UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DESIGN AND ANALYSIS OF A DYNAMIC IGG SEROPOSITIVITY STUDY FOR COVID-19 USING MIP AND BAYESIAN INFERENCE

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN GESTIÓN DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

SIMÓN GRASS ARAYA

PROFESOR GUÍA:
DENIS SAURÉ VALENZUELA

MIEMBROS DE LA COMISIÓN:
LEONARDO BASSO SOTZ
MIGUEL LUIS O'RYAN GALLARDO
JUAN PABLO TORRES TORRETTI
CHARLES THRAVES CORTÉS-MONROY

SANTIAGO DE CHILE
2022

Resumen

Diseño y análisis de un estudio dinámico de seropositividad de IgG para COVID-19 mediante MIP e inferencia Bayesiana

En esta tesis se resolverá el problema que surge al tratar de analizar la presencia de una variable binaria en una población dados diferentes factores y resolviendo un MIP que buscaba lograr la mayor representatividad posible de la muestra. En este caso particular, el problema planteado fue conocer la presencia de anticuerpos para SARS-CoV-2 en la población de Chile, teniendo en cuenta diferentes parámetros biológicos y no biológicos. La implementación de los modelos implicó la realización de pruebas de IgG; donde un resultado positivo indicaría la presencia de anticuerpos en el sujeto, ayudando tanto a disminuir la probabilidad de contraer SARS-CoV-2 como a disminuir la gravedad de este en caso de ser contraído.
El primer modelo que presentamos busca lograr la máxima representatividad de la población para los centros urbanos de Chile, utilizando las zonas censales como parámetro geográfico para medir la representatividad geográfica, además de otros factores como la edad y las comorbilidades. Los resultados del primer modelo muestran que fue posible obtener una muestra representativa mucho mayor. A modo de ejemplo, *Gran Santiago* mostró una utilización teórica del 84% de los resultados (12957 de 15404 muestras) a partir del 13 de julio, una mejora importante teniendo en cuenta que el marco temporal anterior tenía 15% de datos utilizables (1182 de 7902)
Futuras implementaciones de un modelo de este tipo deberían buscar la mayor flexibilidad posible en la reasignación de los lugares de recolección de muestras, ya que este factor demostró ser la mayor limitación a la hora de cerrar la brecha entre la implementación y el modelo teórico, por lo que mejorarlo aumentaría en gran medida la posibilidad de adaptarse a los datos recolectados y obtener una muestra más representativa.
El segundo modelo presentado en la tesis pretende analizar la muestra recolectada, con el fin de estimar la probabilidad de detectar la presencia de IgG asumiendo una prueba perfecta. Utiliza variables biológicas como la edad y la comorbilidad, así como variables no biológicas como el método de transporte y la frecuencia del mismo. Combina estos factores como una regresión logística para estimar la probabilidad descrita. Se utiliza un enfoque bayesiano y el algoritmo Marcov Chain Monte Carlo para ajustar el modelo. Nuestros resultados muestran una notable diferencia en la presencia esperada de IgG entre individuos vacunados y no vacunados, así como una considerable diferencia entre vacunas, donde BNT162b12 muestra una mayor seroprevalencia.
Futuras implementaciones de un modelo de este tipo deberían tratar de optimizar tanto el código como el hardware utilizado, con el objetivo de perfeccionar los resultados y reducir la complejidad temporal del algoritmo.

Abstract

This thesis will solve the problem that arises when trying to analyze the presence of a binary variable in a population given different factors and solving a MIP that sought to achieve the biggest representative sample possible. In this particular case, the problem presented was understanding the presence of antibodies for SARS-CoV-2 in the population of Chile, taking into consideration different biological and non-biological parameters. The implementation of the models involved testing for IgG; having a positive result that would indicate the presence of antibodies in the subject, helping in both lowering the probability of contracting SARS-CoV-2 as well as lessening the severity of it if contracted.

The first model we present seeks to achieve the maximum representative sample of the population for urban centers in Chile, using census zones as a geographical parameter to measure geographical representativeness, as well as other factors such as age and comorbidities. Results from the first model show that it was possible to obtain a much larger representative sample. As an example, *Gran Santiago* showed a theoretical usage of 84% of the results (12957 out of 15404 samples) as of July 13th, an important improvement considering the prior time frame had a usable data of 15% (1182 out of 7902)
Future implementations of a model of this kind should seek as much flexibility as possible in the reallocation of sites to collect samples, as this factor proved to be the biggest limitation at closing the gap between the implementation and the theoretical model, hence improving it would greatly increase the possibility to adapt to the collected data and get a larger representative sample.

The second model presented in the thesis seeks to analyze the sample collected, in order to estimate the probability to detect the presence of IgG assuming a perfect test. It used biological variables such as age and comorbidity, as well as non-biological variables such as method of transportation and frequency of transportation. It combines these factors as a logistic regression to estimate the probability described. Using a bayesian approach and Marcov Chain Monte Carlo algorithm to fit the model. Our results show a notable difference in the expected presence of IgG between vaccinated and not vaccinated individuals, as well as a considerable difference between vaccines, where BNT162b12 shows higher seroprevalence.
Future implementations of a model of this kind should seek to optimize both the code and hardware used, aiming to refine results and lower the algorithm's time complexity.

This thesis is dedicated to my family, without them there is not a single achievement in my life that would have been possible

Acknowledgments

First, I would like to thank my family, who supported me not only during this process but also during my whole education, instilling in me the hunger for knowledge that allowed the development of my personality and skills.

I would also like to thank Denis Sauré, who with a single course was able to get me hooked with Operation Management and later became the teacher who guided me through this whole thesis.

I would like to thank Leonardo Basso and all the other contributors that made the implementation of this combat against COVID-19 possible.

Last but not least, I would like to thank my friends, who made these 6 years of my life much better, making my passage through Engineering much more enjoyable.

Table of Content

List of Figures

Chapter 1

Introduction

SARS-CoV-2 is a virus that generated a pandemic starting in December 2019 in Wuhan China. Countries all over the world had to make tough decisions to control the situation as well as possible. An important factor in decision-making is not only to understand the virus and to create a vaccine, but to understand the current state of the population in order to defend itself against the virus.

Chile, as many other countries, took emergency measures to stop the problem as it was investigated, starting with a lockdown, and evolving into social distancing and sanitary measures. As this evolved, availability of information for decision-making was crucial.

Two main factors help decision-making in different ways. The first one consists in the current number of people infected with SARS-CoV-2, the number of interactions and severity of cases those present, helping decisions such as hospitalization capacity in different sectors. The second factor is understanding the presence of antibodies in the population, helping estimate in particular the severity that the first factor could bring given those interactions.

This thesis will focus on the latter, by testing for IgG antibodies, having a positive result that would indicate the presence of antibodies in the subject and helping in both lowering the probability of contracting SARS-CoV-2 as well as lessening the severity of it if contracted.

In particular, the objective of the study was to collect data as close to the Chilean population as possible in order to help this decision-making in a national scale.

The use of this information depends both on data collection as well as its analysis, therefore the following chapters will be mainly divided in two parts; the first one refers to a model to collect data as representative of the population as possible, and the second is a model that helps estimate the positivity of IgG antibodies in the population.

For the first part, we used a model that, as is seen in chapter 3 section 2, aims to improve the maximum representative data. As an example, *Gran Santiago* could have enabled the usage of 84% of the data (12957 out of 15404 samples), an important improvement considering that the samples prior to the implementation of the model had a usable data of 15% (1182 out of 7902)

The second part uses a model that, as is seen in chapter 4 section 2, shows a notable difference in the expected presence of IgG between vaccinated and not vaccinated individuals, as well as a considerable difference in the vaccines, as BNT162b2 presents higher seroprevalence through time, putting the usefulness of the vaccine in evidence.

It is important to note that this is done in the context of a real serological study, and the results obtained were in fact implemented. This was possible with the help of health authorities; by assigning the logistic responsibilities of installing the selected sites that recollect samples of IgG antibodies to the corresponding Health Services[1].

[1]State entities that have assigned territories in which they are responsible for the implementation of integrated actions for the promotion, protection, and recovery of health.

Chapter 2

Objectives and Background

2.1 Objectives

2.1.1 General Objectives

This thesis resolves the problem of sampling a population so as to match theoretical population data by the use of Mixed integer optimization as well as the use of this data, logistic regression, and Bayesian inference in order to estimate the presence of antibodies in the population.

2.1.2 Specific Objectives

1. Localization model
   (a) Improve the maximum representative sample retrievable from the data collected.
   (b) Help the decision-making of health authorities, particularly in selecting the locations at which one should collect data through time.
   (c) Show the application of the model and corresponding analysis to inspire future replication in similar use cases.
2. Seroprevalence Model
   (a) Gather analysis that generates insight about which factors of the individuals can be related to the expected seroprevalence.
   (b) Give insight about what is the situation of seroprevalence for the population of Chile.
   (c) Show the application of the model and corresponding analysis to inspire future replication in similar use cases.

2.2 Background

The implementation of this operation was made possible with the help of *Redes Asistenciales*[1], which supported the coordination of the corresponding Health Services. Each Health Service (see Figure 2.1 in the appendix for the full list) was in charge of the installation of a testing station in their corresponding territory, with the exception of *Araucanía Sur* which had capacity to manage the installation of two testing stations, and the Health Services *Metropolitano Central, Norte* and *Occidental* which did not participate in the study as entities, having that territory for site installation directly managed by *Redes Asistenciales* with two different testing station to be allocated.

The territory of the 29 Health Services was separated into 21 urban centers and 86 counties, in which individuals were mainly vaccinated with Coronavac and BNT16b2 in a 2:1 ratio approximately.

To show Health Services where to install the sites, a map was provided via web offering the solution of the model with the recommended location for site installation (an example can be seen as Figure 2.2 in the Appendix). Meetings were offered to explain the usage of the tools provided, as well as to discuss the exact position of the site to be installed.

Individuals that participated in the study provided information via a form that was filled in real time by people of the corresponding Health Service, as can be seen in the appendix as Figure 2.3.

[1]State entity whose mission is to regulate and supervise the operation of health networks.

Chapter 3

Localization Model

3.1 Data Collection Method

From March 12, 2021, 28 testing stations for SARS-CoV-2 IgG detection were installed in hotspots based on cellular-phone mobility tracking within the most populated cities in Chile. Each testing station was assigned to a Health Service with the corresponding jurisdiction over the area.
In each station, individuals were invited to do a lateral flow test (LFT) by finger prick voluntarily and respond to a questionnaire on sociodemographic characteristics, vaccination status (including type of vaccine if one was received), variables associated with SARS-CoV-2 exposure, and comorbidities.

For the localization model, the initial and main objective was to have representative geographical data. A good representation means that samples for each county are similar, percentage-wise, to the official census percentages for each urban center.
For this section, individuals were excluded if they did not live in any urban center in Chile (noting the fact that there are certain counties that do not belong to any). Further analysis of data in latter chapters will exclude individuals if they were younger than 18 years, had no declared gender, had an invalid IgG test result, had previously tested positive for SARS-CoV-2 infection on PCR, could not recall their vaccination status, and further analysis will even exclude individuals had been immunized against COVID-19 with vaccines other than CoronaVac or BNT162b2.
Data reported for all the following chapters corresponds to people tested up to July 2nd, 2021 for every urban center other than *Gran Santiago*, for which we report data collected up to December 13th.

3.2 Localization model

An optimization model was used mixed-integer program (MIP) based on weekly analysis of national mobile phone mobility data, facilitated by Chile's largest telecommunications agency (*Empresa Nacional de Telecomunicaciones*, Santiago, Chile), to select sites with high traffic volume and wide county-level distribution of people.
The model aims at maximizing the size of a representative sample according to the geographical distribution, at county granularity, for each urban center. The model used was as follows:

Parameters

- I: set of census zones.
- J: set of counties.
- $m_{(i,j,t,tb)}$: expected number of samples from county j obtained in the census zone i during day t in time block tb.
- p_j: population of county j, relative to the total of the urban center.
- B_{tb}: number of testing sites to be allocated on each time block tb.
- T: set of days of the week (Monday to Friday). Not including weekends
- TB Time Block, T1 for mornings and T2 for afternoons

We need to decide in what census zone should the sites be located. We define for $i \in I$, $t \in T$ and $tb \in TB$ the decision variable

$$x_{i,t,tb} = \begin{cases} 1 & \text{if a site is assigned to zone } i \text{ in day } t \text{ and time block } tb \\ 0 & \sim \end{cases}.$$

For a given allocation, we compute the number of samples to obtain from the county $j \in J$ as

$$y_j = \sum_{i,t,tb \in I,T,TB} x_i \, m_{(i,j,t,tb)}.$$

Then, to obtain a representative sample of the urban center of size n, we have to collect at least $n\,p_j$ samples from county j for each county in J. Then, given the allocation of sites, the size of the representative sample is the maximum value of n that satisfies the condition

$$n\,p_j \leq y_j, \forall j \in J.$$

We want to maximize the size of the representative sample obtained by the allocation of testing sites. With this, we had the following optimization problem

$$\max n \tag{3.1a}$$

$$s.t. \quad n\,p_j \leq \sum_{i,t,tb \in I,T,TB} x_{i,t,tb}\, m_{(i,j,t,tb)} \forall j \tag{3.1b}$$

$$\sum_{i \in I} x_{i,t,tb} \leq B \quad \forall t, tb \tag{3.1c}$$

$$x_{i,t,tb} \in \{0,1\}. \tag{3.1d}$$

Because the theoretical and real data collected were understandably not the same, the model was solved on a rolling horizon basis with a weekly plan for resolving, weighting in the recollected data in the model, re-calculating the sites to test accordingly.
This was done by considering the data already recollected on the right side of restriction 3.1b. In that way, the optimization problem would factor in the current distribution of samples taken and weight in that into the approximation of the best representative data in the future. Then, the size of a sample from j is:

$$\sum_{i,t,tb \in I,T,TB} x_i\, m_{(i,j,t,tb)} + k(j).$$

Where $k(j)$ is the current number of samples obtained for j. With this, the optimization problem becomes:

$$\begin{aligned} & \max n & \\ s.t. \quad & n\,p_j & \leq \sum_{i,t,tb \in I,T,TB} x_{i,t,tb}\, m_{(i,j,t,tb)} + k(j). \quad \forall j \\ & \textstyle\sum_{i \in I} x_{i,t,tb} & \leq B \quad \forall t, tb \\ & x_{i,t,tb} & \in \{0,1\}. \end{aligned}$$

As samples were collected, another difference was noted: the mobility data was underestimating the internal movement of individuals within a county. This endogenous movement represented people that moved within census zones of the same county, and as higher internal movement occurred; data collection for a given $\hat{j}$ had more samples of individuals from $\hat{j}$ than expected. This meant that an adaptation to the model was necessary in order to correct the estimation and data recollection, to account for this endogenous movement.
To correct the proportion of internal versus external movement of the recollected data, an analysis was made in contrast to the data given by telecommunications. This adjustment consisted in comparing the data collected in the census zone with the theoretical data to measure the proportion of endogenous movement. Then to adapt to this, the model used that difference in proportion as a factor to re-normalized the distribution of endogenous versus non-endogenous movement accordingly.

Because of this dynamic adaptation, incorporating the current data recollected as a factor, the model was run periodically to adjust to the best solution given the current sample. This allowed a plan for data recollection that was limited by the capacity to change locations of each Health Service.

Figure 3.1 shows an example of a particular solution. The data used was July 13 and for a more clear representation of how the solution gets a good representative sample size over time, a whole month of solution follows. To accomplish this, the solution of the week before assumes an optimistic data recollection of 50 samples by time block, using that theoretical result to add into the existing sample for the next week's solution.

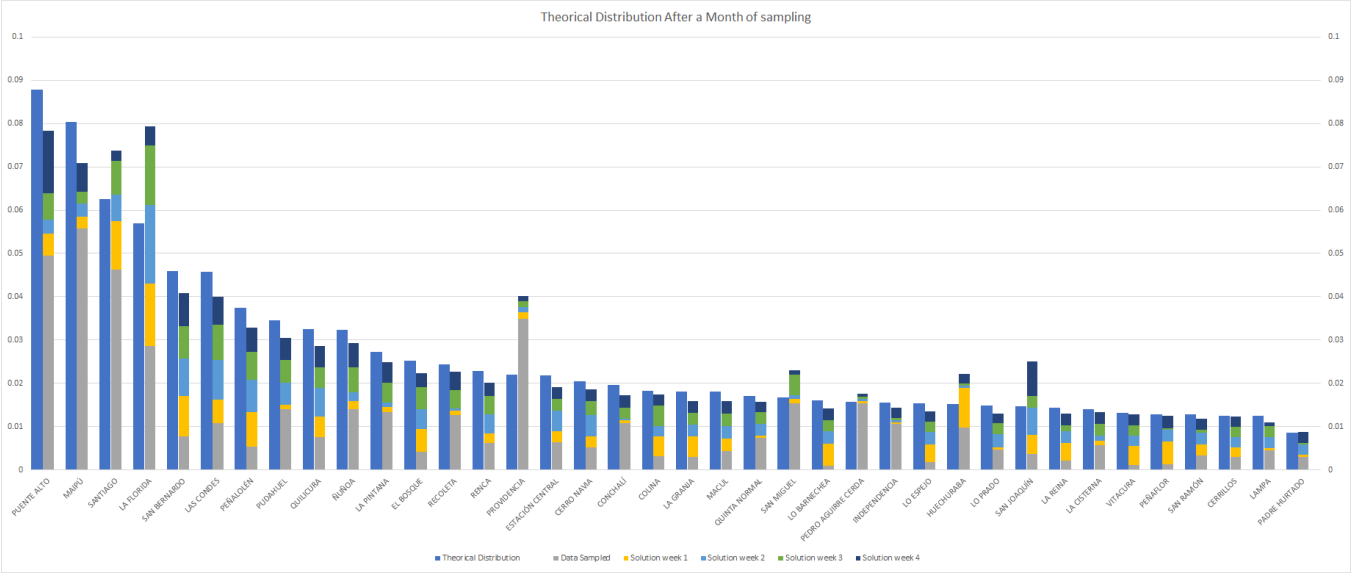

Figure 3.1: Gran Santiago Urban Center Solution

The blue bars represent the real distribution of the urban center based on census data, the gray portion of the bars represent the sampled data, the bar's yellow portion represent the samples that should be collected based on the solution of the first week, and the bar's light blue, green, and dark blue portions correspond to the samples that should be collected on the second, third, and fourth week correspondingly.
To put the result into perspective; this allows a usable representative data of 12957 out of the total sampled data of 15404 (84%), whereas before that month the current usable data was 1182 out of the 7902 (15%).

Figure 3.2 shows how the distribution looked for that date, before the solution was deployed.

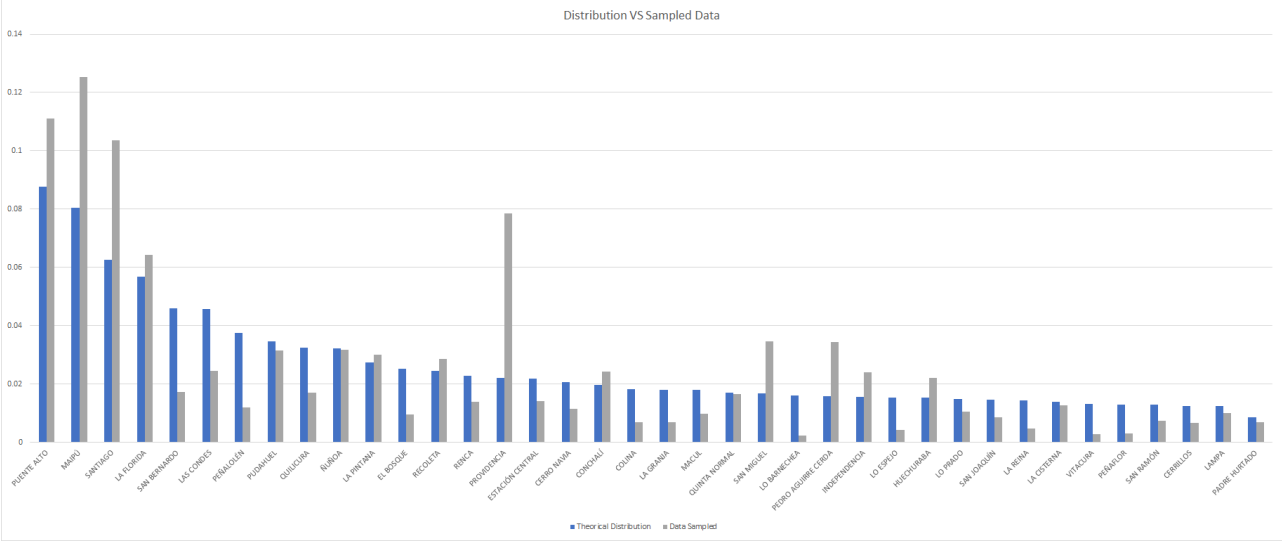

Figure 3.2: Gran Santiago Urban Center Before Solution

Figure 3.3 shows the amount of data sampled for each county given the previous solution.

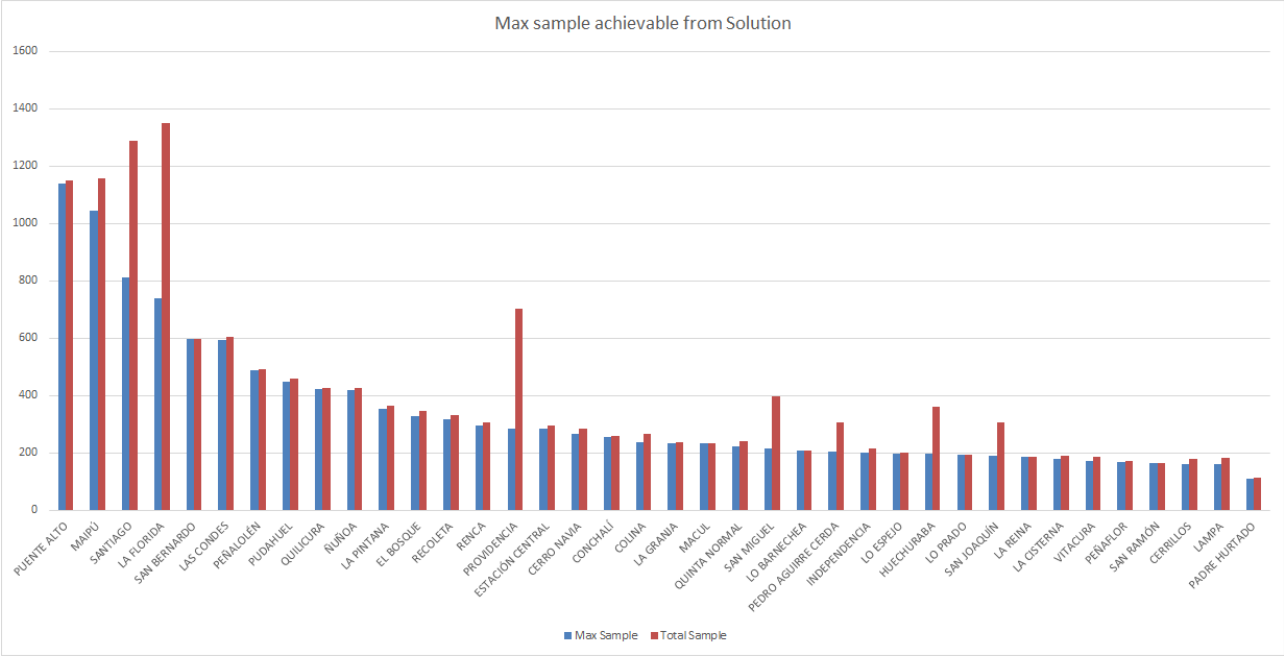

Figure 3.3: Gran Santiago Urban Center Total Sample

Here, the blue bar represents the maximum number of representative samples for each county and the red bar the total number of samples for each county.

