% of the population immunized [2], is not reached in any of the region of Chile and that many cases remain undetected.

1 Data

DEIS deaths data [1] contains Covid-19 deaths by age and by sex: confirmed by PCR suspicious. The first death occurred on March 16, 2020 and the last one on January 10, 2021. In the following we denote CD the confirmed deaths and AD all registered Covid-19 deaths. In Table 1 are reported AD, CD and excess mortality computed as the number of deaths from March 16, 2020 to January 10, 2021 minus the median of the number of deaths occured in the same period during the last 4 years.

In total and in most of the 16 regions of Chile, excess mortality is between CD and AD, but sometimes closer to CD as in Metropolitana and sometimes closer to AD as in Arica y Parinacota. There are three regions (Aysén, Los Ríos and Ñuble) where excess mortality is greater than AD, and one region (Magallanes), where excess mortality is significantly below CD. These anomalies occur in sparsely populated areas and may be explained by a statistical fluctuation in the annual number of deaths or by poor reporting of Covid-19 deaths.

Seroprevalence probability distribution inferred by our model is very sensitive with respect to the number of deaths registered. Therefore, we are going to train it separately with CD and AD. Then we will build a mixed model (MM) as the mean of the two seroprevalence probability distributions.

2 Model

The following model is strongly inspired from [4]. We denote $\mathrm{IFR}_{s,a}$ the infection fatality rate by sex s (Male or Female) and by age group a. Age groups are composed of 16 groups of 5 years intervals from 0 to 80 years and a final group of people older than 80. $\mathrm{IFR}_{s,a}$ are assumed constant by region and by time, and may be considered as spatial and temporal averages. The seroprevalence in a region r is denoted λ_r. It is assumed to be independent of age group and of sex. Let $N_{s,a,r}$ be the number of people of sex s, age group a in the region r. We assume that deaths observed $D_{s,a,r}$ are distributed as:

$$D_{s,a,r} \sim \mathrm{Poisson}(N_{s,a,r}\lambda_r \mathrm{IFR}_{s,a}).$$

Unlike to what is done in [4], we do not have a serological study to refine the model. We have chosen to use hospitalization data by age and by sex in Chile (unfortunately not available by region). Let $H_{s,a,\text{Chile}}$ be the cumulative number of people hospitalized by age and by sex in Chile. The mean delay from the onset of symptoms to hospitalization is 4 days [6, p.10 supplementary materials], whereas the mean delay to seroconversion is 10 days, and to deaths is 20 days. This should lead us to use data from hospitalized patients as of December 27, but they are only updated every 3 days. We therefore chose to use data of December 25, 2020. We assume that $H_{s,a,\text{Chile}}$ is distributed as

$$H_{s,a,\text{Chile}} \sim \text{Poisson}(N_{s,a,\text{Chile}} \lambda_{\text{Chile}} \text{IHR}_{s,a}), \tag{1}$$

where $N_{s,a,\text{Chile}}$ is the population in Chile by age and by sex, λ_{Chile} is the seroprevalence in Chile, $\text{IHR}_{\text{s,a}}$ is the probability of being hospitalized if infected, by age and by sex.. We link λ_r and λ_{Chile} with the following equality:

$$\lambda_{\text{Chile}} = \sum_r \lambda_r \frac{\sum_{s,a} N_{s,a,r}}{\sum_{s,a} N_{s,a,\text{Chile}}}.$$

Then, we seek to estimate posterior distribution $p(\lambda_r, \lambda_{\text{Chile}}, \text{IFR}_{s,a}, \text{IHR}_{s,a} | D_{s,a,r}, H_{s,a,\text{Chile}})$. The data $N_{s,a,r}$ is given by Instituto Nacional de Estadísticas de Chile. We set a uniform prior for λ_r from 0.01 to 0.99, a uniform prior for $\text{IFR}_{s,a}$ coming from estimations of [4, Supplementary Information p.22] and a uniform prior for $H_{s,a}$ coming from the data of Diamond Princess [5].

The model was trained with a variant of Hamiltonian Monte Carlo method called No-U-Turn sampler available in the package PyStan. We used 4 chains of 10 000 iterations and a burn-in of 5000.

