ISSN 2683-9288

Science Reviews
from the end of the world

Volume 2
Number 2
April 2021

Delia Velekson - “Bosque petrificado” (1993)

Science Reviews
from the end of the world

Science Reviews - from the end of the world **is a quarterly publication that aims at providing authoritative reviews on hot research topics developed mainly by scientists that carry out their work far away from the main centers of science. Its research reviews are short, concise, critical and easy-reading articles describing the state of the art on a chosen hot topic, with focus on the research carried out by the authors of the article. These articles are commissioned by invitation and are accessible not only to hardcore specialists, but also to a wider readership of researchers interested in learning about the state-of-the-art in the reviewed subject. The reviews cover all fields of science and are written exclusively in English. They are refereed by peers of international prestige and the evaluation process follows standard international procedures.**

Centro de Estudios sobre Ciencia, Desarrollo y Educación Superior
538 Pueyrredón Av. - 2° C – Second building
Buenos Aires, Argentina - C1032ABS
(54 11) 4963-7878/8811
sciencereviews@centroredes.org.ar
www.scirevfew.net

Vol. 2, No. 2
April 2021

ISSN 2683-9288

Our cover: *Bosque petrificado* (1993), by Delia Velekson. From the “Artistas Plásticos con la ciencia” collection. Reproduced by permission from the Comisión Administradora Permanente de la Exposición de Arte Centro Atómico Constituyentes, Comisión Nacional de Energía Atómica.

Science Reviews
from the end of the world

Table of Contents

INTRODUCTION

IN THIS ISSUE

ARTICLES

INSTRUCTIONS FOR AUTHORS

Science Reviews
from the end of the world

Vol. 2, No. 2, April 2021 //

INTRODUCTION

Science in the Time of COVID-19

With the arrival of the pandemic, scientists from every country rearranged their priorities, and Argentina was not the exception. A myriad of scientific papers of variable quality were produced all over the world, and rushed for publication on the knowledge that in lieu of prompt diffusion many would have soon become obsolete. Science Reviews from the end of the world was very cautious; after all we publish only reviews of significant work, often carried out during long time spans.

Controversy has arisen on the validity of mathematical models to describe the spread of the pandemic, on the testing procedures and the ways to achieve significant tracking of the spread of the virus, on the validity of various therapies, on the nature and performance of various vaccines, on social and political issues about lockdown policies, etc.

In this issue we present just two papers, addressing the first issues in the above paragraph, mathematical models and testing strategies. They represent very thorough work that had clear application in the design of pandemic management in Argentina. Papers of equivalent quality on the other issues have been commissioned, and shall be published when available.

The research teams led by Guillermo Durán and by Roberto Etchenique represent well the paradigm of current research, as the result of the effort of multidisciplinary teams focused on problem solving.

We hope these papers prove of value to all readers.

Miguel A. Blesa

Bio

Miguel A. Blesa

Ph D in Chemistry, La Plata University. Formerly: Chemistry Manager (Atomic Energy Commission); Federal Undersecretary of Science and Technology; Senior Researcher (CONICET); Full Professor (San Martín University); President (Argentine Associaition for the Advancement of Science). Past President, Interciencia Associaition. Member of the National Academy of Science (Córdoba), National Academy of Exact, Natural and Physical Sciences (ANCEFN) and Academy of Sciences of Latin America. TWAS Prize in Chemistry (2004). Konex Merit Diploma in Chemsitry (2003). President for Argentina in the Bilateral French Argentine Cooperation Program (ECOS-SUD). Main research interests: chemistry of metal oxide particles immersed in water; environmental chemistry.

IN THIS ISSUE

Mathematical and Computational Initiatives from the University of Buenos Aires to Contribute to Decision-Making in the Context of COVID-19 in Argentina

Received: 12/28/2020 – Approved for publication: 03/21/2021

Mehrnoosh Arrar, Laouen Belloli, Ana M. Bianco, Leonardo Boechi, Rodrigo Castro, Guillermo Durán, Roberto Etchenique, Natalia B. Fernández, Luciana Ferrer, Diego Garbervetsky, Rodrigo Goldsmit, Carolina Grillo Vidal, Juan E. Kamienkowski, Pablo Laciana, Esteban Lanzarotti, Mario E. Lozano, Rodrigo Maidana, Mauricio Mendiluce, Sol Minoldo, Ezequiel Pecker-Marcosig, Leonardo Pepino, Armando E. Puerta, Rodrigo Quiroga, Guillermo Solovey, Marina Valdora and Mariano Zapatero

Pool Strategy for Surveillance Testing of SARS-CoV-2

Received: 12/26/2020 – Approved for publication: 02/23/2021

Felipe Marceca, Luciana Rocha-Viegas, Nicolás Pregi, Gabriela Barbas, Daniela Hozbor, Adali Pecci and Roberto Etchenique