3.3 Limitations to Data recollection

There are several factors that limited data collection in different ways. In a perfect scenario the model could be as dynamic as having morning and noon time blocks given that the movement in the morning and afternoons tends to vary by factors such as labor time schedules. In this scenario, the flexibility could allow the program to be run after the morning data was recollected to evaluate a change in location for the afternoon.

In reality, this flexibility was not possible since every Health Service had to make administrative requests before changing location, which took more than a labor day. This meant, in the long run, that changes could only be done at most on a weekly basis.

Another Limitation was related to the specific location in the census zone. Each site had to be safe for the people working there, have a roof and bathroom nearby, and preferably be a public space, as it would facilitate the installment of the testing station.
This meant that after having the results for the optimal set of census zones, an analysis was made in each census zone in order to understand what was the driving factor for that movement, in order to find the exact location for the site; this could be for example a shopping mall or individual store in the zone that generated that movement.

The model then worked as follows:
A month plan was made with weekly flexibility. This meant that the second week of the model assumed that the theoretical data of the first one was met, the third week assumed the first two weeks of data were achieved, and so on. This allowed Health Services to prepare all the paperwork necessary for the installation of sites in the selected places.
After each week the program was run again, this time with the real data collected replacing the theoretical data of the past week, to evaluate possible changes.
The optimization problem would then be

$$\begin{aligned} & \max n & \\ s.t. \quad & n\,p_j & \leq \sum_{i,\in I} x_i\, m_{(i,j)} + k(j). \quad \forall j \\ & \textstyle\sum_{i\in I} x_i & \leq B \\ & x_i & \in \{0,1\}. \end{aligned}$$

Where there is no longer t nor tb. As for the mobility data that helped the selection, a weekly average for each site was made, converting the time block daily granularity to an average movement by week.
As the representative sample size for several urban centers increased, other objectives could be considered. For example, as vaccination status would be a relevant parameter in future analysis; locating testing stations in sites where there was a higher number of unvaccinated individuals would increase the amount of samples that had that characteristic (as unvaccinated individuals were a minority). Other factors that make a sample representative were assumed to be met at first (gender, age), since the data was randomly sampled, but were then later confirmed in data analysis to ensure representation, and also looking for further insights into certain groups of people being more keen to participate in the study (with no

conclusive evidence that this was happening).
As weeks passed and some urban centers did not have a considerable number of weeks left, a variation was presented to the model where $m_{(i,j)}$ would now be multiplied by the approximate weeks left of study to assume no further changes would be able to be made. As for the next week's calculation, the problem would assume the same thing, having one week subtracted in that multiplication and assuming the samples of the first week were met. This could allow a more robust solution for the remaining weeks of study as it assumed time was limited, and represented a less "greedy" approach.

3.4 Results

Considering the limitations stated in the previous section, optimal data collection was not achievable for many practical reasons. In addition, the model assumed 50 collected samples per time block (being this an optimistic assumption, based on the results of the urban center of *Gran Santiago* that varied between 25-75 test for each time block approximately). Noting that this is the urban center with the highest population and that even though 50 is half of the range mentioned, the average number of samples collected lowered as the study continued.

As of 2 of July, the total number of data recollected was 66625 for all urban centers except *Gran Santiago*, which continued sampling data consistently for a longer period of time, the data sampled as of 13 of December for the urban center of *Gran Santiago* was 23444 as a total, of which 18491 was useful for analysis (meeting the criteria mentioned before).

The distribution of data recollected for each urban center for those dates resulted as follows:

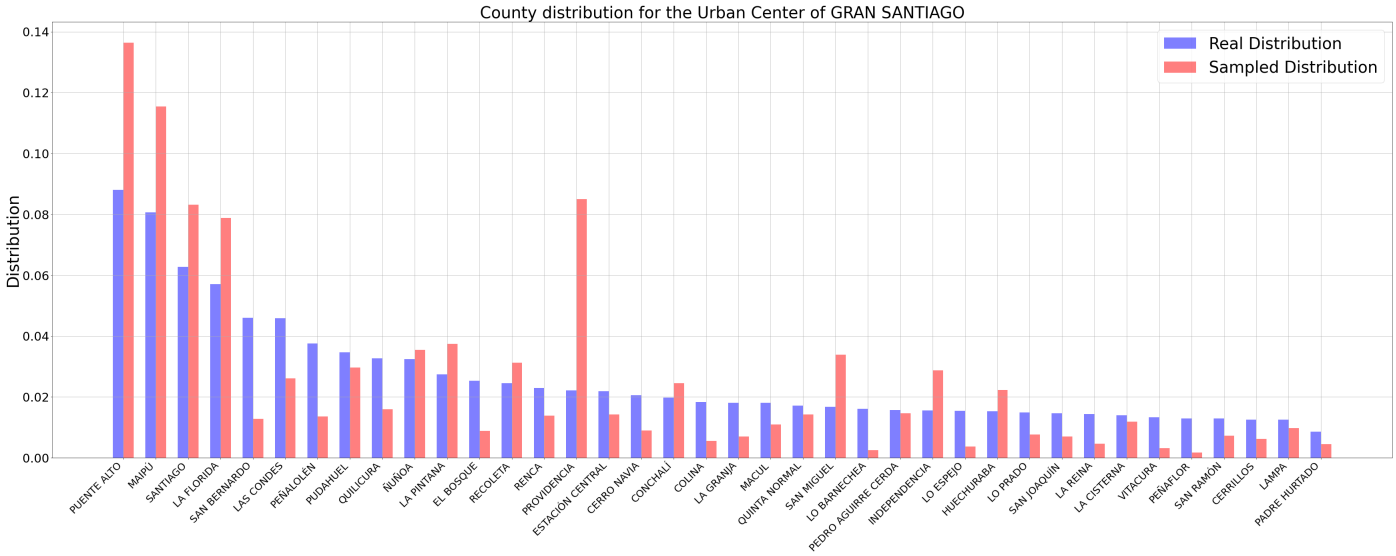

Figure 3.4: Gran Santiago Urban Center Distribution

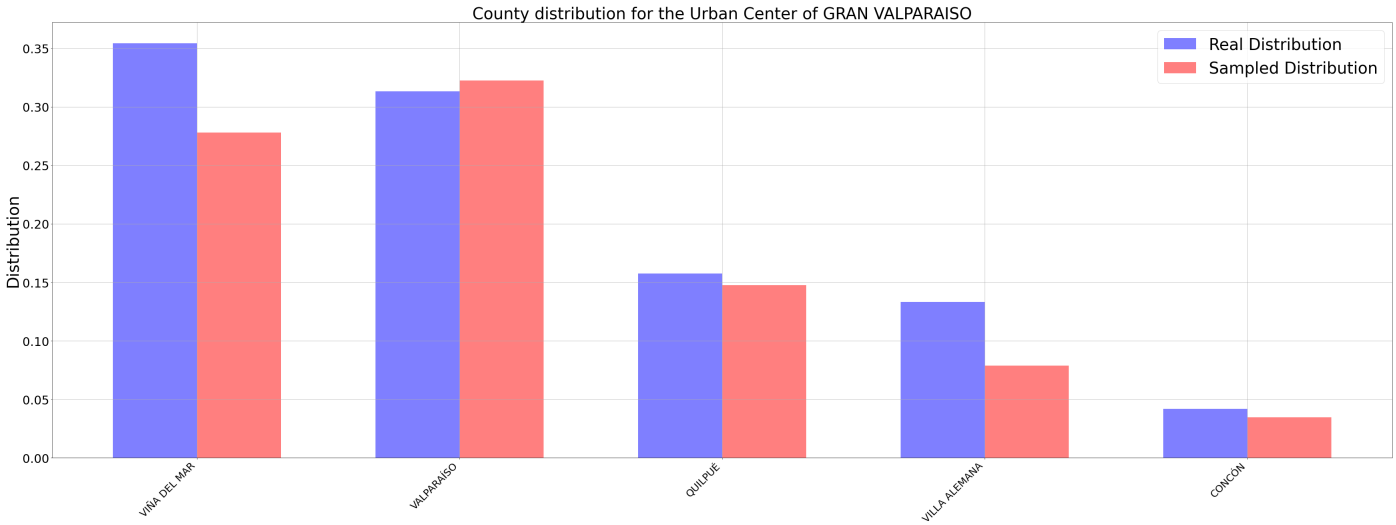

Figure 3.5: Gran Valparaíso Urban Center Distribution

This allows for a maximum data of 773 for Santiago, which again could be compared to what the maximum theoretical data shown with the 13 of July example to show the potential that flexibility could provide. This can be put into perspective as shown in Figure 3.6 below

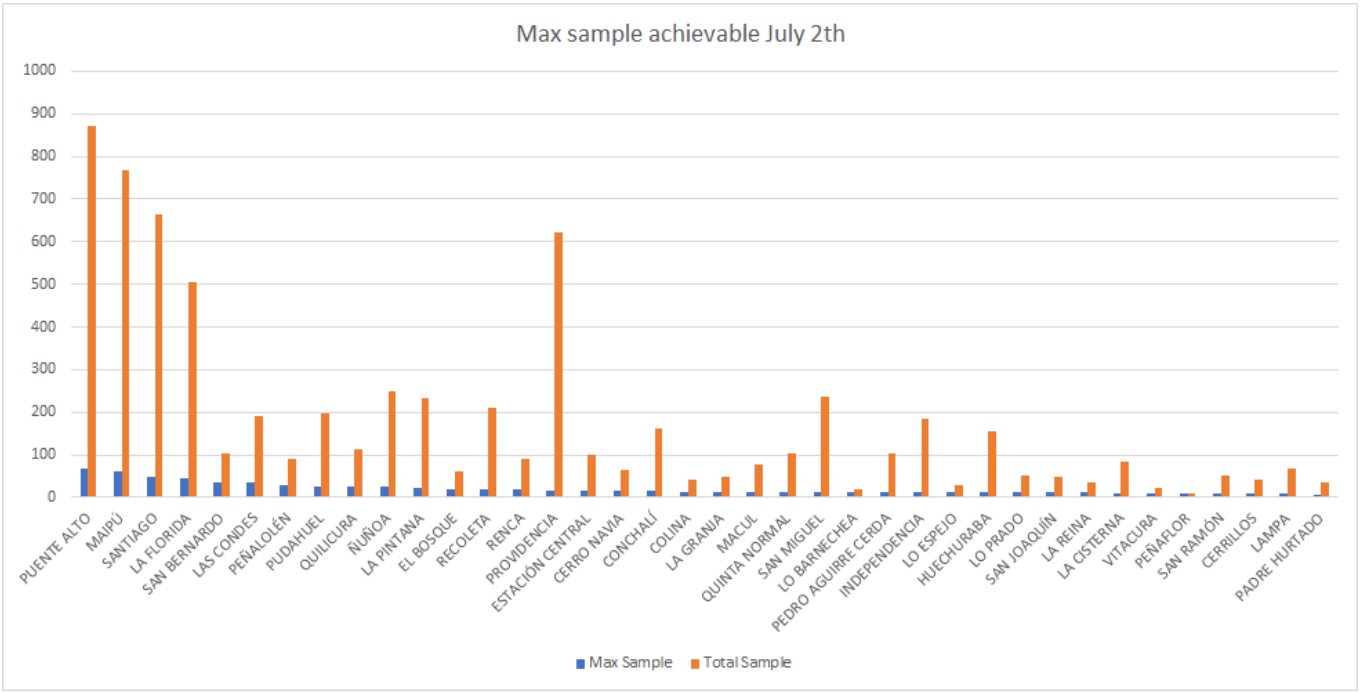

Figure 3.6: Gran Santiago Urban Center Distribution

Of the 66625 samples taken as of July 2nd 2021, only 53501 were useful for further analysis due to reasons such as not following the minimum age restriction, having incomplete or wrongly filled data in the formulary, having an invalid result in the test. For a reference of the characteristics of those 53501 samples, see Figure 3.7 in the appendix.

As for Santiago, which continued taking tests further on, the 23444 samples became

18491 when discarding tests that did not follow the criteria described above, the detail of these samples can be seen as Figure 3.8 in the appendix.

The Distribution for Santiago as of that date goes as follows:

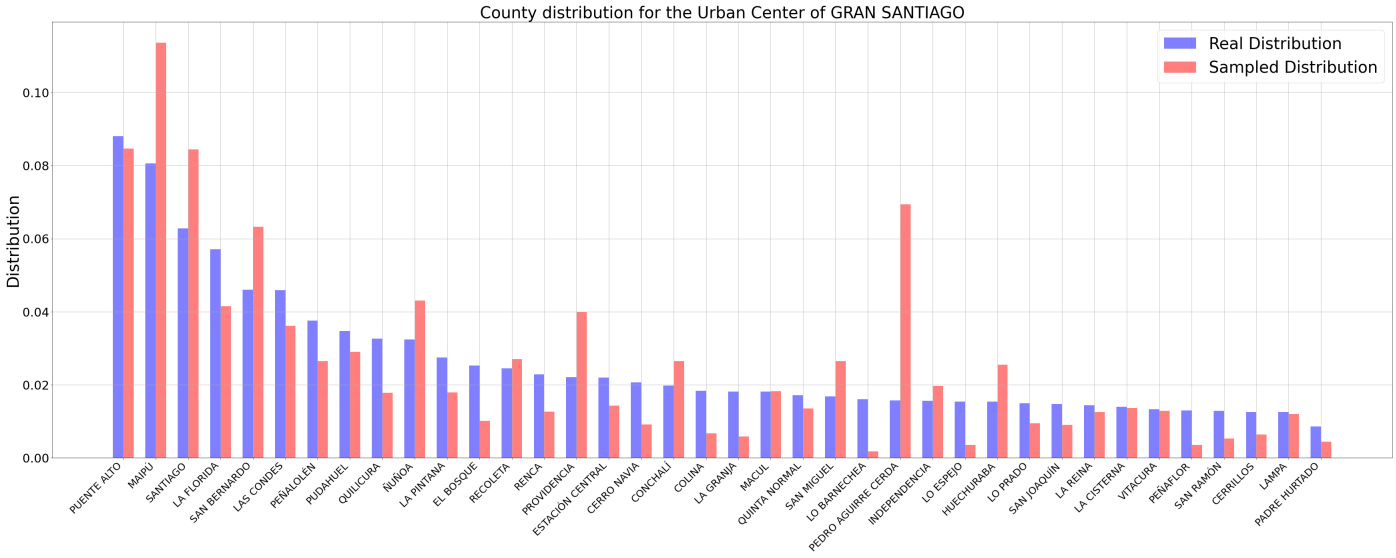

Figure 3.9: Gran Santiago Urban Center Distribution

It can be seen that, even though the over-representation of the most populated counties *Puente Alto*, *Santiago* and *La Florida* was reduced from July 2nd to December 13 (Figure 3.4 to Figure 3.9), the underrepresented counties have not gained enough representation as the solution was offering. This, again, due to a lack of flexibility when changing sites, over sampling the sites for being there a longer period of time. (See Figures 3.10 to 3.28 in the appendix to see the distributions up to July 2nd for all the other urban centers)

A few conclusions can be drawn from this section:

1. The total number of data collected as of December 13th was 103917 out of which 83198 samples belong to the urban centers of the study. This number could be improved by monitoring the average samples taken at each testing station and seeking to increase the number via three methods:
    - Reduce the amount of time the testing station is vacant (if that happens), by prioritizing sites with a higher mobility over sites with higher representativity.
    - Optimize the process at which individuals are tested to reduce the average sample time and consequentially increase the capacity of the station.
    - Increase the number of tests that can be done simultaneously, either by having more personnel or by increasing the amount of samples a health worker can take at the same time.
2. Allow higher flexibility in logistic and administrative factors should be a priority, as it limits the quality of the solution substantially.

The most important conclusion; the model can be implemented in a real scenario and achieve results that better understand the population's situation (for this particular case) and help decision-making.

Chapter 4

Seroprevalence Model

4.1 Seroprevalence

From March 12, 2021, 28 testing stations for SARS-CoV-2 IgG detection were installed in hotspots based on cellular-phone mobility tracking within the most populated cities in Chile. Each testing station was assigned to a Health Service with the corresponding jurisdiction over the area.
In each station, individuals were invited to do a lateral flow test (LFT) by finger prick voluntarily and respond to a questionnaire on sociodemographic characteristics, vaccination status (including type of vaccine if one was received), variables associated with SARS-CoV-2 exposure, and comorbidities.

For the Seroprevalence model, the main objective was to represent the presence of IgG for the population. This model used both biological and non-biological traits of the population as factors that can be related to variation in the presence of IgG for both vaccinated and unvaccinated individuals.
For this section, individuals were excluded if they did not live in any urban center in Chile (noting the fact that there are certain counties that do not belong to any). For this Section, individuals were excluded if they were younger than 18 years, had no declared gender, had an invalid IgG test result, had previously tested positive for SARS-CoV-2 infection on PCR, could not recall their vaccination status, immunization other than CoronaVac, BNT162b2 or ChAdOx1

Data reported for all the following chapters corresponds to people tested up to July 2nd, 2021 for every urban center other than *Gran Santiago*, for which we report data collected up to December 13th.

4.2 Model

We model the probability of the presence of IgG assuming a perfect test. We let $P(x)$ be the result of a perfect test. Note that IgG can come both from a vaccine or by contracting the virus.
In this section we use the following notation/definitions

- y: relevant demographic variables such as Age, Gender, and Comorbidities
- s: vaccine related variables such as vaccination status, type of vaccine, dates after vaccination, etc.
- $x = (y, s)$: individual's profile (y demographic variables, s vaccination status).
- $P_n(y)$: probability of detectable IgG due to virus exposition and not to the vaccine.
- $P_v(x)$: probability of detectable IgG due to vaccination.
- $v(x) :=$ type of vaccine.
- V: Set of vaccines.

For an unvaccinated individual we set, $x = (y, \emptyset)$, and define $P_v(y, \emptyset) = 0$. With this:

$$P(x) \quad = \quad P_v(x) + (1 - P_v(x)) \cdot P_n(y)$$

Note that the only vaccines in question are CoronaVac and BNT162b2, then further analysis can be made by separating both of them. The model is then:

$$\begin{aligned} P(x) \quad = \quad & \mathbb{1}_{\{v(x)=CoronaVac\}} \cdot P_{vCoronaVac}(y, s) \\ & + \mathbb{1}_{\{v(x)=BNT162b2\}} \cdot P_{vBNT162b2}(y, s) \\ & + (1 - \mathbb{1}_{\{v(x)=CoronaVac\}} \cdot P_{vCoronaVac}(y, s)) \\ & \cdot (1 - \mathbb{1}_{\{v(x)=BNT162b2\}} \cdot P_{vBNT162b2}(y, s)) \cdot P_n(y) \end{aligned}$$

Using (Lancet,2021)[11] paper, we calibrate

$$P_v(y, s) = \frac{\exp(f(y, s))}{1 + \exp(f(y, s))} \quad , \quad f(y, s) = \log \frac{P_v(y, s)}{1 - P_v(y, s)}$$

Where $f(y, s)$ depends only on s, and the components in y associated with age, gender, and comorbidities.

With regard to $P_n(y)$, it depends on variables such as: age, gender, comorbidities, perception of exposition, as well as the time frame and place of where the test realized. That is,

$$f(y, s) = \beta_t^v + \beta_{gender}^v + \beta_{age}^v + \beta_{comorb}^v$$

Here v represents the vaccine, t the number of fortnights since the first dose

For the unvaccinated,

$$P_n(y) = \frac{\exp(g(y))}{1+\exp(g(y))} \quad , \quad g(y) = \log\frac{P_n(y)}{1-P_n(y)}$$

Where $g(y)$ depends only on relevant demographic variables represented by y.

$$g(y) = \alpha_t + \alpha_{gender} + \alpha_{age} + \alpha_{comorb} + \alpha_{freq} + \alpha_{exp}$$

Then $h(y,s)$ has both α and β to account for variables represented in $f(y,s)$ and $g(y)$

$$h(y,s) = \alpha_t + \alpha_{gender} + \alpha_{age} + \alpha comorb + \alpha freq + \alpha_{exp} + \sum_{v\in V} \mathbb{1}_{\{v(x)=v\}} \cdot (\beta_t^v + \beta_{comorb}^v)$$

With,

$$P(y,s) = \frac{\exp(h(y,s))}{1+\exp(h(y,s))} \quad , \quad h(y,s) = \log\frac{P(y,s)}{1-P(y,s)}$$

This model was applied for two datasets, the first one considered data from the 6 most populated urban centers (that contain more than 50% of the population) from March 12th to July 2nd and added the variable *Urban Center* to the definition of $P_n(y)$, and the second one considered only the capital (*Gran Santiago*), using data from March 12th to December 13th and variables such as *Type of transport* and *Frequency of Transport* were added to $P_n(y)$.

Because *Gran Santiago* had data up to a later date than the rest of the urban centers, a different model could be made that included information that was collected for samples in later stages. This happened because as time passed, Chilean authorities started supplying booster shots for people. This booster doses could either be homologous or heterologous, so another model measuring the effect of the booster dose could be made. Since the most common vaccine was CoronaVac (as can be seen in Figures 3.7 and 3.8 of the appendix) this vaccine could be further disaggregated into boosted and not boosted samples, as it had a better chance of showing statistical significance. The main booster vaccines used in people that had CoronaVac as their first two doses were ChAdOx1, and BNT162b2.

The model then had the following modification:

$$P(x) = P_{vCoronaVac}(y,s) + P_{vBNT162b2}(y,s) + (1-P_{vCoronaVac}(y,s)) \cdot (1-P_{vBNT162b2}(y,s)) \cdot P_n(y)$$

Where:

$$\begin{aligned} P_{vCoronaVac}(y,s) &= P_{vChAdOx1}(y,s) + P_{vBNT162b2}(y,s) \\ &\quad + (1-P_{vChAdOx1}(y,s)) \cdot (1-P_{vBNT162b2}(y,s)) \cdot P_{NoBooster}(y) \end{aligned}$$

4.2.1 Bayesian Approach

We used a Bayesian approach, which helps the perception that taking into consideration information of individuals that are alike would help when building a model to predict positivity.
Then the logistic regression would be a Discriminative model, where we can directly estimate the posterior probability with the data and use maximum likelihood to estimate the parameters
For this particular case, because there was no prior notion about the distribution of the variables, a Normal(0,100) distribution was set for all of them as well as setting all parameters to return a real value.
Marcov Chain Monte Carlo (MCMC) was used as the algorithm to estimate the model's parameters
Lastly, the simulations associated were run for 1000 iterations and four chains for all models except the last one, that given the complexity and time of resolution, only had two chains. We used a Laptop With an *Intel(R) Core(TM) i7-950H CPU @ 2.60GHz*, simulations took from 6 hours to 30 hours depending on the complexity of the model.