We also estimate seroprevalance during the time $\lambda_r(t)$, using the time series of deaths by region $D_r(t)$ with the formula

$$\lambda_r(t) = \lambda_r \frac{\sum_{i=1}^{t} S_r(i)}{\sum_{i=1}^{T} D_r(i)}, \tag{2}$$

where T is the final time of observation (here December 31, 2020, because of the mean delay from seroconversion to death of 10 days), and $S_r(i)$ is the number of deaths which would have been seroconverted at time i if they had survived. The quantity $S_r(i)$ is computed from the probability laws of the delay between the onset of symptoms to seroconversion, and the onset of symptoms to death, see [4] for details.

3 Probability of death by age

In [4], only data of death people younger than 65 are used. The authors observe that in countries that have a reliable register of deaths the risk of deaths is log-linear by age from 30, when nursing home deaths are excluded. In Figure 4.3, we plot the risk of dying from Covid-19 by age for each region in Chile, i.e.

$$\log_{10}\left(\frac{\sum_{s=1}^{2} D_{s,a,r}}{\sum_{s=1}^{2} N_{s,a,r}}\right). \tag{3}$$

The log-linear relation from 30 years old generally holds in most of the regions, apart from some regions as Los Ríos or Aysén where few deaths occurred.

There is no outlier for the elderly, which seems to indicate that there was no massive outbreak in nursing homes as there where in Europe and that the older's deaths are well counted, unlike in other South American countries. We have therefore chosen, unlike [4] to train our model with all ages. This allows us to use more data and it reduces the variance of our estimation.

4 Results

4.1 Seroprevalence in Chile

We present in Table 2, seroprevalences on December 31, 2020 by region in Chile estimated with the three different models from Covid-19 deaths occurred until January 10, 2021.

Figure 2 shows different 95% credible interval of seroprevalance with the different models. We can observe that the estimation AD and CD have important differences. The mixed model (MM, green color in the figure) takes into account of the uncertainty between AD and CD. Figure 3 shows time series of seroprevalence for Chile and Metropolitana region with the mixed model.

We will be able to reduce these credible interval when serological study [3] will be available in Chile. For that we can use the method described in [4, equations 3 and 4]. Moreover, we did not model a waning of antibodies along the time, because the waning law is not yet well established in the literature. Our estimate of seroprevalence can be at least interpreted as the number of people who have already been infected with the virus, even if a proportion no longer has detectable antibodies.

4.2 Infection fatality rate (IFR) and infection hospitalisation rate (IHR)

Figure 4 shows prior distributions of IFRs, which were chosen uniform as mentioned in Section 2, and posterior distributions for AD and CD models. We observe that posterior distributions differ significantly for most of age groups. Figure 5, shows the global IFR in Chile for AD and AC model. Global IFR are very similar between models and is estimated about 0.6%. IHR which is the probability of being hospitalised when infected by SARS-CoV-2 is estimated between 2.1% and 2.4% (see Figure 6). Both IFR and IHR change along time with new treatment or hospitalization capacity. They have to be interpreted here as time average.

4.3 Covid-19 cases detected in September

From the seroconversion time series $(\lambda_r(t))_{t\in[0,T]}$ of the mixed model and knowing the distribution probability between the onset of symptoms and seroconversion, we can retrospectively deduce the number of people infected over time. The time series of positive PCR by regions is available in `https://www.minciencia.gob.cl/COVID19`. This allows us to estimate the number of cases that were detected given a period of time. Unfortunately, the information about the onset of symptoms of detected cases is only available at the national level. Therefore, we have chosen a time period where the number of cases is almost constant to avoid taking into account the unknown delay of days between the onset of symptoms and PCR test. We excluded Aysén of our study because there were too few cases in September.

In Table 3, it can be seen that at the national level, between 20.3% and 24.2% of the cases are detected which is largely insufficient to control the epidemic with contact tracing. The Metropolitan region has a particularly low detection rate ($[12.2\%, 14.2\%]$) while Magallanes region's is the highest ($[56.8\%, 79.5\%]$).