Science Reviews
from the end of the world

Vol. 2, No. 2, April 2021 // pp. 6-40 /

Mathematical and Computational Initiatives from the University of Buenos Aires to Contribute to Decision-Making in the Context of COVID-19 in Argentina

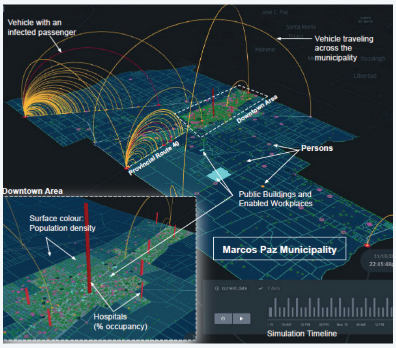
Mehrnoosh Arrar[1], Laouen Belloli[2], Ana M. Bianco[1], Leonardo Boechi[1], Rodrigo Castro[2], Guillermo Durán[1*], Roberto Etchenique[3], Natalia B. Fernández[4], Luciana Ferrer[2], Diego Garbervetsky[2], Rodrigo Goldsmit[6], Carolina Grillo Vidal[6], Juan E. Kamienkowski[2], Pablo Laciana[2], Esteban Lanzarotti[2], Mario E. Lozano[7], Rodrigo Maidana[8], Mauricio Mendiluce[6], Sol Minoldo[5], Ezequiel Pecker-Marcosig[2], Leonardo Pepino[2], Armando E. Puerta[2], Rodrigo Quiroga[9], Guillermo Solovey[1], Marina Valdora[1] and Mariano Zapatero[2]

[1] CONICET-Universidad de Buenos Aires. Instituto de Cálculo (IC). Buenos Aires, Argentina
[2] CONICET-Universidad de Buenos Aires. Instituto de Ciencias de la Computación (ICC). Buenos Aires, Argentina
[3] CONICET-Universidad de Buenos Aires. DQIAyQF - Instituto de Química-Física de los Materiales, Medio Ambiente y Energía (INQUIMAE). Buenos Aires, Argentina
[4] CONICET-Universidad de Buenos Aires. Instituto de Biociencias, Biotecnología y Biología Traslacional (iB3). Buenos Aires, Argentina
[5] CONICET-Universidad de Córdoba. Centro de Investigaciones y Estudios sobre Cultura y Sociedad (CIECS). Córdoba, Argentina
[6] FCEyN-Universidad de Buenos Aires. Buenos Aires, Argentina
[7] Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Quilmes, Argentina
[8] FCE, Universidad Nacional de La Plata, La Plata, Argentina
[9] CONICET-Universidad Nacional de Córdoba, Instituto De Investigaciones en Físico-Química de Córdoba (INFIQC), Córdoba, Argentina

** Corresponding author. E-mail: gduran@dm.uba.ar*

Abstract

With the arrival of the pandemic in Argentina in March 2020, a working group of scientists from two institutes belonging to the Faculty of Exact and Natural Sciences of the University of Buenos Aires and CONICET, together with colleagues from different academic institutions in the country, decided to put forth our experience and knowledge in data science and associated disciplines, towards helping with decision-making in the context of COVID-19. Data analysis within Argentina and other countries, scenario simulation, as well as rapid response projects- mainly in the province of Buenos Aires- were all within the scope of our aim. This review article outlines some of the activities carried out by our team throughout these pandemic months.



Keywords: data-driven COVID-19 response, agent-based simulations, effective reproductive number, prognosis models, Buenos Aires

Introduction

The COVID-19 pandemic, caused by the infection of the SARS-CoV-2 virus, has caused over one million deaths worldwide during 2020 (World Health Organization, n.d.). Thousands of academic groups around the world have been working on issues related to the pandemic since the beginning of 2020; the group of Neil Ferguson at Imperial College London being of the most cited ("Report 1 - Estimating the Potential Total Number of Novel Coronavirus (2019-nCoV) Cases in Wuhan City, China" n.d.)) along with the Franco-Spanish engineer Tomás Pueyo, with his renowned anecdote of the hammer and dance as key tools to face COVID-19 (Pueyo 2020).

Due to the great difficulty that health organizations have faced in finding effective treatments and vaccines, and thus containing the progression of the pandemic, new approaches have emerged that seek to use information (previously existing or generated during the pandemic) to help the decision-making process of institutional stakeholders. These are multidisciplinary approaches that consider logistical, economic, political and social aspects.

From the applied math and computer science institutes -Instituto de Cálculo (IC) and the Instituto de Ciencias de la Computación (ICC)- of the Faculty of Exact and Natural Sciences of the University of Buenos Aires (FCEyN-UBA) and CONICET, a multidisciplinary group with extensive experience in the development and research of mathematical-computational tools, who, alongside scientists from other national academic institutions, established collaborations with different jurisdictions, predominantly in the Province of Buenos Aires, in order to facilitate decision-making during the pandemic.