4.3 Results

4.3.1 Model Results

Instance I: For the first instance the variables studied are as follows,

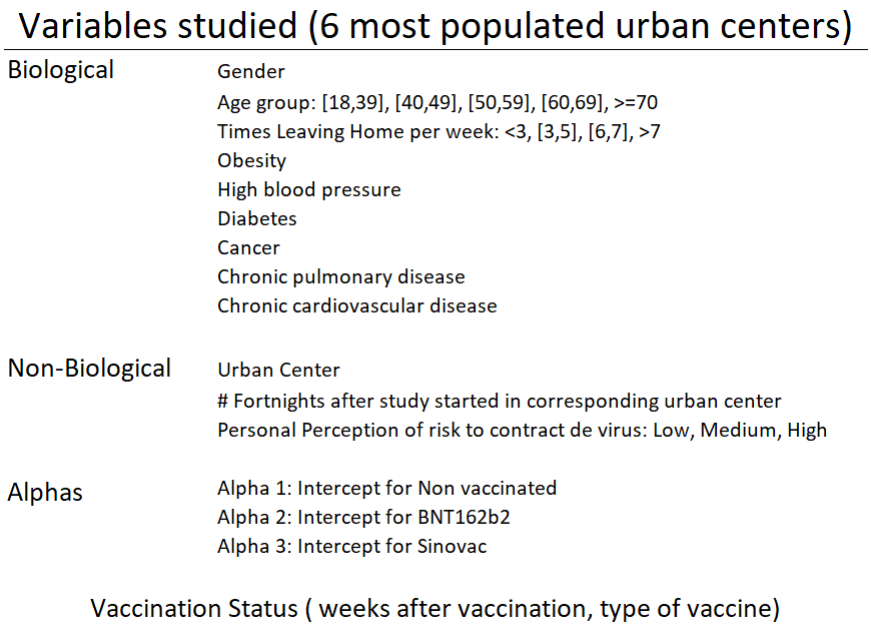

Figure 4.1: Variables 6 most populated urban centers

Figure 4.1 shows the betas of the model for the main urban centers from March 12th till July 2nd, can be seen as Figure 4.2 in the appendix.

A non-exhaustive Figure follows to show some of the statistically relevant variables:

Variable	Mean	SD	2.50%	97.50%	Variable	Mean	SD	2.50%	97.50%
Fortnight of study * Valparaíso	0.3224	0.0437	0.2362	0.4091	BNT162b2 * Fortnight 4 After Vac.	2.817	0.3329	2.204	3.544
Fortnight of study * Concepción	0.1406	0.06379	0.01434	0.2693	BNT162b2 * Fortnight 5After Vac.	3.339	0.5238	2.379	4.422
Fortnight of study * Valdivia	0.2361	0.0497	0.1406	0.3343	BNT162b2 * Fortnight 6 After Vac.	2.684	0.4795	1.796	3.7
Fortnight of study * Puerto Montt & Varas	0.2229	0.04722	0.127	0.3132	BNT162b2 * Fortnight 7 After Vac.	3.987	1.348	1.945	7.25
Valparaíso	-1.391	0.2251	-1.837	-0.9597	BNT162b2 * Fortnight 8 After Vac.	3.503	1.28	1.565	6.506
Concepción	-1.647	0.2846	-2.236	-1.095	BNT162b2 * Fortnight 10 After Vac.	2.651	1.384	0.4458	5.92
Puerto Montt & Varas	-0.9964	0.2109	-1.397	-0.5901	CoronaVac * Gender	0.2505	0.04711	0.1622	0.3446
Age Range (50,59)	0.3954	0.12	0.1581	0.6252	CoronaVac * Age Range (50,59)	-0.1929	0.07063	-0.3338	-0.05096
Age Range (60,69)	0.5754	0.1681	0.2269	0.8886	CoronaVac * Age Range (60,69)	-0.3309	0.09613	-0.5209	-0.148
Age Range (70+)	0.6043	0.2355	0.1041	1.022	CoronaVac * Age Range (70+)	-0.6874	0.1461	-0.9775	-0.414
Times Leaving Home per week (7+)	-0.7089	0.1589	-1.024	-0.4096	CoronaVac * Fortnight 1 After Vac.	0.5869	0.07417	0.4463	0.736
Medium Perception of Risk	0.2655	0.07643	0.1229	0.417	CoronaVac * Fortnight 2 After Vac.	0.9909	0.07358	0.8542	1.14
High Perception of Risk	1.166	0.08675	0.9975	1.337	CoronaVac * Fortnight 3 After Vac.	1.045	0.07491	0.9058	1.197
BNT162b2 * Age Range (60,69)	-1.47	0.3048	-2.072	-0.8763	CoronaVac * Fortnight 4 After Vac.	0.8524	0.08028	0.6968	1.011
BNT162b2 * Age Range (70+)	-2.934	0.5743	-4.178	-1.859	CoronaVac * Fortnight 5After Vac.	0.5581	0.09282	0.3758	0.7408
BNT162b2 * Fortnight 1 After Vac.	1.694	0.1467	1.41	1.969	CoronaVac * Fortnight 10 After Vac.	-381700	461200	-1434000	-10940
BNT162b2 * Fortnight 2 After Vac.	2.481	0.2214	2.061	2.922	CoronaVac * Obesity	-0.1933	0.09157	-0.3739	-0.01369
BNT162b2 * Fortnight 3 After Vac.	2.781	0.3	2.22	3.399	CoronaVac * Diabetes	-0.2733	0.06921	-0.4066	-0.1412

Figure 4.3: Statistically relevant variables for Instance I [non-exhaustive]

Our results show a positive statistical significance in values between urban centers by fortnight. There are several factors that could be causing this. Mainly indirect factors such as population density, the distribution of infection that could affect the future spread of the virus and the use of different methods of transportation and other contact related factors that could not be captured by the other variables in this particular model. This can be further seen when comparing the *Fortnight of study * Urban Center* value with the urban center value on its own, seeing that the latter is negative in comparison to the base while the other improves fortnight by fortnight. As for the *Age Range* in comparison to the base (18-39) a positive significant difference can be observed in values as the age increases, meaning that for older ages, there is a decrease in the positivity of the test. For the fortnights 1 through 6 there can be seen an increase in the positivity in comparison to the base case, meaning that as the study continued, more positivity was expected, attributed arguably to the indirect situation of the number of people with the virus as well as many other non-controllable factors. For comorbidities, we can see a significant decrease in positivity for obese and diabetic participants in comparison to the base group of the same category.
Instance II:

As of December 13th for *Gran Santiago's* urban center the following Figure shows the variables analyzed, and the betas of the model can be seen as Figure 4.5 in the appendix:

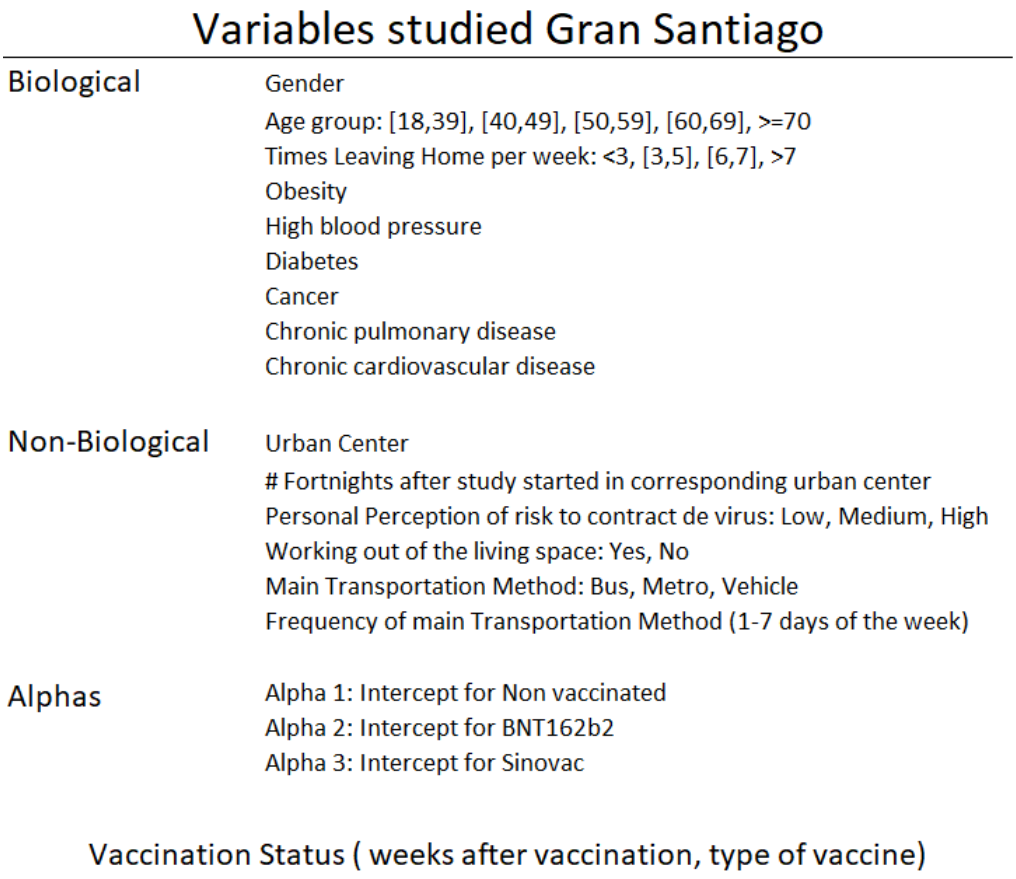

Figure 4.4: Variables December 13th Gran Santiago

A non-exhaustive Figure follows to show some of the statistically relevant variables:

Variable	Mean	SD	2.50%	97.50%	Variable	Mean	SD	2.50%	97.50%
Working Outside (Yes)	-0.114	0.101	-0.315	0.087	CoronaVac * Gender	0.257	0.082	0.092	0.420
Transport by Metro	0.694	0.250	0.198	1.180	CoronaVac * Age Range (70+)	0.587	0.177	0.213	0.902
Study Fortnight	0.112	0.011	0.090	0.132	CoronaVac * Fortnight 1 After Vac.	0.775	0.148	0.494	1.070
Times Leaving Home per week (3,5)	0.281	0.129	0.024	0.526	CoronaVac * Fortnight 2 After Vac.	0.911	0.151	0.627	1.210
Times Leaving Home per week (7+)	-0.607	0.250	-1.100	-0.142	CoronaVac * Fortnight 3 After Vac.	0.793	0.153	0.498	1.110
High Perception of Risk	1.040	0.113	0.814	1.270	CoronaVac * Fortnight 4 After Vac.	0.699	0.150	0.399	0.990
BNT162b2 * Age Range (60,69)	-1.670	0.404	-2.510	-0.909	CoronaVac * Fortnight 5After Vac.	0.435	0.149	0.149	0.739
BNT162b2 * Fortnight 1 After Vac.	1.140	0.330	0.492	1.800	CoronaVac * Fortnight 8 After Vac.	-0.491	0.171	-0.828	-0.165
BNT162b2 * Fortnight 2 After Vac.	2.010	0.463	1.100	2.950	CoronaVac * Fortnight 9 After Vac.	-0.615	0.186	-1.000	-0.243
BNT162b2 * Fortnight 3 After Vac.	1.740	0.429	0.934	2.630	CoronaVac * Fortnight 10 After Vac.	-0.621	0.213	-1.070	-0.245
BNT162b2 * Fortnight 4 After Vac.	2.550	0.578	1.500	3.810	CoronaVac * Fortnight 11 After Vac.	-1.140	0.281	-1.770	-0.664
BNT162b2 * Fortnight 5After Vac.	4.320	1.311	2.370	7.460	CoronaVac * Fortnight 13 After Vac.	0.661	0.184	0.297	1.000
BNT162b2 * Fortnight 6 After Vac.	2.690	0.675	1.490	4.140	CoronaVac * Fortnight 14 After Vac.	1.390	0.197	1.010	1.780
BNT162b2 * Fortnight 7 After Vac.	2.160	0.606	1.100	3.450	CoronaVac * Fortnight 15 After Vac.	1.950	0.210	1.560	2.380
BNT162b2 * Fortnight 8 After Vac.	2.220	0.640	1.060	3.620	CoronaVac * Fortnight 16 After Vac.	2.270	0.243	1.780	2.740
BNT162b2 * Fortnight 9 After Vac.	1.520	0.591	0.442	2.720	CoronaVac * Fortnight 17 After Vac.	2.540	0.350	1.880	3.230
BNT162b2 * Fortnight 10 After Vac.	0.977	0.402	0.206	1.770	CoronaVac * Fortnight 18 After Vac.	1.990	0.331	1.370	2.650
BNT162b2 * Fortnight 11 After Vac.	2.220	0.681	0.998	3.720	CoronaVac * Hypertension	-0.260	0.074	-0.397	-0.117
BNT162b2 * Fortnight 15 After Vac.	1.470	0.761	0.023	3.050	CoronaVac * Diabetes	-0.216	0.094	-0.398	-0.031
BNT162b2 * Fortnight 16 After Vac.	79.900	61.240	5.280	233.000					

Figure 4.6: Statistically relevant variables for Instance II [non-exhaustive]

Transportation: A statistical difference can be seen in the positivity of people who use the subway as a transportation method, not achieving a statistical difference in the frequency of this usage. This could be explained by the high interaction that transportation method presents, increasing the chances of contracting the virus and therefore presenting more positivity in the test.

Fortnights of study: There is a positive statistical difference in the number of fortnights after the study started, this could be caused by the general infection of the percentage of population as time passes. Another positive statistical difference can be seen in the people that went out 3-5 times a week compared to the people that left their homes less than 3 times a week, this could be explained by how leaving home increases contact with other people, increasing the chances of contracting the virus and therefore having antibodies for it.

Frequency of movement: A negative statistical difference can be seen in the group of people that went out more than 7 times a week, which can not be explained in a direct way since the distribution of method of transportation (which would be the most relevant explanation) is similar between all the frequencies as can be seen in the following Figure noting that the base case is walking as a method of transportation.

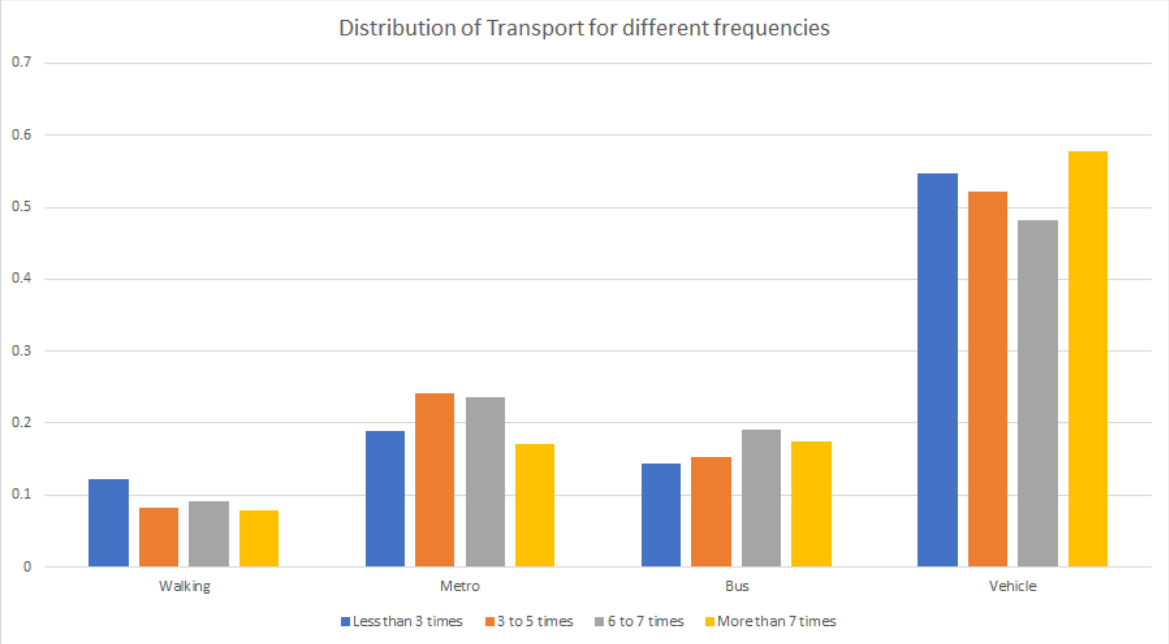

Figure 4.7: Distribution of transport for each frequency

Personal Evaluation: A positive statistical difference in seroprevalence can be observed in individuals that evaluate themselves as having a high risk of contraction in comparison to individuals that evaluate themselves as low risk of infection. This is probably due to them actually getting infected and consequently having more defenses against the virus.

Age: All the age groups vaccinated with BNT162b2 have lower seroprevalence in comparison to the youngest age group (18-39) but only the 60-69 age group had statistical difference.

This can be due to the capacity of younger individuals to maintain more antibodies, as several studies suggest[12]

Fortnights after vaccination: A positive statistical difference can be found in the fortnights 1-11 and 15-16 after vaccination with BNT162b2 in comparison to the first one where the ones with no statistical difference were also positive and where no strict trend could be found between fortnights. Suggesting either a bigger number of antibodies generated at first, making it possible to still pass the thresh hold of positivity after several fortnights of decreasing antibodies, or a better maintenance of the generated antibodies with that particular vaccine.

Gender: A positive statistical difference can be found in women vaccinated with CoronaVac versus male individuals, which can be supported by other studies[8].

Age with CoronaVac: A positive statistical difference can be found in people of age 70+ vaccinated with CoronaVac in comparison to people from 18-39 years of age, with no direct interpretation.

Fortnights after vaccination with CoronaVac: A positive statistical difference can be found in people 1-5 and 13-18 fortnights after vaccination with CoronaVac and negative statistical difference in fortnights 8-11. This can be due to the creation of antibodies at first and later decrease of it due to time, later growing again because of the use of a Reinforcement vaccine.

Comorbidities: A negative statistical difference can be seen in people with High blood pressure and Diabetes in comparison to people without it and the same vaccine.

As for the model that had the booster shot as a factor the variables are the same as the model shown before, adding the Interaction with both booster vaccines as well as $Alpha4$ that represents the intercept for individuals that received ChAdOx1 as their vaccine and $Alpha5$ to represent the intercept for individuals that received BNT162b2 as their booster vaccine, the results for the betas can be seen as Figure 4.8 in the appendix

A non-exhaustive Figure follows to show some of the statistically relevant variables:

Variable	Mean	SD	2.50%	97.50%	Variable	Mean	SD	2.50%	97.50%
Transport by Metro	0.457	0.177	0.148	0.811	Booster ChAdOx1 * Fortnight 10 After Vac.	-16.900	6.532	-28.980	-6.137
Gender	0.201	0.051	0.098	0.299	Booster ChAdOx1 * Fortnight 11 After Vac.	-6.694	2.595	-11.230	-0.705
Study Fortnight	0.165	0.007	0.155	0.178	Booster ChAdOx1 * Fortnight 14 After Vac.	5.293	2.607	0.122	10.840
Age Range (40,49)	0.388	0.073	0.243	0.529	Booster ChAdOx1 * Fortnight 15 After Vac.	-9.684	2.091	-13.180	-7.004
Age Range (50,59)	0.503	0.057	0.378	0.594	Booster ChAdOx1 * Fortnight 17 After Vac.	-19.340	4.660	-27.460	-11.830
Age Range (60,69)	0.247	0.090	0.046	0.383	Booster BNT162b2 * Gender	-6.356	3.097	-12.040	-1.189
Age Range (70+)	0.477	0.107	0.286	0.657	Booster BNT162b2 * Age Range (40,49)	-17.170	10.046	-35.890	-3.062
Times Leaving Home per week (3,5)	0.285	0.082	0.135	0.433	Booster BNT162b2 * Age Range (50,59)	7.429	4.139	3.017	16.530
High Perception of Risk	0.481	0.097	0.297	0.641	Booster BNT162b2 * Age Range (70+)	-6.384	2.570	-9.758	-0.335
BNT162b2 * Age Range (60,69)	-4.427	1.636	-7.667	-2.381	Booster BNT162b2 * Fortnight 1 After Vac.	-4.053	2.572	-9.586	-0.714
BNT162b2 * Age Range (70+)	8.080	2.295	4.103	11.850	Booster BNT162b2 * Fortnight 2 After Vac.	15.500	4.926	8.000	22.300
BNT162b2 * Obesity	-6.979	3.911	-14.090	-1.476	Booster BNT162b2 * Fortnight 3 After Vac.	23.450	6.816	7.975	32.760
BNT162b2 * Hypertension	-7.022	3.634	-13.710	-1.134	Booster BNT162b2 * Fortnight 5After Vac.	7.606	3.818	1.869	12.720
BNT162b2 * Chronic Pulmonary D.	-8.913	4.495	-16.790	-3.350	Booster BNT162b2 * Fortnight 7 After Vac.	-9.278	3.221	-13.910	-3.712
BNT162b2 * Chronic Cardiovascular D.	11.170	2.628	4.349	15.040	Booster BNT162b2 * Fortnight 8 After Vac.	4.512	3.220	0.398	11.360
CoronaVac * Gender	-0.447	0.091	-0.595	-0.244	Booster BNT162b2 * Fortnight 9 After Vac.	-6.265	3.083	-12.800	-1.644
CoronaVac * Age Range (40,49)	0.191	0.111	0.023	0.422	Booster BNT162b2 * Fortnight 10 After Vac.	-11.170	8.245	-23.550	-0.729
CoronaVac * Hypertension	0.385	0.086	0.250	0.541	Booster BNT162b2 * Fortnight 11 After Vac.	10.660	3.997	3.069	18.750
CoronaVac * Diabetes	0.359	0.164	0.088	0.696	Booster BNT162b2 * Fortnight 12 After Vac.	12.710	4.767	1.865	19.690
CoronaVac * Chronic Cardiovascular D.	-0.539	0.224	-0.886	-0.151	Booster BNT162b2 * Fortnight 13 After Vac.	-12.140	2.998	-15.400	-5.197
Booster ChAdOx1 * Gender	-9.402	3.394	-14.520	-3.026	Booster BNT162b2 * Fortnight 15 After Vac.	-7.325	2.839	-14.970	-2.773
Booster ChAdOx1 * Age Range (40,49)	9.683	2.551	5.711	14.140	Booster BNT162b2 * Fortnight 16 After Vac.	8.720	4.108	1.873	16.190
Booster ChAdOx1 * Age Range (50,59)	7.471	2.867	1.360	13.040	Booster BNT162b2 * Fortnight 18 After Vac.	16.710	3.073	9.549	23.140
Booster ChAdOx1 * Fortnight 3 After Vac	-15.810	2.848	-21.890	-11.970	Booster BNT162b2 * Obesity	-13.290	5.880	-21.750	-2.259
Booster ChAdOx1 * Fortnight 5After Vac	-16.380	8.373	-32.880	-5.965	Booster BNT162b2 * Chronic Pulmonary D.	-17.760	5.817	-26.120	-6.489
Booster ChAdOx1 * Fortnight 7 After Vac	14.640	3.671	5.888	19.490					

Figure 4.9: Statistically relevant variables for Instance II-Booster [non-exhaustive]

Transportation: Just as the prior model showed, there can be seen a statistical difference in the positivity of people who use the Subway as a transportation method, not achieving a statistical difference in the frequency of this usage. This could be explained by the high interaction that transportation method presents, increasing the chances of contracting the virus and therefore presenting more positivity in the test.
Gender: There can be seen a positive statistical difference in gender where the base scenario corresponds to males, meaning female individuals are expected to have higher antibodies, which again can be supported by other studies[8].