Data and code

Data and the code are available at `https://github.com/AntoineBraultChile/seroprevalenceCovidChile`.

Acknowledgement

I thank Héctor Ramírez y Joaquín Fontbona for their careful reading and interesting discussions regarding the content of this article. I also thank Maïko de Montrond who corrected some mispelling. Antoine Brault was supported by Basal Program CMM ANID PIA AFB 170001, from ANID (Chile).

References

[1] *DEIS*. `https://deis.minsal.cl/`. Accessed: 2021-01-01.

[2] Arnaud Fontanet and Simon Cauchemez. "COVID-19 herd immunity: where are we?" In: *Nature Reviews Immunology* 20.10 (2020), pp. 583–584.

[3] *ISCI Seroprevalencia*. `https://covidanalytics.isci.cl/seroprevalencia/`. Accessed: 2021-01-01.

[4] Megan O'Driscoll et al. "Age-specific mortality and immunity patterns of SARS-CoV-2". In: *Nature* (2020), pp. 1–6.

[5] Henrik Salje et al. "Estimating the burden of SARS-CoV-2 in France". In: *Science* (2020).

[6] Patrick G. T. Walker et al. "The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries". In: *Science* 369.6502 (2020), pp. 413–422. ISSN: 0036-8075. DOI: `10.1126/science.abc0035`. eprint: `https://science.sciencemag.org/content/369/6502/413.full.pdf`. URL: `https://science.sciencemag.org/content/369/6502/413`.

Region	DEIS PCR confirmed deaths (CD)	DEIS all Covid-19 deaths (AD)	Excess deaths
Tarapacá	309	368	332.5
Antofagasta	600	800	602.5
Atacama	117	152	97.0
Coquimbo	315	441	371.5
Valparaíso	1303	1912	1600.5
O'Higgins	618	916	768.0
Maule	627	841	677.5
Biobío	917	1176	1015.5
Araucanía	356	592	502.0
Los Lagos	298	464	419.5
Aysén	18	25	84.0
Magallanes	244	314	187.5
Metropolitana	11030	14180	10966.5
Los Ríos	119	171	244.0
Arica y Parinacota	241	306	297.0
Ñuble	238	309	392.0
Chile	17350	22967	18557.5

Table 1: Comparaison of AD, CD and excess deaths by region in from March 16, 2020 to January 10, 2021.

Seroprevalence (%) estimated on December 31, 2020 (95% CI)

Region	Mixed model	CD model	AD model
Tarapacá	[14.6, 18.0]	[14.3, 17.7]	[14.8, 18.1]
Antofagasta	[16.6, 21.5]	[16.4, 19.1]	[19.1, 21.8]
Atacama	[5.1, 7.7]	[4.9, 6.9]	[5.8, 7.9]
Coquimbo	[4.7, 6.9]	[4.6, 5.6]	[5.9, 7.0]
Valparaíso	[7.9, 11.3]	[7.8, 8.7]	[10.5, 11.4]
O'Higgins	[7.8, 11.6]	[17.7, 8.9]	[10.4, 11.7]
Maule	[6.9, 9.3]	[6.8, 7.9]	[8.3, 9.4]
Biobío	[7.2, 9.1]	[7.1, 8.0]	[8.3, 9.2]
Araucanía	[4.1, 7.2]	[4.1, 5.0]	[6.3, 7.3]
Los Lagos	[4.3, 7.1]	[4.2, 5.3]	[6.0, 7.2]
Aysén	[1.8, 4.4]	[1.6, 4.0]	[2.1, 4.6]
Magallanes	[17.4, 24.1]	[17.0, 21.5]	[19.9, 24.5]
Metropolitana	[20.9, 24.5]	[20.9, 21.5]	[23.8, 24.6]
Los Ríos	[3.3, 5.5]	[3.2, 4.5]	[4.2, 5.7]
Arica y Parinacota	[13.2, 18.0]	[13.0, 16.5]	[14.8, 18.4]
Ñuble	[5.0, 7.1]	[4.9, 6.2]	[5.9, 7.2]
Chile	[13.4, 16.0]	[13.4, 13.7]	[15.7, 16.0]

Table 2: 95% credible interval of the seroprevalence distribution of the three models.