Fortnights of study: Just as the model before there can also be seen a positive statistical difference in the number of fortnights after the study started, this could be caused by the general infection of the percentage of population as time passes.
Age: There can also be seen a positive statistical difference in every age group compared to the base case of 18 to 39 years of age, with no clear explanation.
Frequency of movement: Just as the model before, another positive statistical difference can be seen in the people that went out 3-5 times a week compared to the people that left their homes less than 3 times a week, this could be explained by how leaving home increases contact with other people, increasing the chances of contracting the virus and therefore having antibodies for it.

Personal Evaluation: Again, individuals that evaluate themselves as having a high risk of contraction present a positive statistical difference in seroprevalence in comparison to the people that evaluate themselves as low risk of infection. This is probably due to them actually getting infected and consequently having more defenses against the virus.

Age and BNT162b2: Individuals with 60-69 years of age vaccinated with BNT162b2 show statistically lower seroprevalence in comparison to the youngest age group (18-39). The group of more than 70 years of age present statistically higher seroprevalence with no clear explanation.
Fortnights of BNT162b2As for fortnights after being vaccinated with BNT162b2, a high number of fortnights present a higher statistically difference versus the base group with no fortnights after vaccination, this could help show a positive progression of the build of antibodies in a short time frame.
Comorbidites and BNT162b2: Individuals with obesity, hypertension, chronic pulmonary disease have a negative statistical difference. Individuals with chronic cardiovascular disease present positive statistical difference versus the base group of vaccinated individuals with no comorbidities.
Fortnights and CoronaVac: For individuals vaccinated with CoronaVac there can be seen a negative statistical difference in antibodies the first 4 fortnights, this would no have a clear explanation. After the sixth fortnight, a positive statistical difference can be seen till the fortnight number 12. This could show the effect of antibodies generated by the vaccine in the short term, and since the 15th fortnight shows negative seroprevalence, a decline could be seen in antibodies in a longer time period, justifying the supply of booster shots.
Comorbidities and CoronaVac: A positive seroprevalence can be observed for individuals with hypertension and diabetes and negative seroprevalence for individuals with chronic cardiovascular disease, contrary to what was observed with individuals that received BNT162b2 with no clear explanation.
For individuals that received booster shot, both individuals that received ChAdOx1 and individuals that received BNT162b2, present negative statistical difference in gender, being male the base. All other variables that show significant difference for individuals with booster shot can not be inferred to have any direct cause.

4.3.2 Estimating Seroprevalence

With the previous information, an estimate of the population's antibodies can be made, as P(x) is different for people with different profiles, then for an accurate estimate of the population, a representative sample of the population can be withdrawn from the used data.
For this model, important factors for this estimation were comorbidities, gender, age, proportion of people vaccinated with CoronaVac and BNT162b2. The proportions of these factors were estimated using different sources such as the census to get the closest to the population's real proportion as possible The model is then as follows

Parameters:

- C Set of comorbidities
- A Set of age ranges
- I Set of Samples.
- I_c Subset of individuals with comorbidity $c \in$ C

- I_G Subset of female individuals.
- I_a Subset of individuals with age range $a \in$ A
- I_{Pf} Subset of individuals vaccinated with BNT162b2.
- I_{Co} Subset of individuals vaccinated with CoronaVac.
- R_c Chilean proportion of individuals with comorbidity $c \in$ C
- R_G Chilean proportion of female individuals.
- R_a Chilean proportion of individuals with age range $a \in$ A
- R_{Pf} Chilean proportion of individuals vaccinated with BNT162b2.
- R_{Co} Chilean proportion of individuals vaccinated with CoronaVac.

Note the missing proportions such as the male gender and individuals without comorbidities would automatically fill the missing proportion of samples taken.

$$x_i = \begin{cases} 1 & \text{if sample i is selected} \\ 0 & \sim \end{cases}.$$

The optimization problem is then as follows

$$\begin{array}{rll} & \max n & \\ s.t. & \sum_{i \in I_s} x_i & \leq R_s\, n \\ & s & \in \{C, G, A, Pf, Co\} \\ & \sum_{i \in I} x_i & = n \\ & x_i & \in \{0, 1\} \\ & n & \in N. \end{array}$$

Where n is the number of representative samples that we are trying to maximize.

Out of the 17935 samples that were left after filtering vaccines other than CoronaVac and BNT16b2 a sample of size 1727 achieved the restrictions perfectly.
The second model was applied, using each of this sample size 1727, P(x) was calculated by calculating each component of the equation

$$P(x) = P_{vCoronaVac}(y,s) + P_{vBNT162b2}(y,s) + (1 - P_{vCoronaVac}(y,s)) \cdot (1 - P_{vBNT162b2}(y,s)) \cdot P_n(y)$$

Noting that both $P_{vCoronaVac}(y,s)$ and $P_{vBNT162b2}$ fulfill the equation

$$\begin{aligned} P_v(y,s) &= \frac{\exp(f(y,s))}{1 + \exp(f(y,s))} \\ f(y,s) &= \log \frac{P_v(y,s)}{1 - P_v(y,s)} \end{aligned}$$

as $P_n(y)$ fulfills

$$\begin{aligned} P_n(y) &= \frac{\exp(g(y))}{1+\exp(g(y))} \\ g(y) &= \log \frac{P_n(y)}{1-P_n(y)} \end{aligned}$$

Using all the variables in $f(y,s)$ and $g(y)$ described before.
After calculating $P(x)$ for each individual of the sample, an average $P(x)$ could be calculated (which in this case was 0,73) as well as graphs showing the evolution of antibodies for this representative data through time.

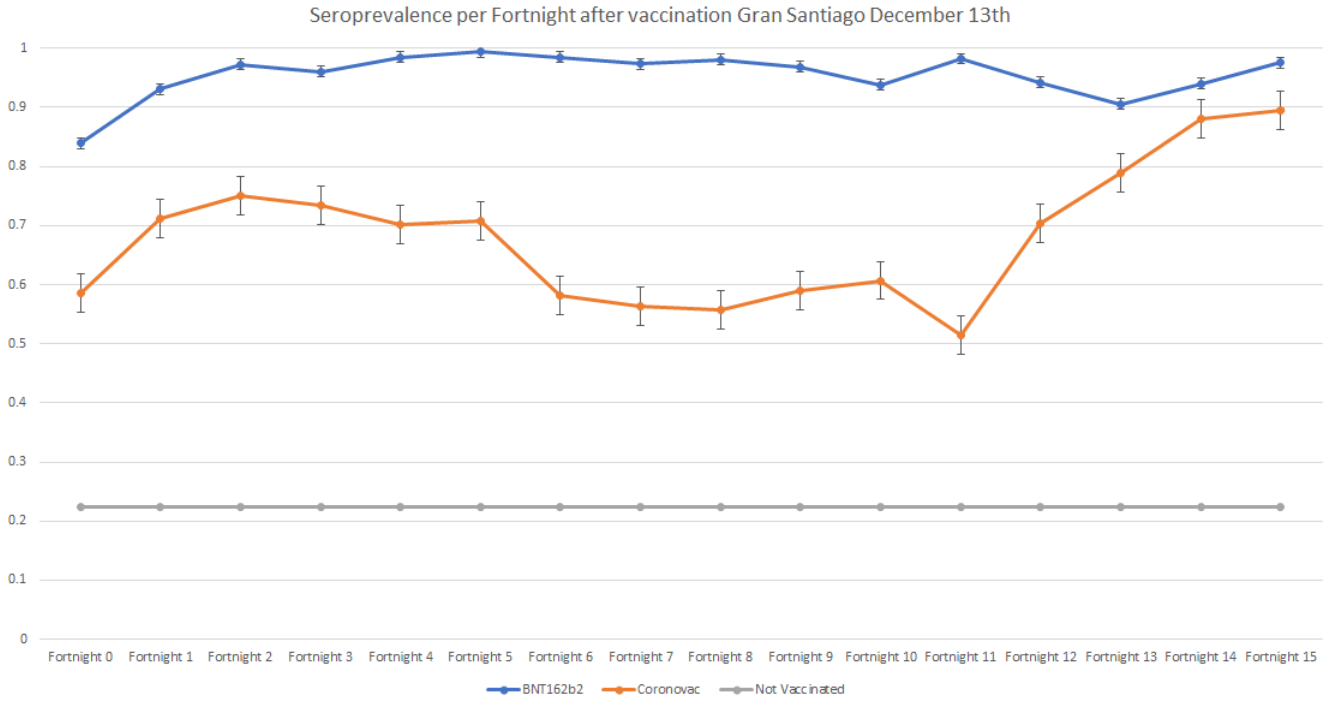

Figure 4.10: Antibodies After Vaccination Gran Santiago

This graph shows the estimate $P(x)$ for individuals of these different vaccination statuses and for each corresponding fortnight after vaccination.
A spike of antibodies can be observed after the eleventh fortnight, this could be explained by the introduction of booster shots to the population months after the first two doses.
To filter this factor out, the model was then applied to samples filtering those who had booster dose (reducing the number to 12055) resulting in a sample size of 1428 that achieved the restrictions perfectly.
Again calculating $P(x)$ for individuals of these different vaccination status and for each corresponding fortnight after vaccination and representing it in the following graph

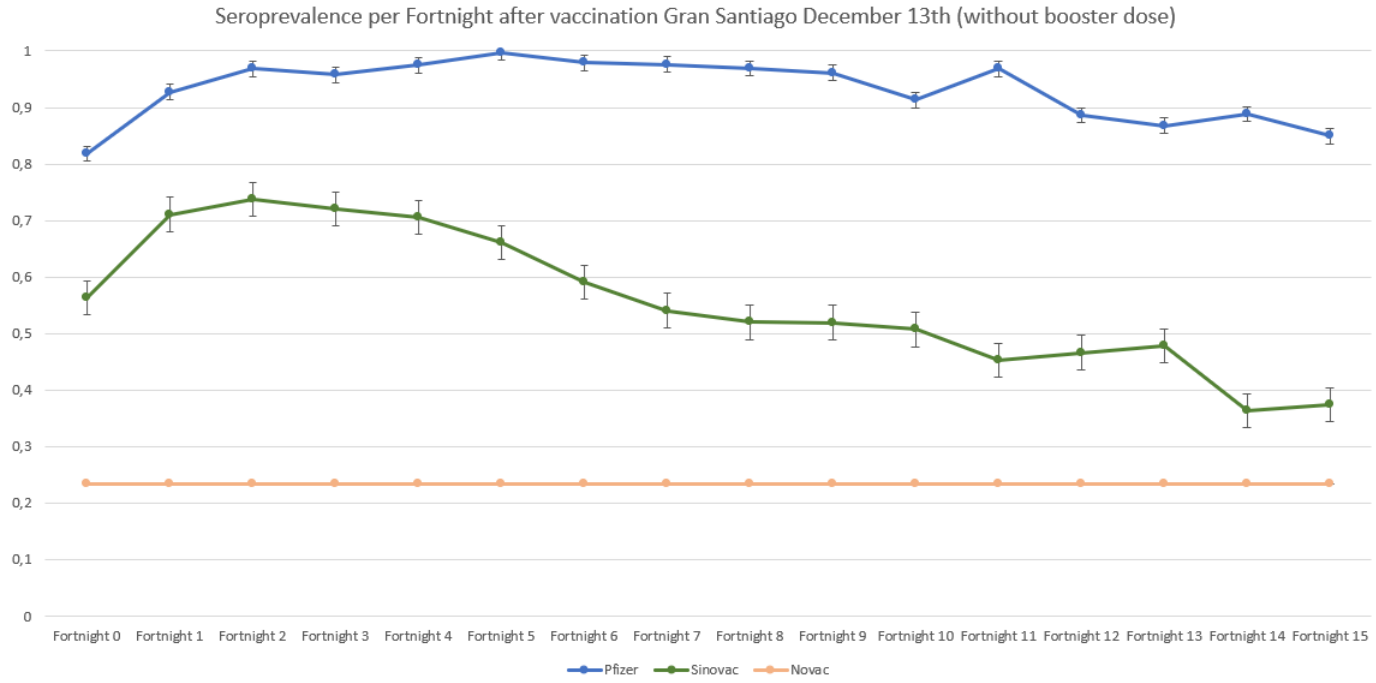

Figure 4.11: Antibodies After Vaccination Gran Santiago

The following graph shows the evolution of the virus by number of fortnights after the study started for the first case that included individuals with booster dose:

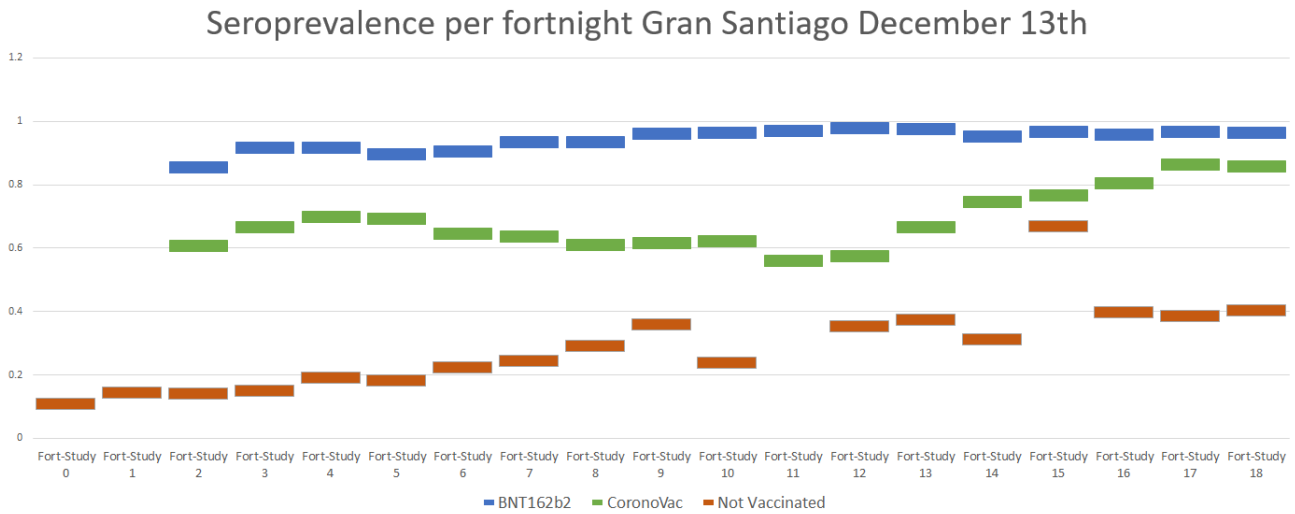

Figure 4.12: Antibodies After the start of the study Gran Santiago

A slight up-trend can be observed for each vaccination status noting that this is, as the model that constructed it suggests, due to the contraction of the virus or antibodies generated by the vaccine. Noting in particular that non vaccinated people can only contract the virus, and as time passes eventually get higher antibodies consequentially.
In both Figure 4.10 and Figure 4.12 a notable difference can be observed between vaccinated

and not vaccinated individuals, helping to lay the usefulness of the vaccine in evidence, as well as a considerable difference in the vaccines as BNT162b2 presents higher seroprevalence through time.

As of the seroprevalence per fortnight filtering samples with booster dose, the graph is as follows

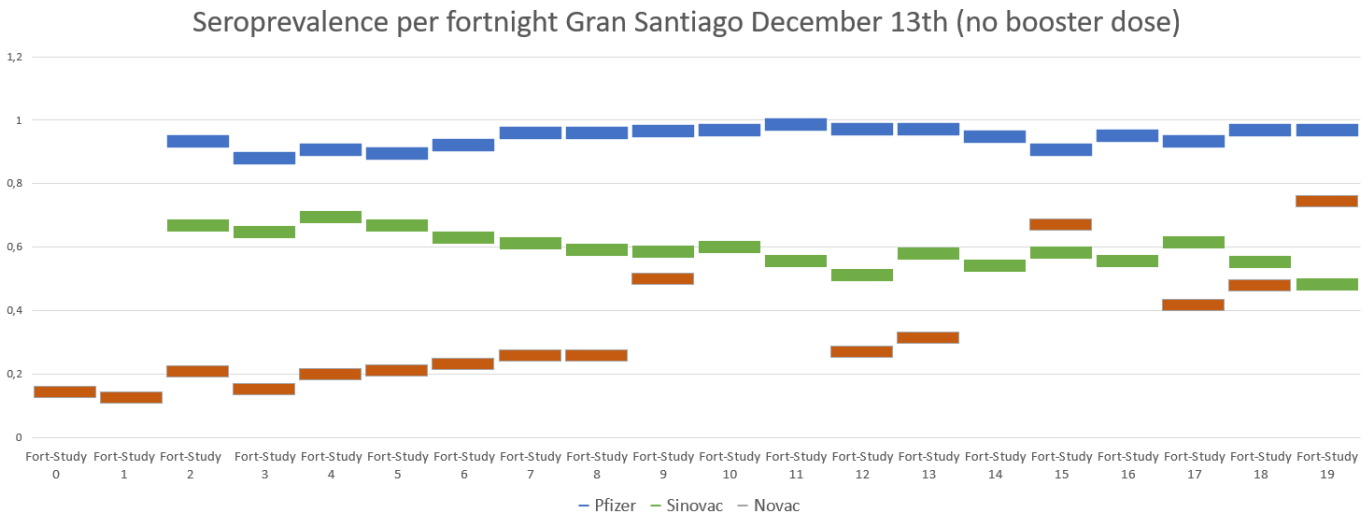

Figure 4.13: Antibodies After the start of the study Gran Santiago

In this Figure, the main difference from Figure 4.12 is the downtrend represented by CoronaVac that could be attributed to the lack of a booster dose.

Chapter 5

Conclusion and General Comments

5.1 Conclusion

This thesis resolves the matching problem of sample testing to a theoretical population data by the usage of a mixed integer optimization model as well as the usage of this data with a logistic regression and simulation in order to estimate the presence of antibodies in the population using the sample.
This general objective was achieved by improving the usable sample size collected by a considerable amount, as well as providing insight into how different biological and non-biological variables could affect the presence of antibodies in individuals.
In particular, the first part of the thesis corresponds to the optimization model based on weekly analysis of national mobile phone mobility data, facilitated by Chile's largest telecommunications agency (*Empresa Nacional de Telecomunicaciones*, Santiago, Chile), to select sites with high traffic volume and wide county-level distribution of people.

The model sought maximizing the representative sample size that corresponds to the geographical distribution at county granularity for each urban center. Taking into consideration parameters such as census zones, counties, expected number of samples, population of the county, number of testing sites and current number of samples obtained and using an auxiliary variable to describe the size of the representative sample for the urban center. Then, a binary variable was used to check if a site was assigned to a census zone in the corresponding time frame.
The model then maximized the auxiliary variable with constrains in the number of assigned sites and the distribution that was sought.
This theoretical model obtained results with significant improvements in the maximum representative data. As an example Gran Santiago could have enabled the usage of 84% of the results (12957 out of 15404 samples), an important improvement considering the samples prior to the implementation of the model that had a usable data of 15% (1182 out of 7902)
The model could not be implemented at it's full potential as changing sites required administrative paperwork behind which did not allow the frequency of change that would otherwise be possible, as well as an additional layer of requirements for the selected sites such as bathrooms and safety that hindered the possibility of sites to be assigned. The model was then

planned in a weekly basis and even so could not change as frequently as desired, achieving consequently worse results. This, could be different if the study is recreated in the future, more on this in the next section (5.2).

The second model sought to predict the probability to detect the presence of IgG assuming a perfect test. It used biological variables such as age and comorbidities, as well as non-biological variables such as method of transportation and frequency of transportation.
It then represented the probability to detect the presence of IgG as a combination of the biological and non-biological variables expressed as a logistic regression.
The model used a bayesian approach and a Marcov Chain Monte Carlo algorithm to fit the model with simulations, obtaining results that show a notable difference in the expected presence of IgG between vaccinated and not vaccinated individuals as well as a considerable difference in the vaccines as BNT162b2 presents higher seroprevalence through time putting the use-fullness of the vaccine in evidence.
Lastly, these results were put into perspective by estimating the expected seroprevalence for the population with a representative sample in order to visualize the seroprevalence through time.

5.2 Future work

If this study or a similar one was replicated in the future, some important factors should be taken into consideration.
First, the localization model has an important logistic and administrative limiting factor that should be discussed and explored to achieve the most flexible implementation possible. This would allow results that would match what the model is able to achieve.
As section 3.4 shows, the impact of a more flexible implementation is considerable in the maximum representative sample achievable.
This can help better decision-making as more specific insight can be retrieved from the corresponding data.

As this data also helps the model used in the logistic regression, better results can be expected there as well.
Another improvement for the second model is the use of a better computer and looking into a more optimized code as it could improve the resolution time and allow a bigger amount of iterations and chains for simulation, that in similar studies could improve the insight achievable of some specific variables.

As a last thought, this thesis and project as a whole generated an impact, as many others in the decision-making in times of crisis. COVID-19 impacted the whole world, and the ability to react and create this kind of projects is truly remarkable.
As a future reference, not only this, but many other studies will help even better speeds in

implementation, noting that the first weeks and months of crisis are key for the development of the situation that the population is facing.
Therefore, future possible implementations of the models shown in this thesis, if in a similar context, should seek a fast and flexible implementation, not only achieving better results, but more importantly a possible even bigger impact in the world.

Bibliography

[1] Instituto nacional de estadísticas. population projections. https://www.ine.cl/estadisticas/sociales/demografia-y-vitales/, Last visited August 2021.

[2] Instituto sistemas complejos de ingeniería. seroprevalence. 2021. https://covidanalytics.isci.cl/seroprevalencia/, Last visited December 2021.

[3] Gurobi optimizer reference manual. https://www.gurobi.com/documentation/9.1/refman/index.html, Visited August 2021.

[4] Informe evaluación de ensayos comerciales subdepartamento de enfermedades virales. https://covidanalytics.isci.cl/wp-content/uploads/2021/03/Informe_ISP.pdf, Visited Oct 2021.

[5] Department of statistics and information, ministry of health chile. official figures covid-19 chile. 2021. https://deis.minsal.cl/, visited September 2021.

[6] Servicios de salud. https://www.minsal.cl/servicios-de-salud/, visited September 2021.

[7] Dami A. Collier, Isabella A. T. M. Ferreira, Prasanti Kotagiri, et al., and N. I. H. R. BioResource4. Age-related immune response heterogeneity to sars-cov-2 vaccine bnt162b2. *Nature*, 596(7872):417–422, Aug 2021.

[8] Stephanie Fischinger, Carolyn M. Boudreau, Audrey L. Butler, Hendrik Streeck, and Galit Alter. Sex differences in vaccine-induced humoral immunity. *Seminars in Immunopathology*, 41(2):239–249, Mar 2019.

[9] Jingxin Li, Aimin Hui, Xiang Zhang, Yumei Yang, and et al. Safety and immunogenicity of the sars-cov-2 bnt162b1 mrna vaccine in younger and older chinese adults: a randomized, placebo-controlled, double-blind phase 1 study. *Nature Medicine*, 27(6):1062–1070, Jun 2021.

[10] Otavio T. Ranzani, Matt Hitchings, Murilo Dorion Nieto, , and et al. Effectiveness of the coronavac vaccine in the elderly population during a p.1 variant-associated epidemic of covid-19 in brazil: A test-negative case-control study. *medRxiv*, 2021.

[11] Denis Sauré, Miguel O'Ryan, Juan Pablo Torres, Marcela Zuniga, Emilio Santelices, and Leonardo J. Basso. Dynamic igg seropositivity after rollout of coronavac and bnt162b2 covid-19 vaccines in chile: a sentinel surveillance study. *The Lancet Infectious Diseases*, 22(1):56–63, Sep 2021.

[12] K. Stiasny, J. H. Aberle, M. Keller, B. Grubeck-Loebenstein, and F. X. Heinz. Age affects quantity but not quality of antibody responses after vaccination with an inactivated flavivirus vaccine against tick-borne encephalitis. *PLoS One*, 7(3):e34145, 2012.

[13] Z. Wu, Y. Hu, M. Xu, Z Chen, and et al. Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine (CoronaVac) in healthy adults aged 60 years and older: a randomised, double-blind, placebo-controlled, phase 1/2 clinical trial. *Lancet Infect Dis*, 21(6):803–812, 06 2021.

Annexes

ANNEX A

Health Service Name	Study Start Date
Aconcagua	March 18th
Antofagasta	March 17th
Araucanía Norte	March 24th
Araucanía Sur	March 12th
Arauco	April 7th
Arica	March 18th
Atacama	June 1st
Aysén	April 6th
Biobío	March 23th
Chiloé	March 18th
Concepción	March 25th
Coquimbo	July 8th
del Libertador B. O'Higgins	April 13th
del Maule	March 16th
Iquique	March 24th
Magallanes	April 12th
Metropolitano Central	June 7th
Metropolitano Norte	June 7th
Metropolitano Occidente	June 7th
Metropolitano Oriente	April 16th
Metropolitano Sur	June 15th
Metropolitano Sur Oriente	April 12th
Ñuble	June 4th
Osorno	April 7th
Reloncaví	March 19th
Talcahuano	March 26th
Valdivia	April 6th
Valparaíso San Antonio	March 23th
Viña del Mar Quillota	March 15th

Figure 2.1: List of Health Services

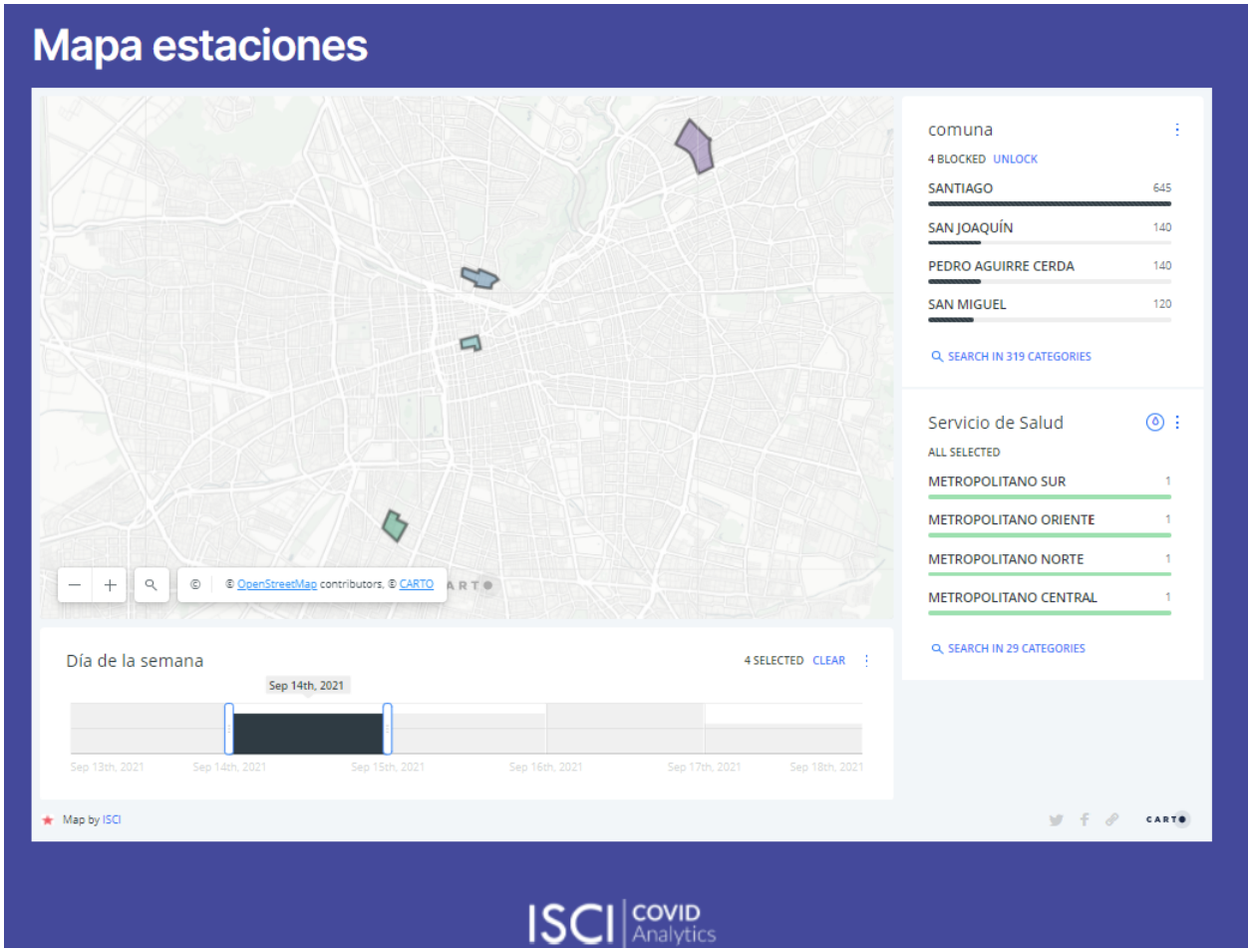

Figure 2.2: Example of how solutions of the model were displayed

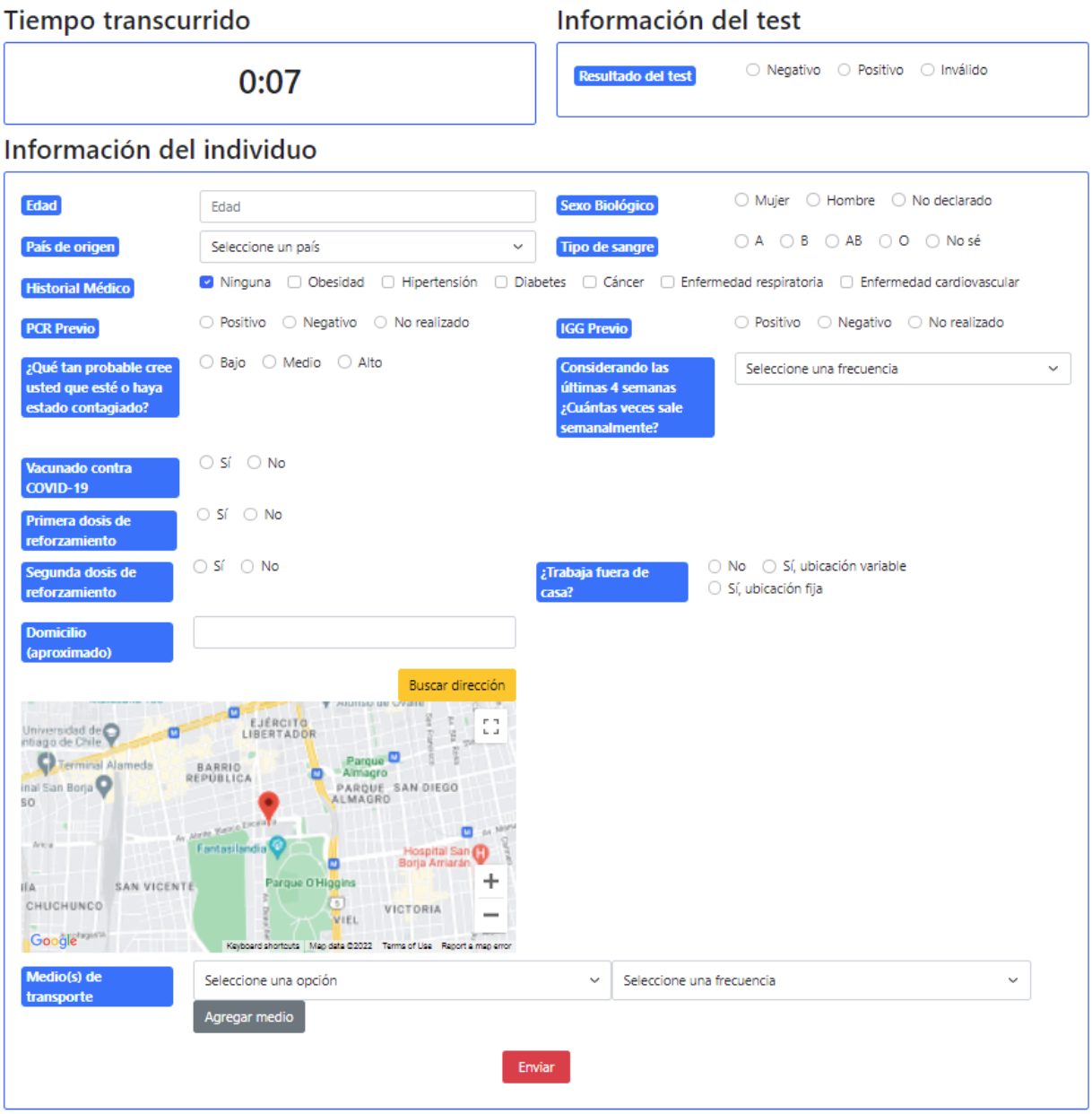

Figure 2.3: Form filled with information of individuals

ANNEX B

Number of usable samples recollected July 2th 2021

	Overall n = 53501	Unvaccinated n = 12897	Coronavac n = 31041	BNT16b2 n = 8751	Other n = 812
Age					
<=39	24096 (45%)	8526 (66.1%)	11503 (37.1%)	3741 (42.7%)	326 (40.1%)
[40,49]	11206 (20.9%)	2595 (20.1%)	6213 (20%)	2197 (25.1%)	201 (24.8%)
[50,59]	9647 (18%)	1362 (10.6%)	5798 (18.7%)	2379 (27.2%)	108 (13.3%)
[60,69]	5834 (10.9%)	315 (2.4%)	5072 (16.3%)	344 (3.9%)	103 (12.7%)
>=70	2718 (5.1%)	99 (0.8%)	2455 (7.9%)	90 (1%)	74 (9.1%)
Gender					
Male	22183 (41.5%)	5745 (44.5%)	12472 (40.2%)	3459 (39.5%)	507 (62.4%)
Female	31318 (58.5%)	7152 (55.5%)	18569 (59.8%)	5292 (60.5%)	305 (37.6%)
Nacionalidad					
Chile	51230 (95.8%)	11989 (93%)	30079 (96.9%)	8401 (96%)	761 (93.7%)
Other	2271 (4.2%)	908 (7%)	962 (3.1%)	350 (4%)	51 (6.3%)
Prevoius positive					
PCR	33115 (61.9%)	830 (6.4%)	1864 (6%)	566 (6.5%)	55 (6.8%)
IgG	802 (1.5%)	75 (0.6%)	588 (1.9%)	124 (1.4%)	15 (1.8%)
Times Leaving Home per week					
<3	17366 (32.5%)	4449 (34.5%)	9827 (31.7%)	2817 (32.2%)	273 (33.6%)
[3,5]	18528 (34.6%)	4318 (33.5%)	10974 (35.4%)	2976 (34%)	260 (32%)
[6,7]	13695 (25.6%)	3181 (24.7%)	7998 (25.8%)	2321 (26.5%)	195 (24%)
>7	3912 (7.3%)	949 (7.4%)	2242 (7.2%)	637 (7.3%)	84 (10.3%)
Comorbidities					
None	38513 (72%)	10910 (84.6%)	20666 (66.6%)	6340 (72.4%)	597 (73.5%)
Obesity	2581 (4.8%)	558 (4.3%)	1527 (4.9%)	436 (5%)	60 (7.4%)
High blood pressure	8339 (15.6%)	710 (5.5%)	6260 (20.2%)	1257 (14.4%)	112 (13.8%)
Diabetes	4117 (7.7%)	387 (3%)	2983 (9.6%)	688 (7.9%)	59 (7.3%)
Cancer	591 (1.1%)	58 (0.4%)	428 (1.4%)	93 (1.1%)	12 (1.5%)
Chronic pulmonary disease	2468 (4.6%)	472 (3.7%)	1578 (5.1%)	382 (4.4%)	36 (4.4%)
Chronic cardiovascular disease	1333 (2.5%)	166 (1.3%)	959 (3.1%)	178 (2%)	30 (3.7%)

Figure 3.7: Baseline characteristics of samples recollected

Number of usable samples recollected in Gran Santiago December 13th 2021

	Overall n = 18491	Unvaccinated n = 1557	Coronavac n = 13494	BNT16b2 n = 2884	Other n = 556
Age					
<=39	7225 (39.1%)	1071 (68.8%)	4422 (32.8%)	1381 (47.9%)	351 (63.1%)
[40,49]	3375 (18.3%)	250 (16.1%)	2342 (17.4%)	659 (22.9%)	124 (22.3%)
[50,59]	3508 (19%)	159 (10.2%)	2618 (19.4%)	688 (23.9%)	43 (7.7%)
[60,69]	2917 (15.8%)	60 (3.9%)	2698 (20%)	136 (4.7%)	23 (4.1%)
>=70	1466 (7.9%)	17 (1.1%)	1414 (10.5%)	20 (0.7%)	15 (2.7%)
Gender					
Male	7827 (42.3%)	746 (47.9%)	5527 (41%)	1161 (40.3%)	393 (70.7%)
Female	10664 (57.7%)	811 (52.1%)	7967 (59%)	1723 (59.7%)	163 (29.3%)
Nacionalidad					
Chile	16861 (91.2%)	1296 (83.2%)	12510 (92.7%)	2577 (89.4%)	478 (86%)
Other	1630 (8.8%)	261 (16.8%)	984 (7.3%)	307 (10.6%)	78 (14%)
Prevoius positive					
PCR	1587 (8.6%)	133 (8.5%)	1110 (8.2%)	289 (10%)	55 (9.9%)
IgG	406 (2.2%)	14 (0.9%)	268 (2%)	111 (3.8%)	13 (2.3%)
Times Leaving Home per week					
<3	5070 (27.4%)	507 (32.6%)	3773 (28%)	695 (24.1%)	95 (17.1%)
[3,5]	6670 (36.1%)	532 (34.2%)	4846 (35.9%)	1089 (37.8%)	203 (36.5%)
[6,7]	4865 (26.3%)	369 (23.7%)	3490 (25.9%)	824 (28.6%)	182 (32.7%)
>7	1886 (10.2%)	149 (9.6%)	1385 (10.3%)	276 (9.6%)	76 (13.7%)
Comorbidities					
None	12420 (67.2%)	1281 (82.3%)	8483 (62.9%)	2187 (75.8%)	469 (84.4%)
Obesity	1327 (7.2%)	90 (5.8%)	958 (7.1%)	241 (8.4%)	38 (6.8%)
High blood pressure	3549 (19.2%)	99 (6.4%)	3096 (22.9%)	321 (11.1%)	33 (5.9%)
Diabetes	1611 (8.7%)	53 (3.4%)	1400 (10.4%)	142 (4.9%)	16 (2.9%)
Cancer	305 (1.6%)	15 (1%)	245 (1.8%)	43 (1.5%)	2 (0.4%)
Chronic pulmonary disease	766 (4.1%)	50 (3.2%)	625 (4.6%)	79 (2.7%)	12 (2.2%)
Chronic cardiovascular disease	476 (2.6%)	24 (1.5%)	405 (3%)	42 (1.5%)	5 (0.9%)