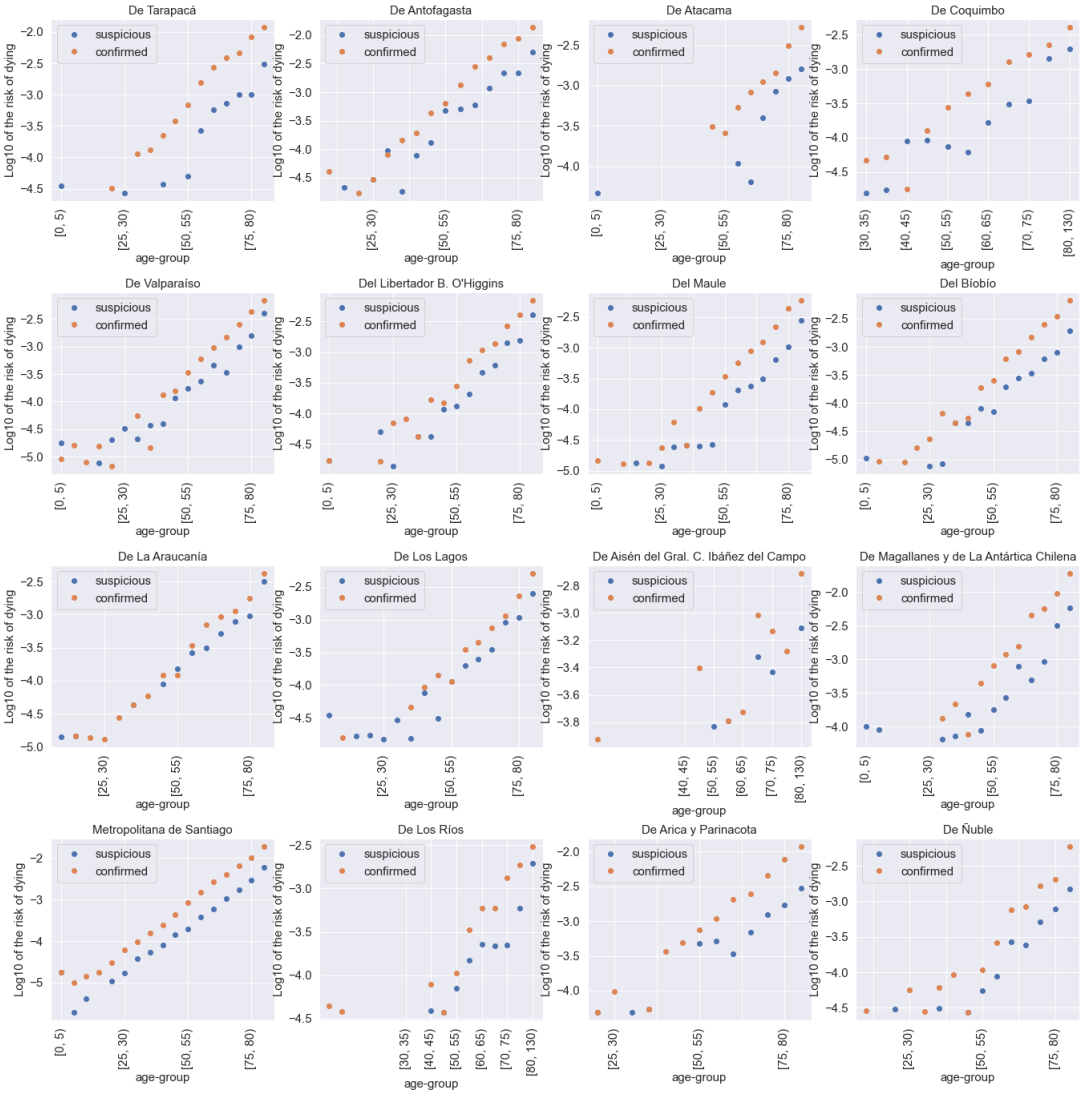

Figure 1: Risk of death by region and by age with PCR confirmed deaths and suspicious deaths that were not confirmed by PCR from DEIS data.