Figure 3.8: Baseline characteristics of samples recollected for Gran Santiago

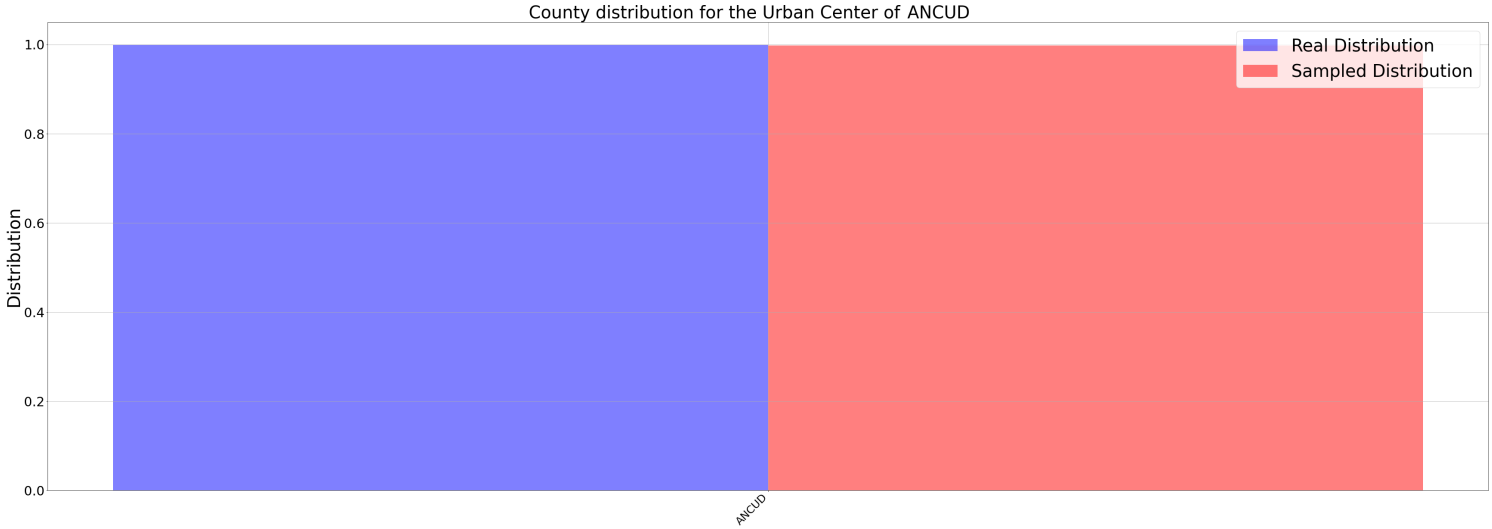

Figure 3.10: Ancud Urban Center Distribution

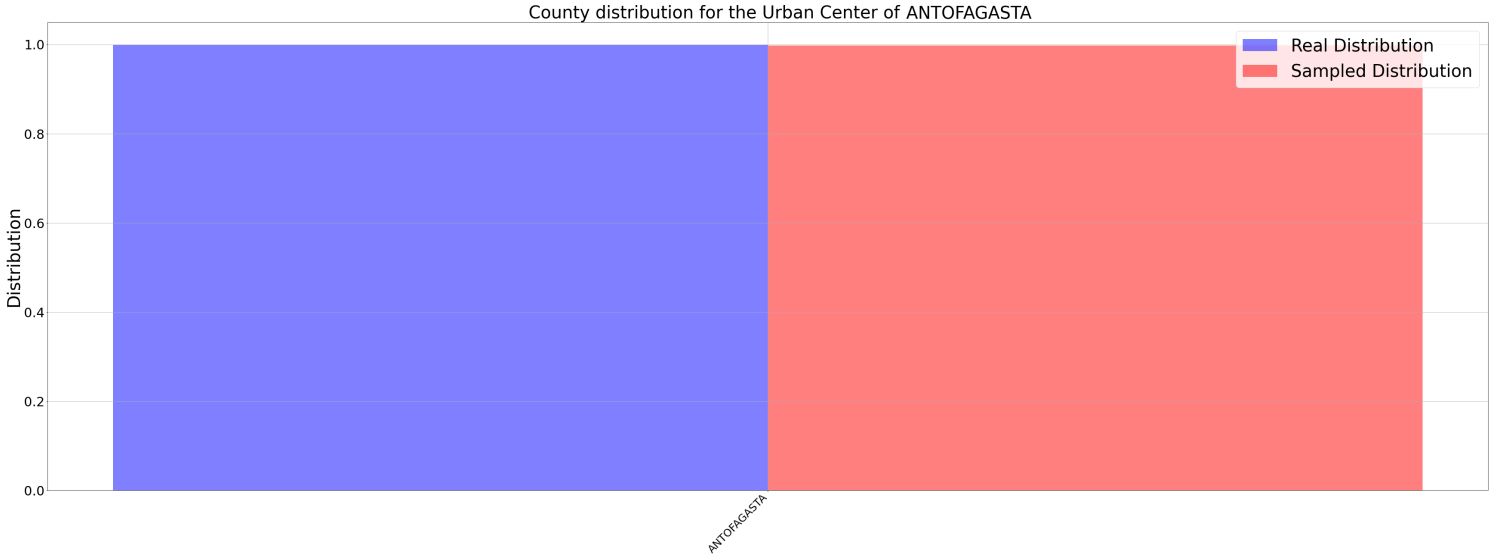

Figure 3.11: Antofagasta Urban Center Distribution

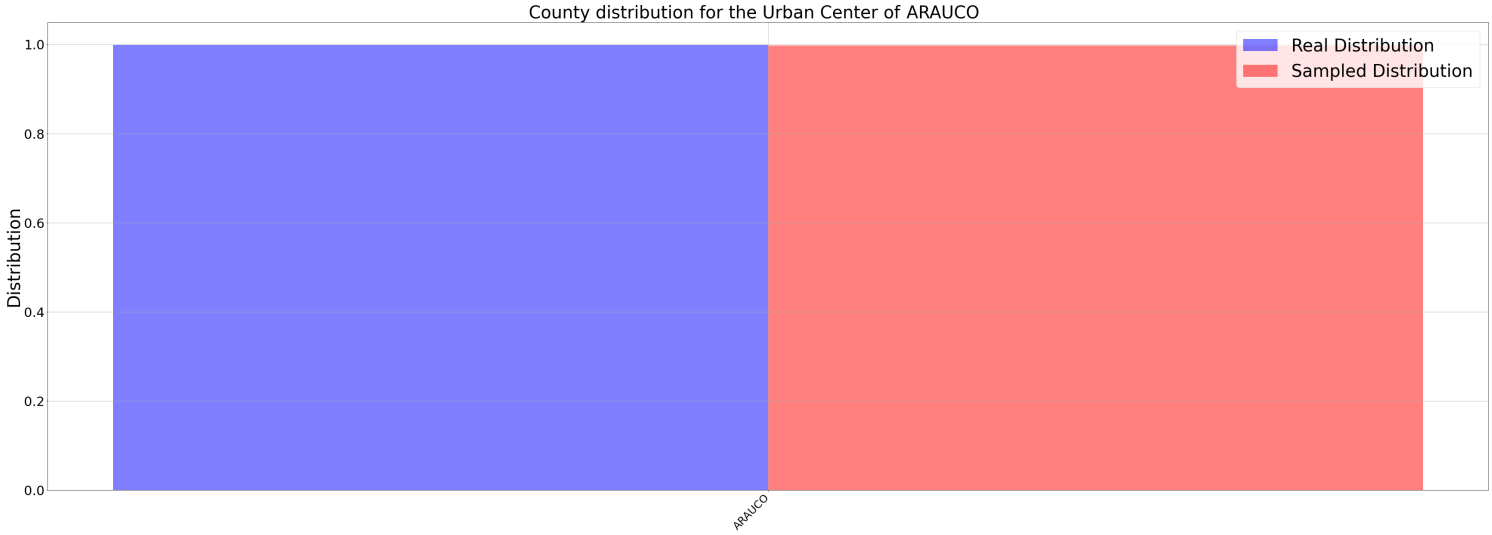

Figure 3.12: Arauco Urban Center Distribution

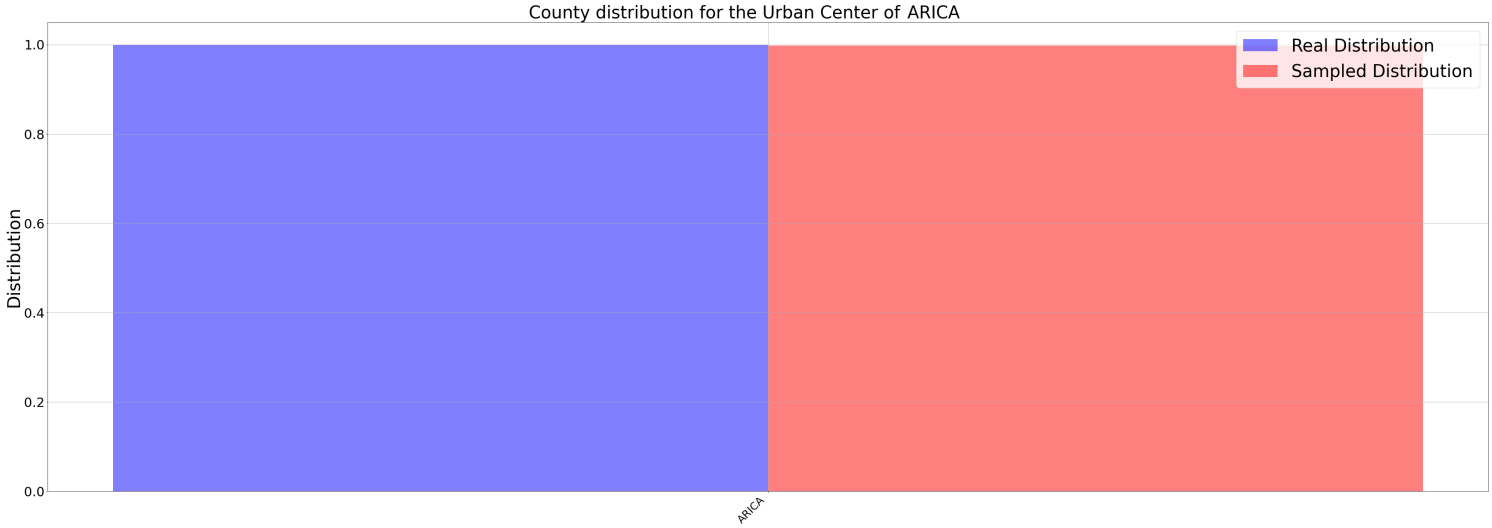

Figure 3.13: Arica Urban Center Distribution

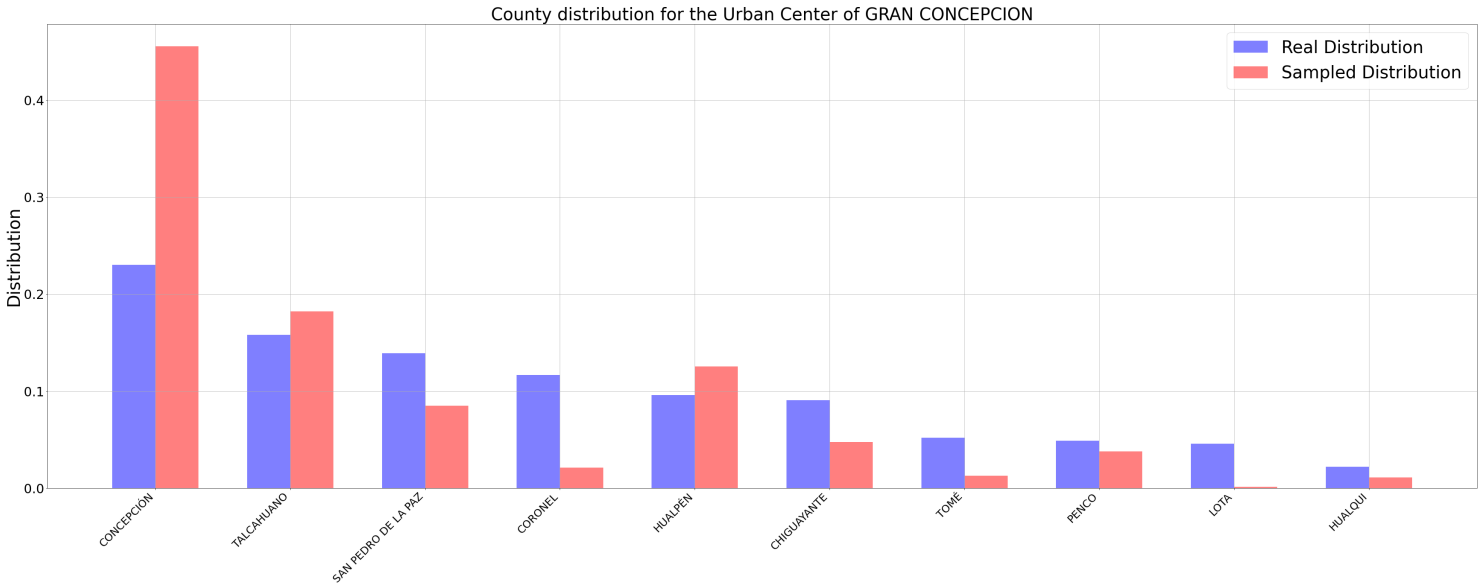

Figure 3.14: Gran Concepción Urban Center Distribution

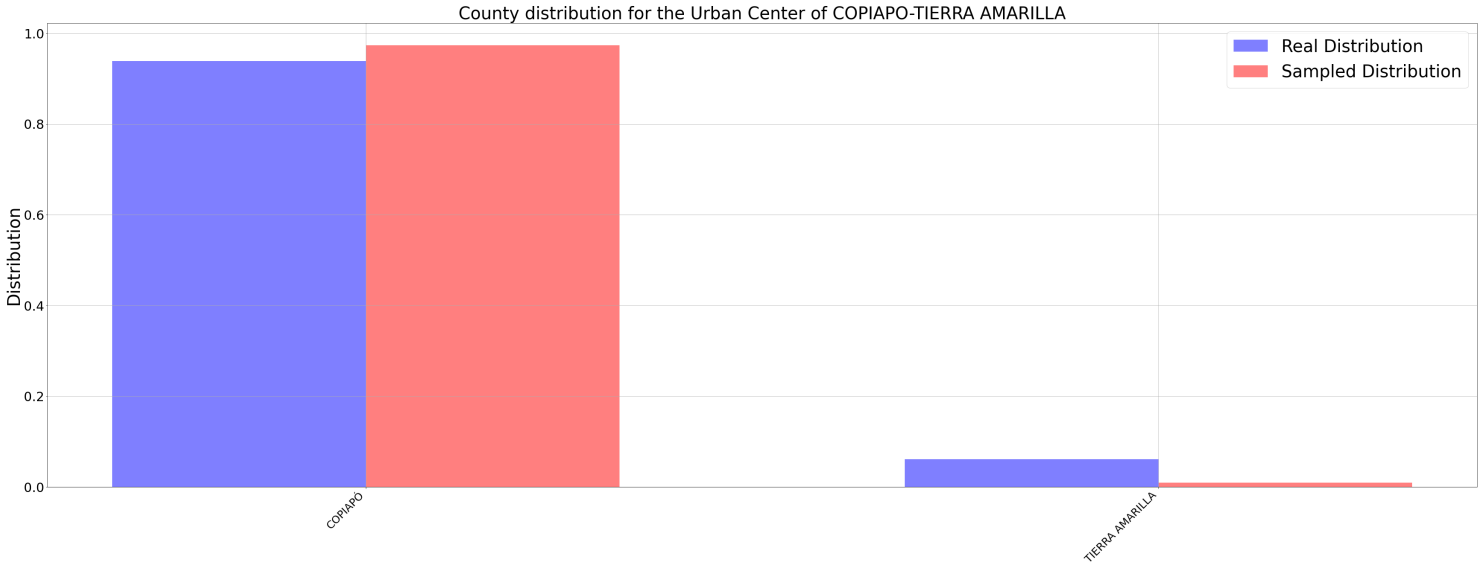

Figure 3.15: Copiapó- Tierra Amarilla Urban Center Distribution

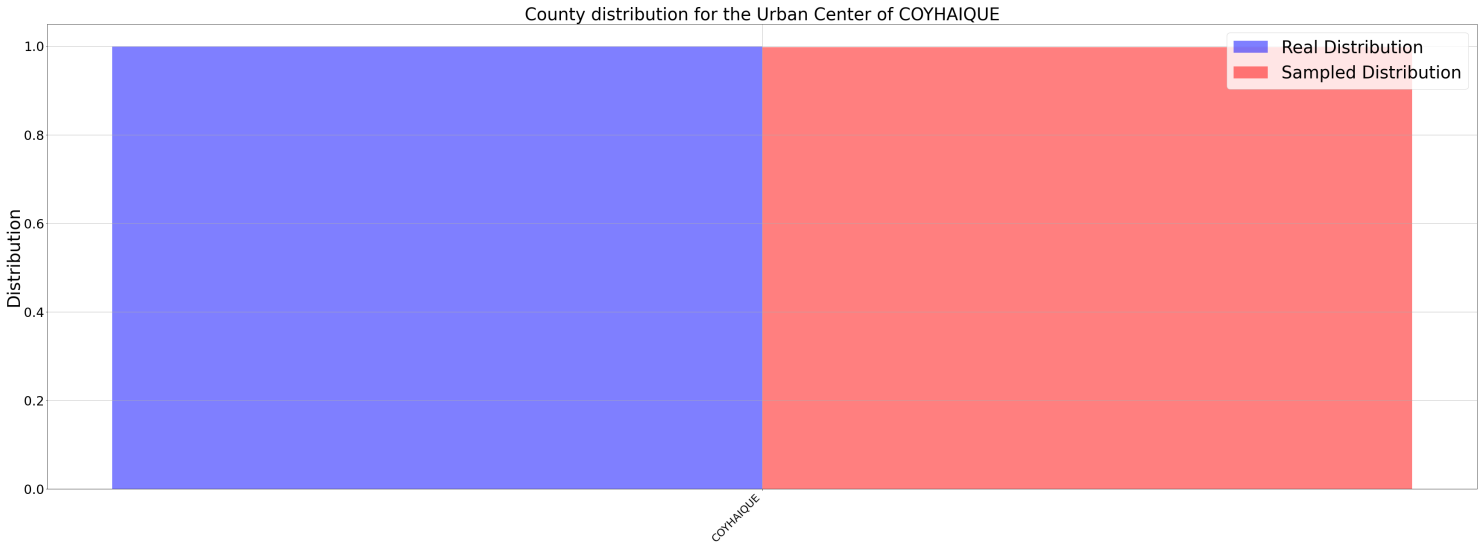

Figure 3.16: Coyhaique Urban Center Distribution

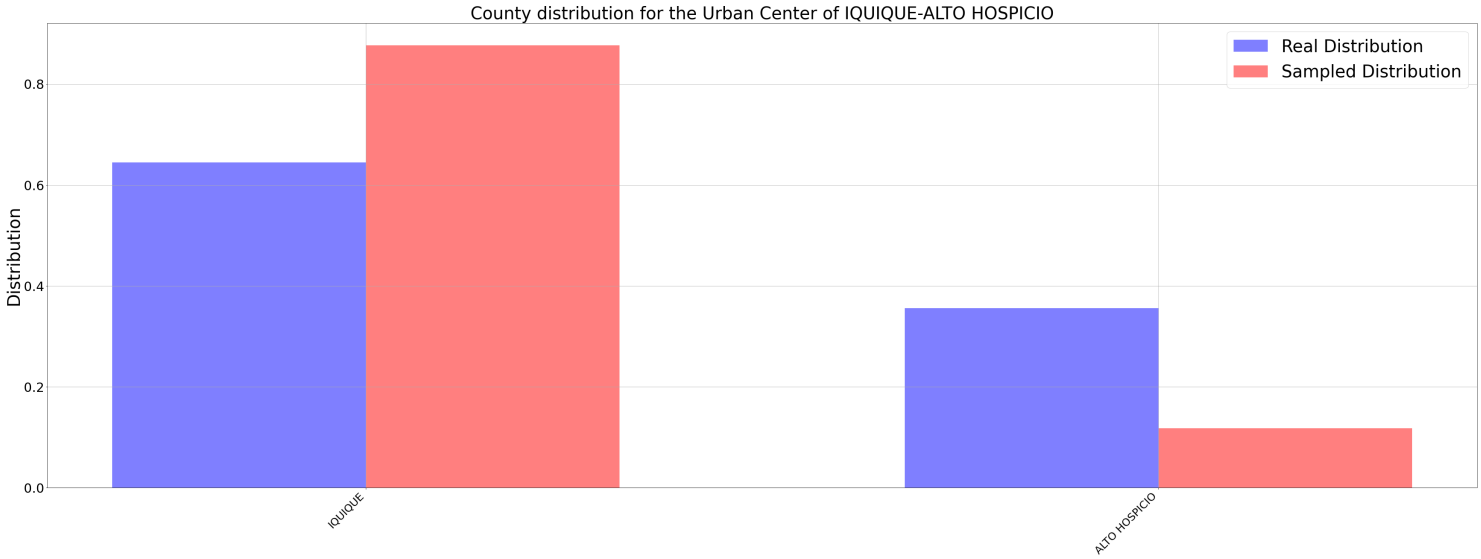

Figure 3.17: Iquique- Alto Hospicio Urban Center Distribution

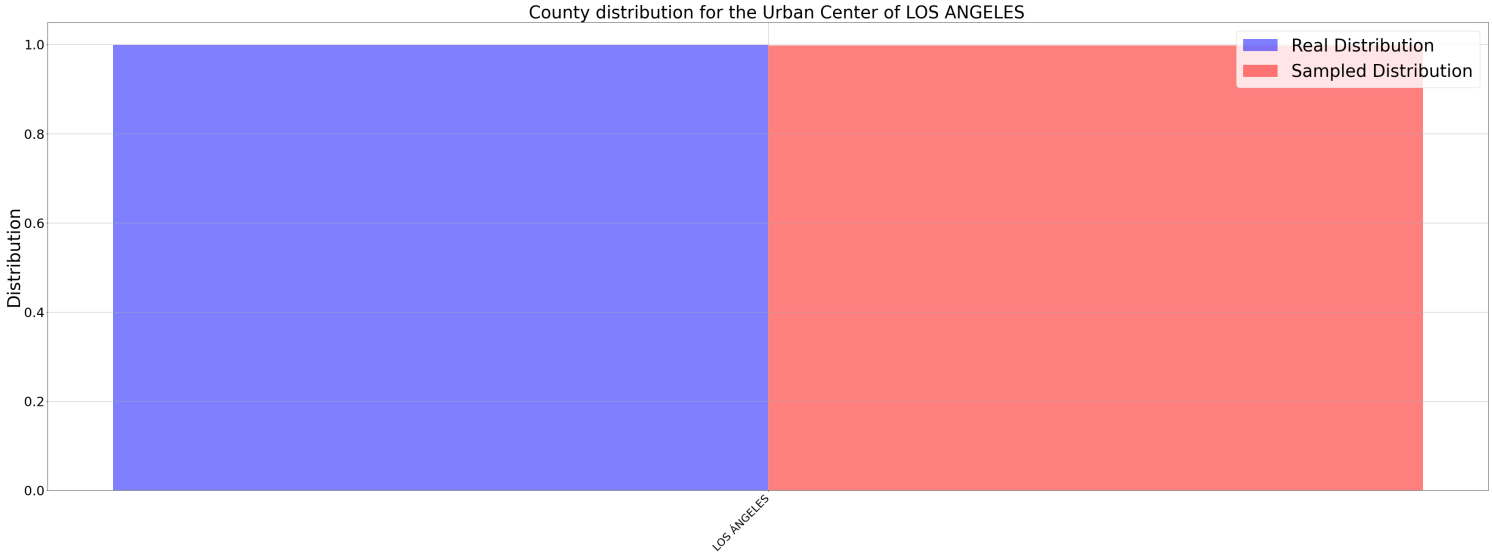

Figure 3.18: Los Ángeles Urban Center Distribution

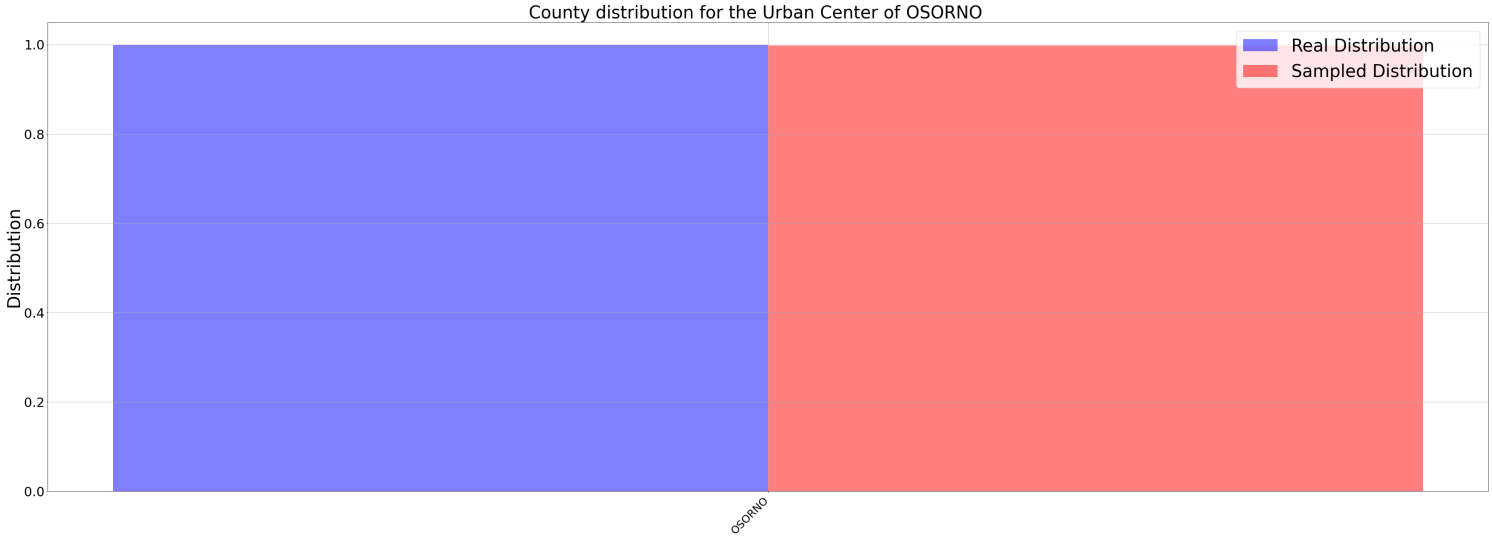

Figure 3.19: Osorno Urban Center Distribution

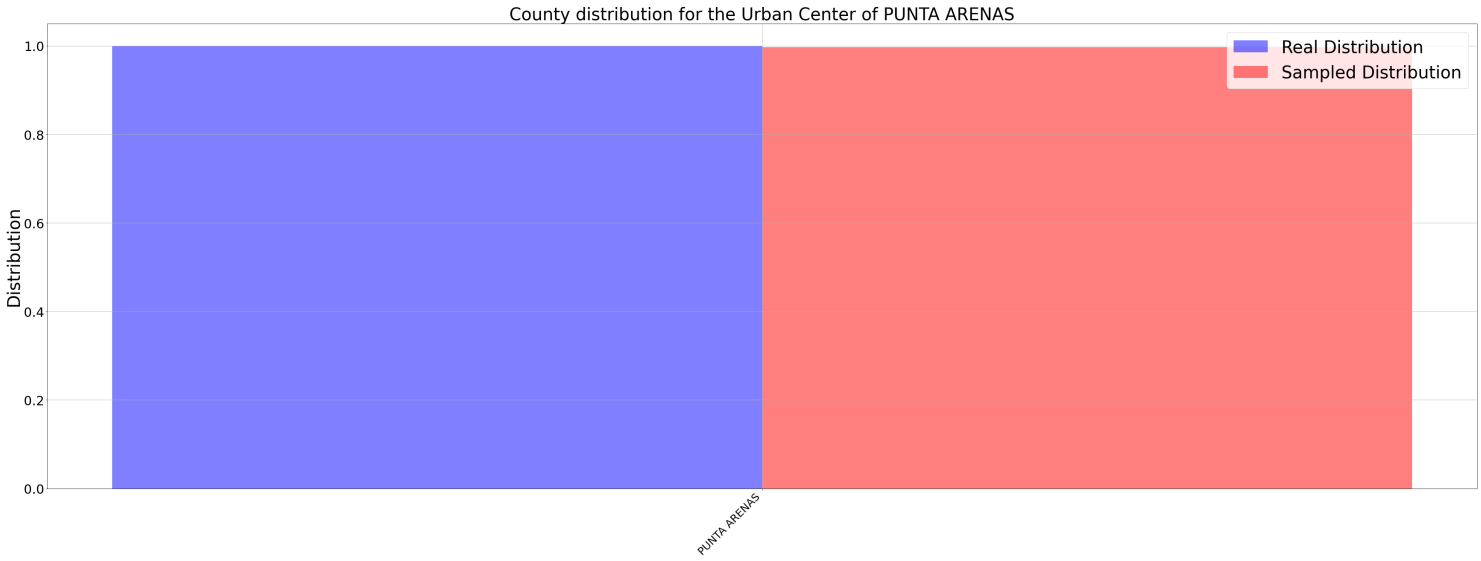

Figure 3.20: Punta Arenas Urban Center Distribution

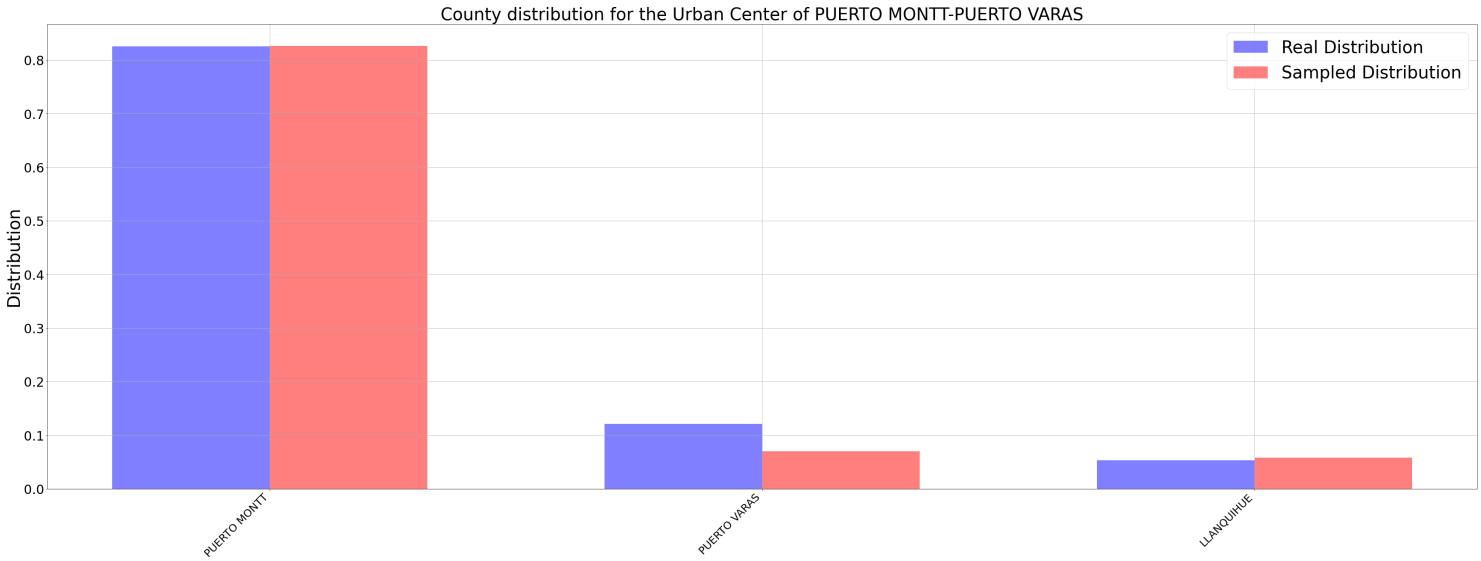

Figure 3.21: Puerto Montt- Puerto Varas Urban Center Distribution

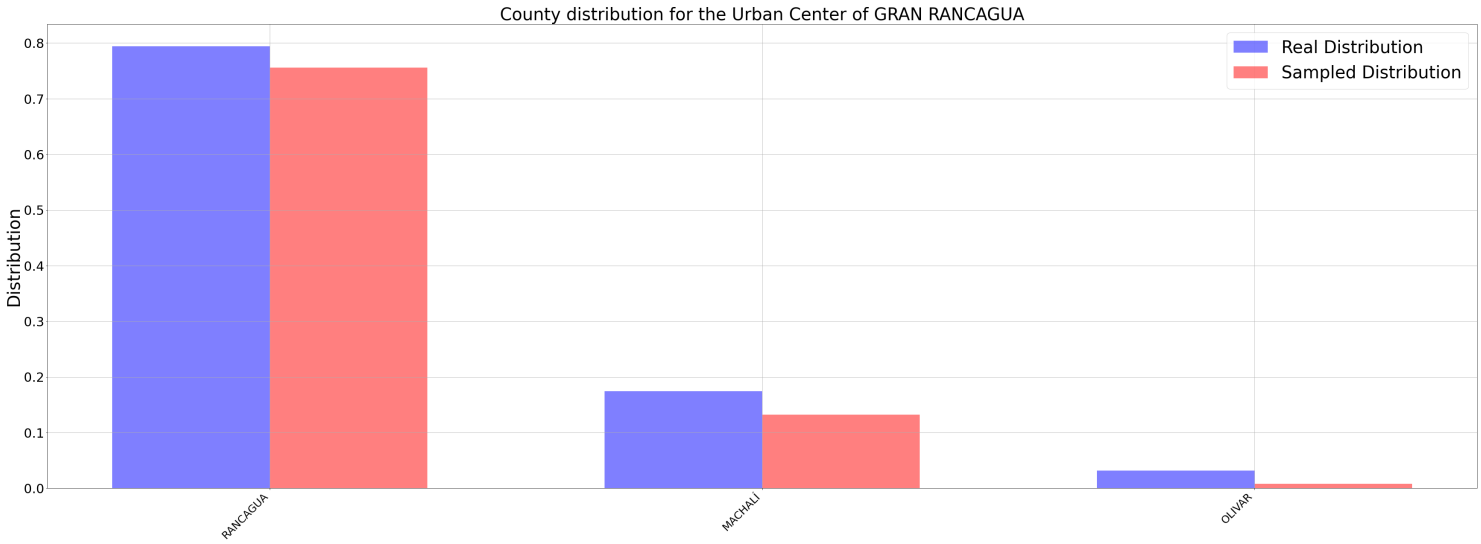

Figure 3.22: Gran Rancagua Urban Center Distribution

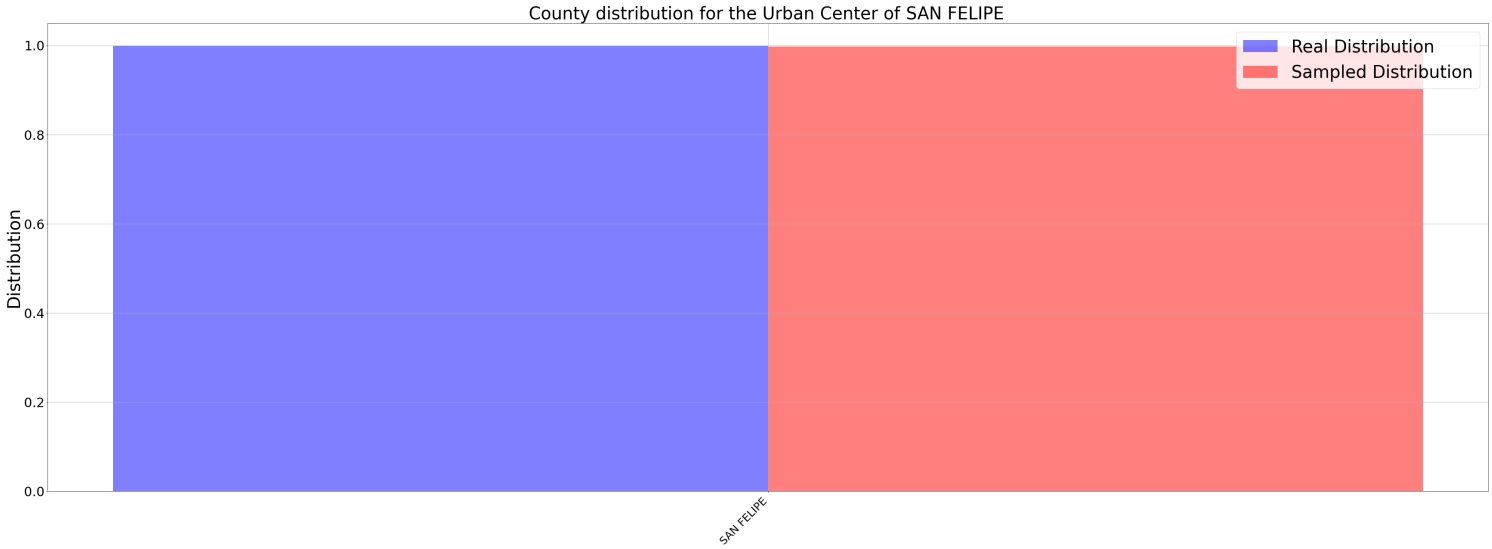

Figure 3.23: San Felipe Urban Center Distribution

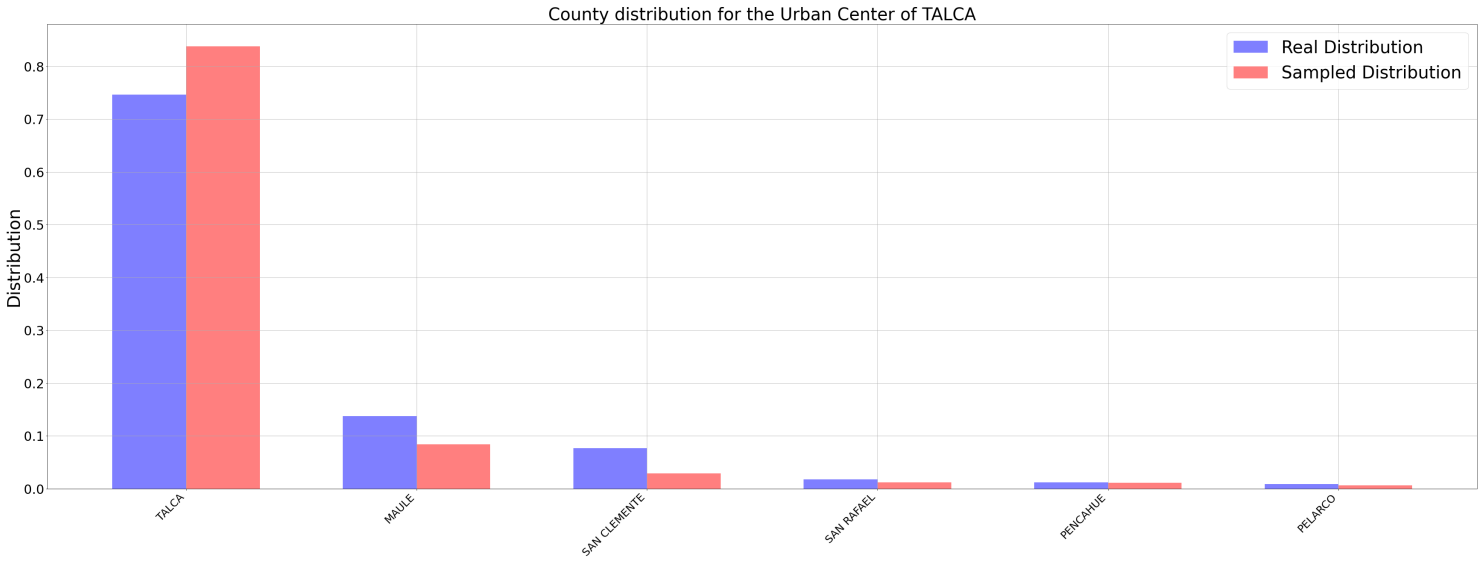

Figure 3.24: Talca Urban Center Distribution

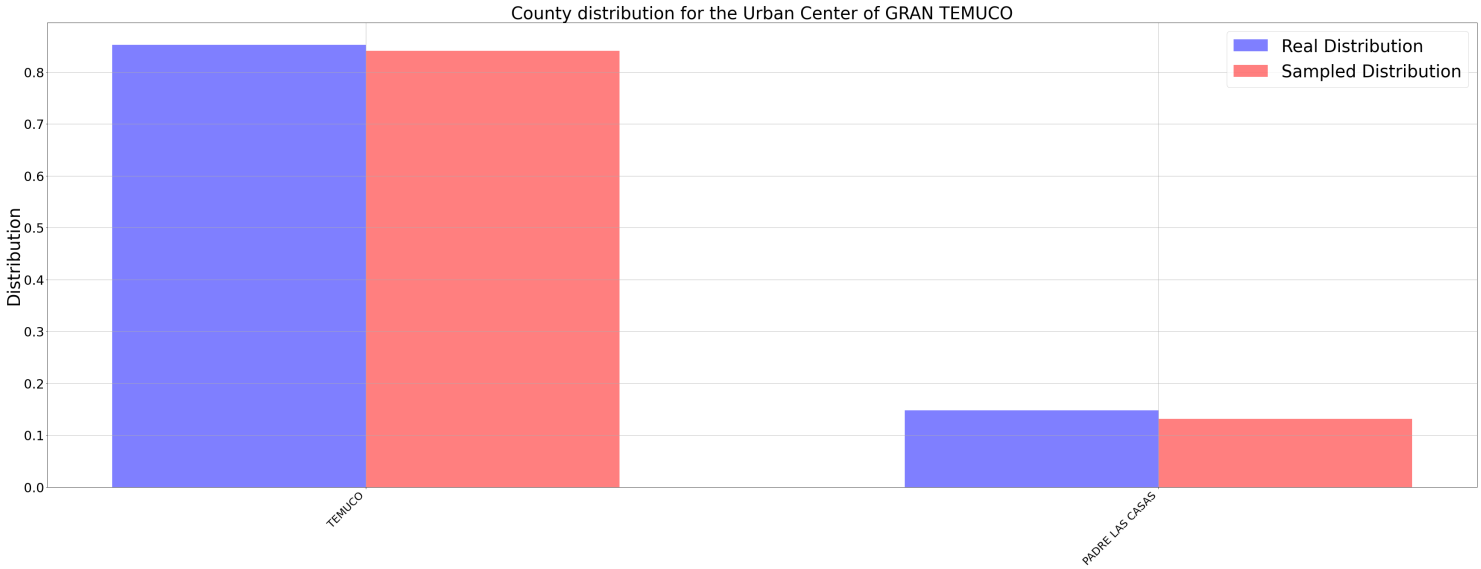

Figure 3.25: Gran Temuco Urban Center Distribution

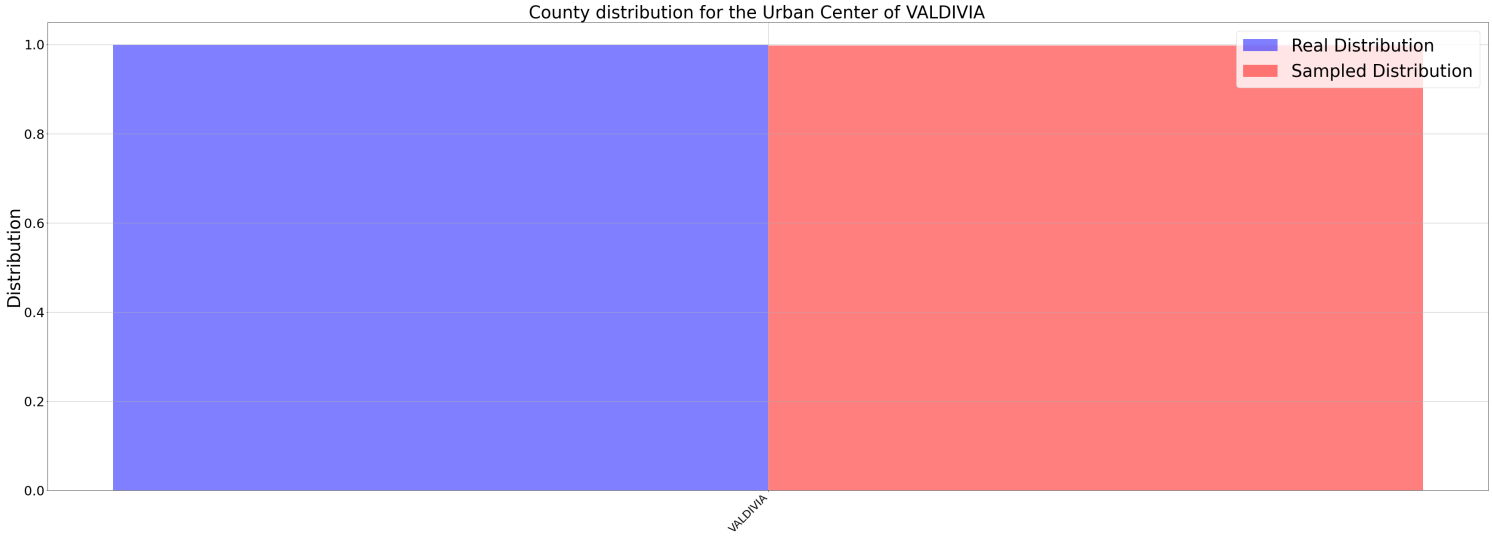

Figure 3.26: Valdivia Urban Center Distribution

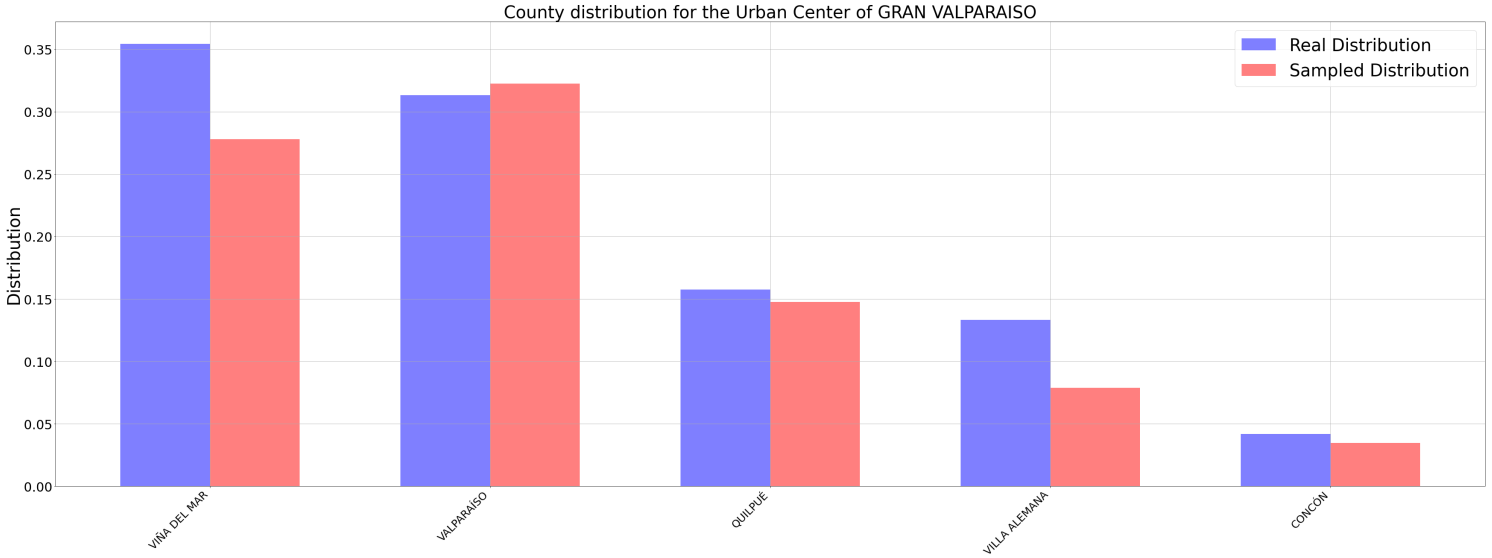

Figure 3.27: Gran Valparaíso Urban Center Distribution

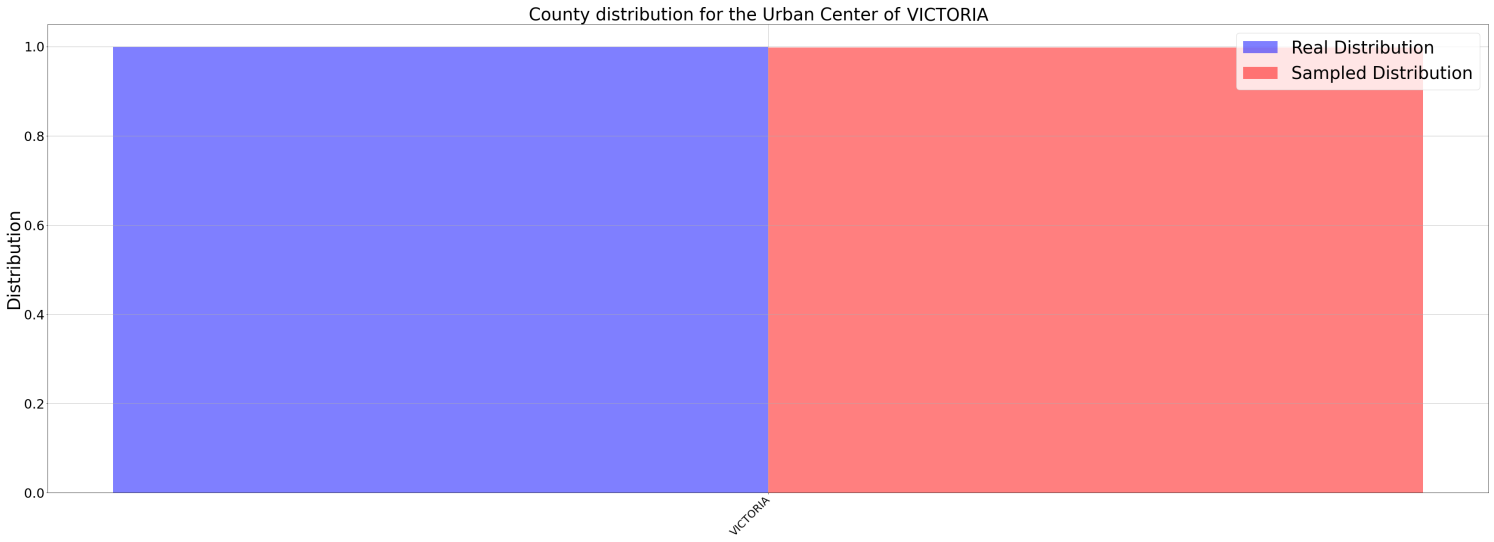

Figure 3.28: Victoria Urban Center Distribution

ANNEX C

Variable	Mean	SD	2.50%	97.50%	Variable	Mean	SD	2.50%	97.50%
Fortnight of study * Valparaíso	0.3224	0.0437	0.2362	0.4091	BNT162b2 * Fortnight 4 After Vac.	2.817	0.3329	2.204	3.544
Fortnight of study * Concepción	0.1406	0.06379	0.01434	0.2693	BNT162b2 * Fortnight 5After Vac.	3.339	0.5238	2.379	4.422
Fortnight of study * Valdivia	0.2361	0.0497	0.1406	0.3343	BNT162b2 * Fortnight 6 After Vac.	2.684	0.4795	1.796	3.7
Fortnight of study * Talca	0.006125	0.04677	-0.08596	0.09539	BNT162b2 * Fortnight 7 After Vac.	3.987	1.348	1.945	7.25
Fortnight of study * Puerto Montt & Varas	0.2229	0.04722	0.127	0.3132	BNT162b2 * Fortnight 8 After Vac.	3.503	1.28	1.565	6.506
alpha1	-1.978	0.1685	-2.301	-1.65	BNT162b2 * Fortnight 9 After Vac.	1.969	1.313	-0.1649	5.028
alpha2	0.4782	0.1361	0.2155	0.7453	BNT162b2 * Fortnight 10 After Vac.	2.651	1.384	0.4458	5.92
alpha3	-0.2927	0.06968	-0.4279	-0.161	BNT162b2 * Obesity	-0.08903	0.2696	-0.5941	0.4557
Gender	0.09751	0.07519	-0.04951	0.2432	BNT162b2 * Hypertension	0.2489	0.2049	-0.1532	0.6524
Fortnight of Study	0.02298	0.02946	-0.03472	0.08024	BNT162b2 * Diabetes	-0.383	0.2394	-0.8495	0.08772
Valparaíso	-1.391	0.2251	-1.837	-0.9597	BNT162b2 * Chronic Pulmonary D.	0.6012	0.3745	-0.09293	1.384
Concepción	-1.647	0.2846	-2.236	-1.095	BNT162b2 * Chronic Cardiovascular D.	-0.3872	0.4707	-1.27	0.5721
Valdivia	-0.2267	0.2668	-0.761	0.2857	CoronaVac * Gender	0.2505	0.04711	0.1622	0.3446
Talca	-0.2211	0.1923	-0.6104	0.1605	CoronaVac * Age Range (40,49)	-0.04588	0.06226	-0.1682	0.07589
Puerto Montt & Varas	-0.9964	0.2109	-1.397	-0.5901	CoronaVac * Age Range (50,59)	-0.1929	0.07063	-0.3338	-0.05096
Age Range (40,49)	0.1048	0.09828	-0.08774	0.2952	CoronaVac * Age Range (60,69)	-0.3309	0.09613	-0.5209	-0.148
Age Range (50,59)	0.3954	0.12	0.1581	0.6252	CoronaVac * Age Range (70+)	-0.6874	0.1461	-0.9775	-0.414
Age Range (60,69)	0.5754	0.1681	0.2269	0.8886	CoronaVac * Fortnight 1 After Vac.	0.5869	0.07417	0.4463	0.736
Age Range (70+)	0.6043	0.2355	0.1041	1.022	CoronaVac * Fortnight 2 After Vac.	0.9909	0.07358	0.8542	1.14
Times Leaving Home per week (3,5)	0.01683	0.07329	-0.1286	0.1546	CoronaVac * Fortnight 3 After Vac.	1.045	0.07491	0.9058	1.197
Times Leaving Home per week (6,7)	-0.08423	0.08654	-0.2547	0.07676	CoronaVac * Fortnight 4 After Vac.	0.8524	0.08028	0.6968	1.011
Times Leaving Home per week (7+)	-0.7089	0.1589	-1.024	-0.4096	CoronaVac * Fortnight 5After Vac.	0.5581	0.09282	0.3758	0.7408
Medium Perception of Risk	0.2655	0.07643	0.1229	0.417	CoronaVac * Fortnight 6 After Vac.	0.1793	0.09474	-0.005577	0.3545
High Perception of Risk	1.166	0.08675	0.9975	1.337	CoronaVac * Fortnight 7 After Vac.	-0.00102	0.1037	-0.1993	0.1917
BNT162b2 * Gender	-0.04025	0.1319	-0.302	0.2191	CoronaVac * Fortnight 8 After Vac.	-0.2763	0.1732	-0.6387	0.0454
BNT162b2 * Age Range (40,49)	-0.1079	0.1563	-0.4169	0.1914	CoronaVac * Fortnight 9 After Vac.	-0.7285	0.4657	-1.825	0.0296
BNT162b2 * Age Range (50,59)	-0.193	0.1709	-0.5229	0.1492	CoronaVac * Fortnight 10 After Vac.	-381700	461200	-1434000	-10940
BNT162b2 * Age Range (60,69)	-1.47	0.3048	-2.072	-0.8763	CoronaVac * Obesity	-0.1933	0.09157	-0.3739	-0.01369
BNT162b2 * Age Range (70+)	-2.934	0.5743	-4.178	-1.859	CoronaVac * Hypertension	-0.01137	0.05474	-0.1164	0.09481
BNT162b2 * Fortnight 1 After Vac.	1.694	0.1467	1.41	1.969	CoronaVac * Diabetes	-0.2733	0.06921	-0.4066	-0.1412
BNT162b2 * Fortnight 2 After Vac.	2.481	0.2214	2.061	2.922	CoronaVac * Chronic Pulmonary D.	0.0384	0.08717	-0.1366	0.2011
BNT162b2 * Fortnight 3 After Vac.	2.781	0.3	2.22	3.399	CoronaVac * Chronic Cardiovascular D.	-0.1795	0.1346	-0.4581	0.08282

Figure 4.2: Variables July 2th most populated urban centers

Variable	Mean	SD	2.50%	97.50%	Variable	Mean	SD	2.50%	97.50%
Working Outside (Yes)	-0.114	0.101	-0.315	0.087	BNT162b2 * Fortnight 13 After Vac.	0.585	0.537	-0.434	1.620
Working Outside (No)	-0.057	0.128	-0.297	0.174	BNT162b2 * Fortnight 14 After Vac.	1.090	0.628	-0.126	2.430
Transport by Vehicle	0.121	0.233	-0.345	0.563	BNT162b2 * Fortnight 15 After Vac.	1.470	0.761	0.023	3.050
Transport by Bus	0.406	0.302	-0.167	1.020	BNT162b2 * Fortnight 16 After Vac.	79.900	61.240	5.280	233.000
Transport by Metro	0.694	0.250	0.198	1.180	BNT162b2 * Fortnight 17 After Vac.	1.010	1.069	-0.830	3.390
Vehicle Frequency	-0.020	0.043	-0.101	0.067	BNT162b2 * Fortnight 18 After Vac.	1.690	1.502	-0.690	5.020
Bus Frequency	-0.032	0.061	-0.151	0.090	BNT162b2 * Obesity	0.338	0.401	-0.422	1.210
Metro Frequency	-0.054	0.049	-0.146	0.048	BNT162b2 * Hypertension	-0.277	0.334	-0.929	0.340
alpha 1	-2.420	0.227	-2.870	-1.960	BNT162b2 * Diabetes	0.601	0.618	-0.525	1.890
alpha 2	1.440	0.234	0.994	1.910	BNT162b2 * Chronic Pulmonary D.	0.651	0.680	-0.546	2.030
alpha 3	-0.451	0.138	-0.720	-0.187	BNT162b2 * Chronic Cardiovascular D.	-0.445	0.804	-1.900	1.240
Gender	0.206	0.119	-0.040	0.447	CoronaVac * Gender	0.257	0.082	0.092	0.420
Study Fortnight	0.112	0.011	0.090	0.132	CoronaVac * Age Range (40,49)	0.156	0.112	-0.071	0.378
Age Range (40,49)	-0.063	0.158	-0.404	0.211	CoronaVac * Age Range (50,59)	0.199	0.121	-0.032	0.430
Age Range (50,59)	0.049	0.165	-0.298	0.355	CoronaVac * Age Range (60,69)	-0.006	0.145	-0.312	0.270
Age Range (60,69)	0.097	0.181	-0.268	0.441	CoronaVac * Age Range (70+)	0.587	0.177	0.213	0.902
Age Range (70+)	-0.976	0.865	-2.720	0.088	CoronaVac * Fortnight 1 After Vac.	0.775	0.148	0.494	1.070
Times Leaving Home per week (3,5)	0.281	0.129	0.024	0.526	CoronaVac * Fortnight 2 After Vac.	0.911	0.151	0.627	1.210
Times Leaving Home per week (6,7)	0.067	0.179	-0.289	0.422	CoronaVac * Fortnight 3 After Vac.	0.793	0.153	0.498	1.110
Times Leaving Home per week (7+)	-0.607	0.250	-1.100	-0.142	CoronaVac * Fortnight 4 After Vac.	0.699	0.150	0.399	0.990
Medium Perception of Risk	-0.008	0.098	-0.214	0.167	CoronaVac * Fortnight 5After Vac.	0.435	0.149	0.149	0.739
High Perception of Risk	1.040	0.113	0.814	1.270	CoronaVac * Fortnight 6 After Vac.	0.045	0.152	-0.264	0.331
BNT162b2 * Gender	-0.290	0.206	-0.698	0.120	CoronaVac * Fortnight 7 After Vac.	-0.228	0.152	-0.520	0.050
BNT162b2 * Age Range (40,49)	-0.052	0.248	-0.527	0.444	CoronaVac * Fortnight 8 After Vac.	-0.491	0.171	-0.828	-0.165
BNT162b2 * Age Range (50,59)	-0.100	0.288	-0.623	0.492	CoronaVac * Fortnight 9 After Vac.	-0.615	0.186	-1.000	-0.243
BNT162b2 * Age Range (60,69)	-1.670	0.404	-2.510	-0.909	CoronaVac * Fortnight 10 After Vac.	-0.621	0.213	-1.070	-0.245
BNT162b2 * Age Range (70+)	-1.360	0.903	-3.170	0.402	CoronaVac * Fortnight 11 After Vac.	-1.140	0.281	-1.770	-0.664
BNT162b2 * Fortnight 1 After Vac.	1.140	0.330	0.492	1.800	CoronaVac * Fortnight 12 After Vac.	0.038	0.196	-0.362	0.396
BNT162b2 * Fortnight 2 After Vac.	2.010	0.463	1.100	2.950	CoronaVac * Fortnight 13 After Vac.	0.661	0.184	0.297	1.000
BNT162b2 * Fortnight 3 After Vac.	1.740	0.429	0.934	2.630	CoronaVac * Fortnight 14 After Vac.	1.390	0.197	1.010	1.780
BNT162b2 * Fortnight 4 After Vac.	2.550	0.578	1.500	3.810	CoronaVac * Fortnight 15 After Vac.	1.950	0.210	1.560	2.380
BNT162b2 * Fortnight 5After Vac.	4.320	1.311	2.370	7.460	CoronaVac * Fortnight 16 After Vac.	2.270	0.243	1.780	2.740
BNT162b2 * Fortnight 6 After Vac.	2.690	0.675	1.490	4.140	CoronaVac * Fortnight 17 After Vac.	2.540	0.350	1.880	3.230
BNT162b2 * Fortnight 7 After Vac.	2.160	0.606	1.100	3.450	CoronaVac * Fortnight 18 After Vac.	1.990	0.331	1.370	2.650
BNT162b2 * Fortnight 8 After Vac.	2.220	0.640	1.060	3.620	CoronaVac * Obesity	0.190	0.107	-0.014	0.392
BNT162b2 * Fortnight 9 After Vac.	1.520	0.591	0.442	2.720	CoronaVac * Hypertension	-0.260	0.074	-0.397	-0.117
BNT162b2 * Fortnight 10 After Vac.	0.977	0.402	0.206	1.770	CoronaVac * Diabetes	-0.216	0.094	-0.398	-0.031
BNT162b2 * Fortnight 11 After Vac.	2.220	0.681	0.998	3.720	CoronaVac * Chronic Pulmonary D.	0.098	0.138	-0.169	0.371
BNT162b2 * Fortnight 12 After Vac.	0.846	0.499	-0.165	1.890	CoronaVac * Chronic Cardiovascular D.	0.102	0.171	-0.274	0.426

Figure 4.5: Logit Results Gran Santiago

Variable	Mean	SD	2.50%	97.50%	Variable	Mean	SD	2.50%	97.50%
Working Outside (Yes)	0.041	0.064	-0.071	0.154	CoronaVac * Fortnight 12 After Vac.	1.133	0.381	0.475	1.819
Working Outside (No)	-0.030	0.083	-0.183	0.111	CoronaVac * Fortnight 13 After Vac.	0.354	0.391	-0.159	1.212
Transport by Vehicle	0.238	0.165	-0.070	0.539	CoronaVac * Fortnight 14 After Vac.	-0.920	0.595	-2.066	0.140
Transport by Bus	0.067	0.203	-0.252	0.443	CoronaVac * Fortnight 15 After Vac.	-17.070	8.790	-27.600	-2.032
Transport by Metro	0.457	0.177	0.148	0.811	CoronaVac * Fortnight 16 After Vac.	-11.990	5.149	-18.040	-2.371
Vehicle Frequency	-0.054	0.029	-0.102	0.001	CoronaVac * Fortnight 17 After Vac.	-11.700	5.768	-24.090	-4.640
Bus Frequency	-0.035	0.041	-0.109	0.031	CoronaVac * Fortnight 18 After Vac.	-5.784	3.522	-11.060	-1.077
Metro Frequency	-0.048	0.033	-0.117	0.011	CoronaVac * Obesity	-0.183	0.177	-0.487	0.143
alpha1	-0.898	0.162	-1.251	-0.657	CoronaVac * Hypertension	0.385	0.086	0.250	0.541
alpha2	-27.560	6.021	-36.790	-18.370	CoronaVac * Diabetes	0.359	0.164	0.088	0.696
alpha3	-2.127	0.348	-2.822	-1.509	CoronaVac * Chronic Pulmonary D.	-0.068	0.186	-0.421	0.245
alpha4	-24.110	3.380	-31.280	-19.940	CoronaVac * Chronic Cardiovascular D.	-0.539	0.224	-0.886	-0.151
alpha5	-32.200	5.496	-40.320	-23.380	Booster ChAdOx1 * Gender	-9.402	3.394	-14.520	-3.026
Gender	0.201	0.051	0.098	0.299	Booster ChAdOx1 * Age Range (40,49)	9.683	2.551	5.711	14.140
Study Fortnight	0.165	0.007	0.155	0.178	Booster ChAdOx1 * Age Range (50,59)	7.471	2.867	1.360	13.040
Age Range (40,49)	0.388	0.073	0.243	0.529	Booster ChAdOx1 * Age Range (60,69)	0.376	2.931	-5.252	4.680
Age Range (50,59)	0.503	0.057	0.378	0.594	Booster ChAdOx1 * Age Range (70+)	-3.087	4.991	-12.620	5.735
Age Range (60,69)	0.247	0.090	0.046	0.383	Booster ChAdOx1 * Fortnight 1 After Vac.	-0.710	5.237	-10.930	7.322
Age Range (70+)	0.477	0.107	0.286	0.657	Booster ChAdOx1 * Fortnight 2 After Vac.	6.410	7.800	-2.697	21.610
Times Leaving Home per week (3,5)	0.285	0.082	0.135	0.433	Booster ChAdOx1 * Fortnight 3 After Vac.	-15.810	2.848	-21.890	-11.970