Region	Percentage of cases detected by PCR in September (95% CI)
Tarapacá	[25.3, 31.6]
Antofagasta	[22.5, 29.3]
Atacama	[31.4, 48.1]
Coquimbo	[23.2, 34.6]
Valparaíso	[15.8, 22.8]
O'Higgins	[21.1, 31.7]
Maule	[26.8, 36.5]
Biobío	[32.4, 41.3]
Araucanía	[28.2, 49.4]
Los Lagos	[32.7, 54.1]
Aysén	
Magallanes	[56.8, 79.5]
Metropolitana	[12.2, 14.2]
Los Ríos	[35.2, 60.0]
Arica y Parinacota	[28.5, 39.2]
Ñuble	[30.2, 43.4]
Chile	[20.3, 24.2]

Table 3: 95% credible interval of cases detected by PCR in September 2020. This estimation comes from the mixed model.

Figure 2: 95% credible interval of seroprevalence by region with the different models (AD: all deaths, CD: confirmed deaths, MM: mixed model).

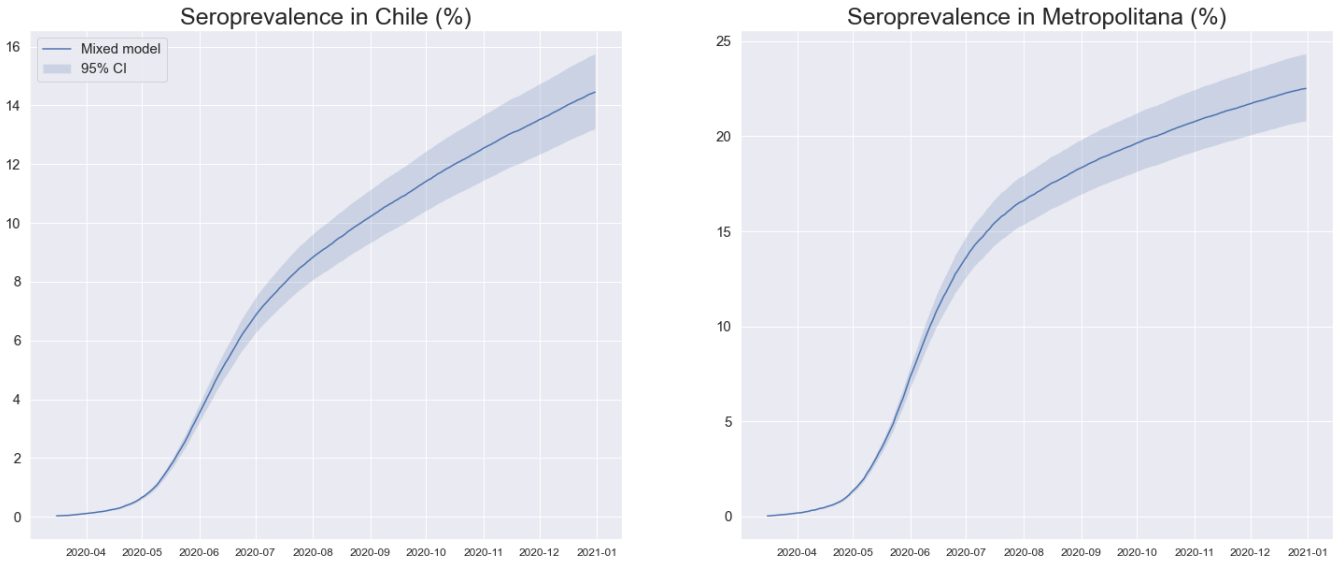

Figure 3: Evolution of seroprevelance in Chile and in the Metropolitana region from March 16 to December 21 2020 with 95% credible interval.