Times Leaving Home per week (6,7)	0.156	0.123	-0.029	0.384	Booster ChAdOx1 * Fortnight 4 After Vac.	-4.065	8.837	-15.420	8.818
Times Leaving Home per week (7+)	-0.147	0.149	-0.414	0.110	Booster ChAdOx1 * Fortnight 5After Vac.	-16.380	8.373	-32.880	-5.965
Medium Perception of Risk	-0.043	0.051	-0.148	0.055	Booster ChAdOx1 * Fortnight 6 After Vac.	2.604	4.530	-5.710	9.928
High Perception of Risk	0.481	0.097	0.297	0.641	Booster ChAdOx1 * Fortnight 7 After Vac.	14.640	3.671	5.888	19.490
BNT162b2 * Gender	0.552	3.254	-7.697	4.373	Booster ChAdOx1 * Fortnight 8 After Vac.	-11.050	6.605	-19.860	0.593
BNT162b2 * Age Range (40,49)	1.630	3.016	-2.317	7.540	Booster ChAdOx1 * Fortnight 9 After Vac.	5.838	6.572	-5.998	13.060
BNT162b2 * Age Range (50,59)	-5.555	5.510	-14.480	3.502	Booster ChAdOx1 * Fortnight 10 After Vac.	-16.900	6.532	-28.980	-6.137
BNT162b2 * Age Range (60,69)	-4.427	1.636	-7.667	-2.381	Booster ChAdOx1 * Fortnight 11 After Vac.	-6.694	2.595	-11.230	-0.705
BNT162b2 * Age Range (70+)	8.080	2.295	4.103	11.850	Booster ChAdOx1 * Fortnight 12 After Vac.	-6.641	6.694	-19.580	3.922
BNT162b2 * Fortnight 1 After Vac.	0.185	2.728	-5.065	5.915	Booster ChAdOx1 * Fortnight 13 After Vac.	1.177	2.280	-3.061	5.216
BNT162b2 * Fortnight 2 After Vac.	-10.120	3.442	-15.910	-4.697	Booster ChAdOx1 * Fortnight 14 After Vac.	5.293	2.607	0.122	10.840
BNT162b2 * Fortnight 3 After Vac.	12.550	4.142	6.210	19.650	Booster ChAdOx1 * Fortnight 15 After Vac.	-9.684	2.091	-13.180	-7.004
BNT162b2 * Fortnight 4 After Vac.	-3.988	6.258	-14.600	5.288	Booster ChAdOx1 * Fortnight 16 After Vac.	-4.565	3.427	-10.890	0.012
BNT162b2 * Fortnight 5After Vac.	3.690	1.929	0.104	6.983	Booster ChAdOx1 * Fortnight 17 After Vac.	-19.340	4.660	-27.460	-11.830
BNT162b2 * Fortnight 6 After Vac.	1.416	2.540	-3.748	5.761	Booster ChAdOx1 * Fortnight 18 After Vac.	3.080	4.418	-2.306	11.970
BNT162b2 * Fortnight 7 After Vac.	-2.756	2.461	-6.591	2.256	Booster ChAdOx1 * Obesity	-3.477	4.141	-10.270	4.117
BNT162b2 * Fortnight 8 After Vac.	-11.400	5.158	-17.370	-1.404	Booster ChAdOx1 * Hypertension	4.540	3.451	-0.841	10.800
BNT162b2 * Fortnight 9 After Vac.	8.445	3.289	2.242	12.690	Booster ChAdOx1 * Diabetes	-1.778	3.381	-8.450	3.029
BNT162b2 * Fortnight 10 After Vac.	0.638	3.735	-5.599	7.626	Booster ChAdOx1 * Chronic Pulmonary D.	-1.451	2.747	-7.390	3.018
BNT162b2 * Fortnight 11 After Vac.	-6.900	5.064	-13.760	0.722	Booster ChAdOx1 * Chronic Cardiovascular	0.458	2.859	-5.336	6.491
BNT162b2 * Fortnight 12 After Vac.	5.526	1.788	2.767	8.480	Booster BNT162b2 * Gender	-6.356	3.097	-12.040	-1.189
BNT162b2 * Fortnight 13 After Vac.	-9.909	2.047	-13.940	-6.743	Booster BNT162b2 * Age Range (40,49)	-17.170	10.046	-35.890	-3.062
BNT162b2 * Fortnight 14 After Vac.	6.660	2.950	0.700	10.570	Booster BNT162b2 * Age Range (50,59)	7.429	4.139	3.017	16.530
BNT162b2 * Fortnight 15 After Vac.	4.677	2.372	0.107	7.889	Booster BNT162b2 * Age Range (60,69)	8.005	4.577	-0.789	14.960
BNT162b2 * Fortnight 16 After Vac.	6.932	1.651	3.394	9.818	Booster BNT162b2 * Age Range (70+)	-6.384	2.570	-9.758	-0.335
BNT162b2 * Fortnight 17 After Vac.	-5.074	3.356	-11.370	0.962	Booster BNT162b2 * Fortnight 1 After Vac.	-4.053	2.572	-9.586	-0.714
BNT162b2 * Fortnight 18 After Vac.	7.417	3.628	0.195	12.660	Booster BNT162b2 * Fortnight 2 After Vac.	15.500	4.926	8.000	22.300
BNT162b2 * Obesity	-6.979	3.911	-14.090	-1.476	Booster BNT162b2 * Fortnight 3 After Vac.	23.450	6.816	7.975	32.760
BNT162b2 * Hypertension	-7.022	3.634	-13.710	-1.134	Booster BNT162b2 * Fortnight 4 After Vac.	11.650	6.077	-0.568	22.910
BNT162b2 * Diabetes	-1.686	5.786	-13.350	4.942	Booster BNT162b2 * Fortnight 5After Vac.	7.606	3.818	1.869	12.720
BNT162b2 * Chronic Pulmonary D.	-8.913	4.495	-16.790	-3.350	Booster BNT162b2 * Fortnight 6 After Vac.	-2.730	3.322	-8.714	3.691
BNT162b2 * Chronic Cardiovascular D.	11.170	2.628	4.349	15.040	Booster BNT162b2 * Fortnight 7 After Vac.	-9.278	3.221	-13.910	-3.712
CoronaVac * Gender	-0.447	0.091	-0.595	-0.244	Booster BNT162b2 * Fortnight 8 After Vac.	4.512	3.220	0.398	11.360
CoronaVac * Age Range (40,49)	0.191	0.111	0.023	0.422	Booster BNT162b2 * Fortnight 9 After Vac.	-6.265	3.083	-12.800	-1.644
CoronaVac * Age Range (50,59)	-0.022	0.122	-0.227	0.180	Booster BNT162b2 * Fortnight 10 After Vac.	-11.170	8.245	-23.550	-0.729
CoronaVac * Age Range (60,69)	-0.049	0.131	-0.307	0.153	Booster BNT162b2 * Fortnight 11 After Vac.	10.660	3.997	3.069	18.750
CoronaVac * Age Range (70+)	0.018	0.148	-0.199	0.254	Booster BNT162b2 * Fortnight 12 After Vac.	12.710	4.767	1.865	19.690
CoronaVac * Fortnight 1 After Vac.	-5.115	1.961	-9.272	-2.826	Booster BNT162b2 * Fortnight 13 After Vac.	-12.140	2.998	-15.400	-5.197
CoronaVac * Fortnight 2 After Vac.	-10.610	4.868	-21.100	-3.312	Booster BNT162b2 * Fortnight 14 After Vac.	1.635	3.318	-4.424	7.068
CoronaVac * Fortnight 3 After Vac.	-4.819	2.104	-9.277	-1.396	Booster BNT162b2 * Fortnight 15 After Vac.	-7.325	2.839	-14.970	-2.773
CoronaVac * Fortnight 4 After Vac.	-7.768	4.216	-15.710	-1.038	Booster BNT162b2 * Fortnight 16 After Vac.	8.720	4.108	1.873	16.190
CoronaVac * Fortnight 5After Vac.	-0.248	0.563	-1.237	0.764	Booster BNT162b2 * Fortnight 17 After Vac.	10.500	4.983	-0.524	15.580
CoronaVac * Fortnight 6 After Vac.	0.880	0.339	0.280	1.473	Booster BNT162b2 * Fortnight 18 After Vac.	16.710	3.073	9.549	23.140
CoronaVac * Fortnight 7 After Vac.	1.287	0.321	0.749	1.994	Booster BNT162b2 * Obesity	-13.290	5.880	-21.750	-2.259
CoronaVac * Fortnight 8 After Vac.	1.601	0.342	1.049	2.096	Booster BNT162b2 * Hypertension	-1.402	2.877	-6.757	3.245
CoronaVac * Fortnight 9 After Vac.	1.737	0.357	1.155	2.451	Booster BNT162b2 * Diabetes	3.572	2.431	-0.748	7.276
CoronaVac * Fortnight 10 After Vac.	1.636	0.357	0.988	2.272	Booster BNT162b2 * Chronic Pulmonary D.	-17.760	5.817	-26.120	-6.489
CoronaVac * Fortnight 11 After Vac.	1.988	0.338	1.368	2.677	Booster BNT162b2 * Chronic Cardiovascula	-3.889	3.931	-12.090	1.340

Figure 4.8: Booster Logit Results Gran Santiago