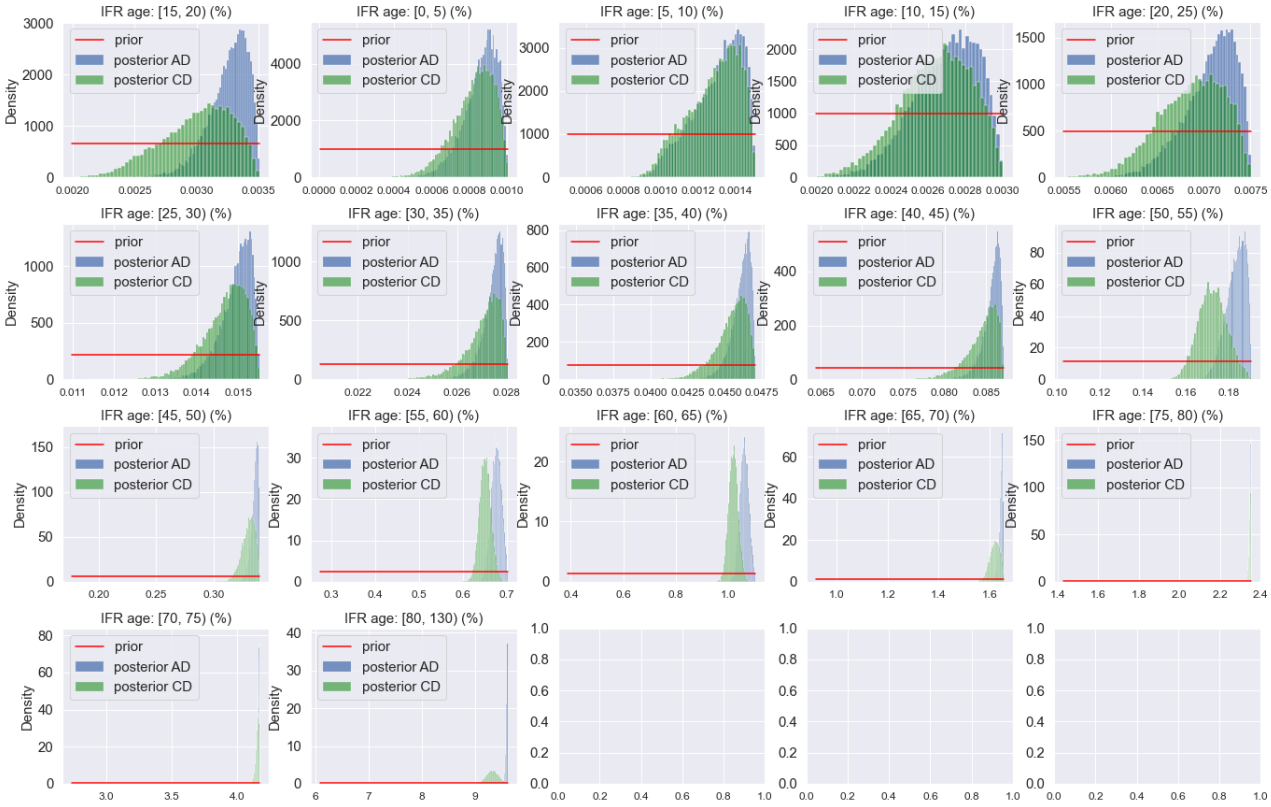

Figure 4: Prior and posteriors for the two models AD and CD of IFR by age in Chile.

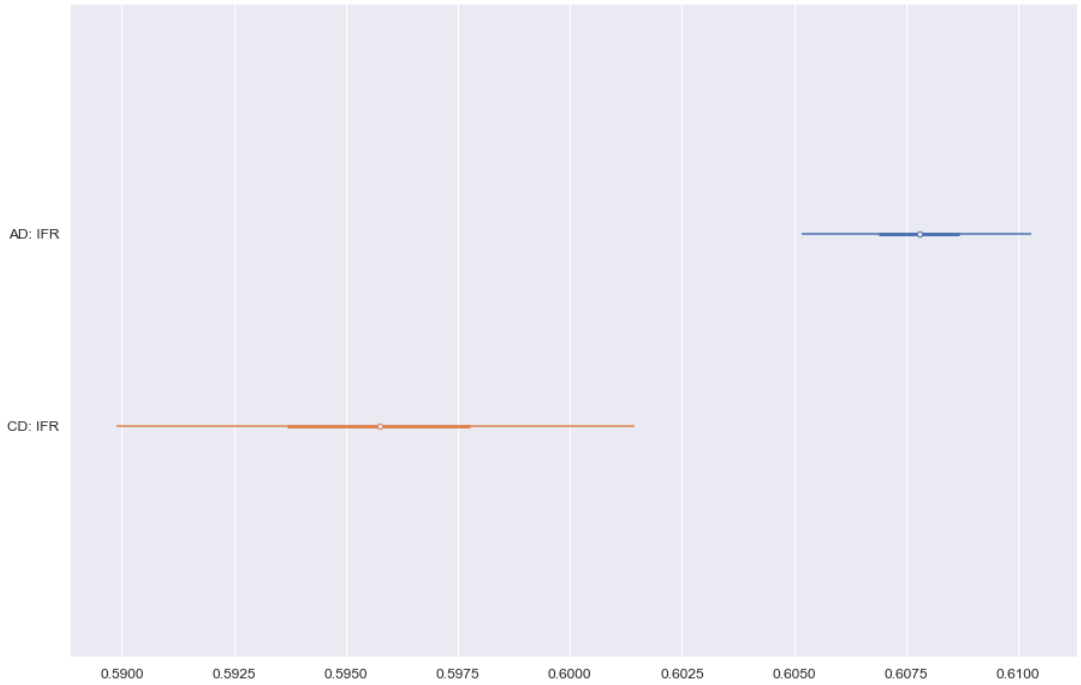

Figure 5: 95% credible interval of global IFR (%) in Chile for the different models (AD: all deaths, CD: confirmed deaths

.

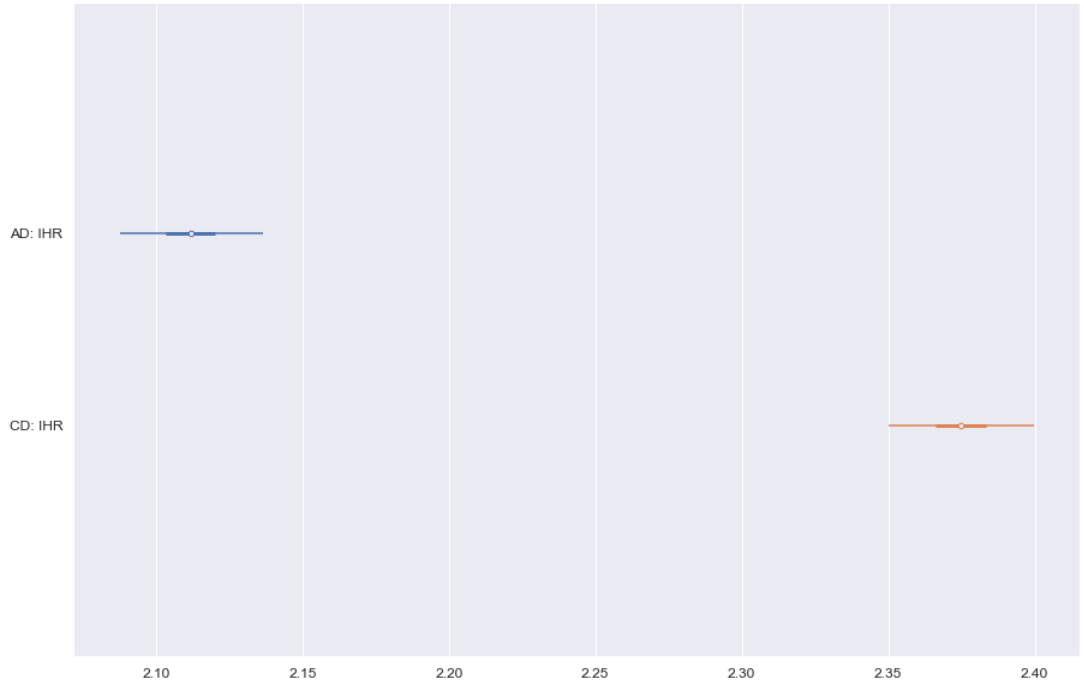

Figure 6: 95% credible interval of global IHR (%) in Chile for the different models (AD: all deaths, CD: confirmed deaths

.