In this article we summarize the main activities carried out by this scientific nucleus. It is a broad set of contributions, ranging from data analysis and recommendations to different districts, to rapid response projects in territory. These developments enable the acquisition of data and organization of available resources; quantify and predict the progress of the pandemic; and facilitate the quantitative analysis of scenarios to predict the effects of certain public policies. Several of the activities respond to specific needs or requests made by the main actors in this pandemic, including hospital directors, health ministry authorities, mayors, and governors, and were designed and co-developed with these actors. Other contributions were academic initiatives that were made public or available to decision makers.

The activities carried out (outlined below) are divided in two parts; in the first we highlight data analysis initiatives, and in the second the rapid response projects.

Data analysis:

- Evaluation of different models for the estimation of key indicators used by the authorities, such as the basic reproductive number (R_0) and the effective number (R_{ef}) (Section 1.1), the doubling time (DT) of the number of cases (Section 1.2).
- Analysis of patient data collected by the Ministry of Health of the Province of Buenos Aires for the development of models for Automatic Prognosis (Section 1.3.1), and to be able to estimate the probability that an infected individual needs hospitalization in Intensive Care Units (ICU) (Section 1.3.2).
- Development of a portal where various indicators of the evolution of the pandemic are presented, at different levels of resolution: national, provincial, municipal and other geographical or administrative groups (Section 1.4)
- Estimates of prevalence at the national, provincial and municipal levels based on Infection Fatality Rate (IFR) estimates and information on death counts (Section 1.5).
- Analysis of different types of immunity and their consequences, considering different indicators (Section 1.6).

Rapid response projects:

- Development of an application that permits allocation of hospital beds within a network of hospitals, and its use in the southeastern region of Greater Buenos Aires (Section 2.1)
- The Covid Prevention Center at the Faculty of Exact and Natural Sciences of the University of Buenos Aires (Section 2.2)
- Agent-based simulations in the Marcos Paz district of the Province of Buenos Aires (Section 2.3)

Fundamental variables such as data privacy, responsible use of data, as well as the urgency of solutions and the limited budget available were all taken into account throughout this challenging and necessary effort to mitigate the negative impact of the pandemic in Argentina.

1. Data analysis

1.1. Estimation of the reproductive number R of the pandemic over time

Background:

A widely accepted metric to characterize incidence is the reproductive number, or R, defined as the average number of encounters with successful transmissions generated by each infectious individual.

At the beginning of an epidemic, the entire population can be considered susceptible and the dynamics of social encounters is not yet affected by sanitary measures. In this situation R assumes its value known as the basic reproductive number, or R_0. Afterwards, when there may be people immune to contagion or the population adopts special hygiene and distancing measures, R is called the effective reproductive number, or R_{ef}, usually satisfying $R_0 \geq R_{ef} \geq 0$

R_{ef} serves as a guideline to synthesize the state of the epidemic in terms of the dynamics of change in the number of active infected: R_{ef} values greater than one represent an exponential (or sub-exponential) growth phase whereas R_{ef} values less than one account for a phase of decline (or suppression) in which the epidemic tends towards its extinction.

Objective:

Our specific objective is to provide a methodology for the estimation of R_{ef} in distinct parts of the country.

Methodology:

There are several techniques to determine the R_{ef} value from the evidence that accumulates as an epidemic develops.

One family of methods is based on models that explicitly define the mechanics of contagion in a population, such as compartmental models (Vynnycky and White 2010). The population is divided into categories (compartments) such as Susceptible (S), Infected (I) and Recovered (R) (in its most compact version known as the SIR model), and a system of differential (continuous time) or difference (discrete time) equations define the dynamics of the evolution from one compartment to another. In this framework, $R_{ef}(t)$ is a variable that depends on system parameters, such as in the following system of differential equations:

$dS/dt = -\beta(t)*I(t)*S(t)$ (average infection rate of susceptibles)

$dI/dt = \beta(t)*I(t)*S(t)-\gamma*I(t)$ (average rate of change of infectious infections)

$dR/dt = \gamma*I(t)$ (average rate of recovery and/or death)

$N=S(t)+I(t)+R(t)$ (constant total population)

$R_{ef}(t)=\beta(t)/\gamma*S(t)$ (effective reproductive number, dimensionless)

Here β (t) expresses the average per-capita infection rate, which is dependent on social behavior, whereas γ is the average rate of recovery and/or death. Given that S (t) is the number of susceptible individuals and that it decreases monotonically as the epidemic progresses, this value is one of the determining factors so that, after a certain time, Ref (t) crosses the threshold of value 1 indicating the beginning of the extinction of the process.

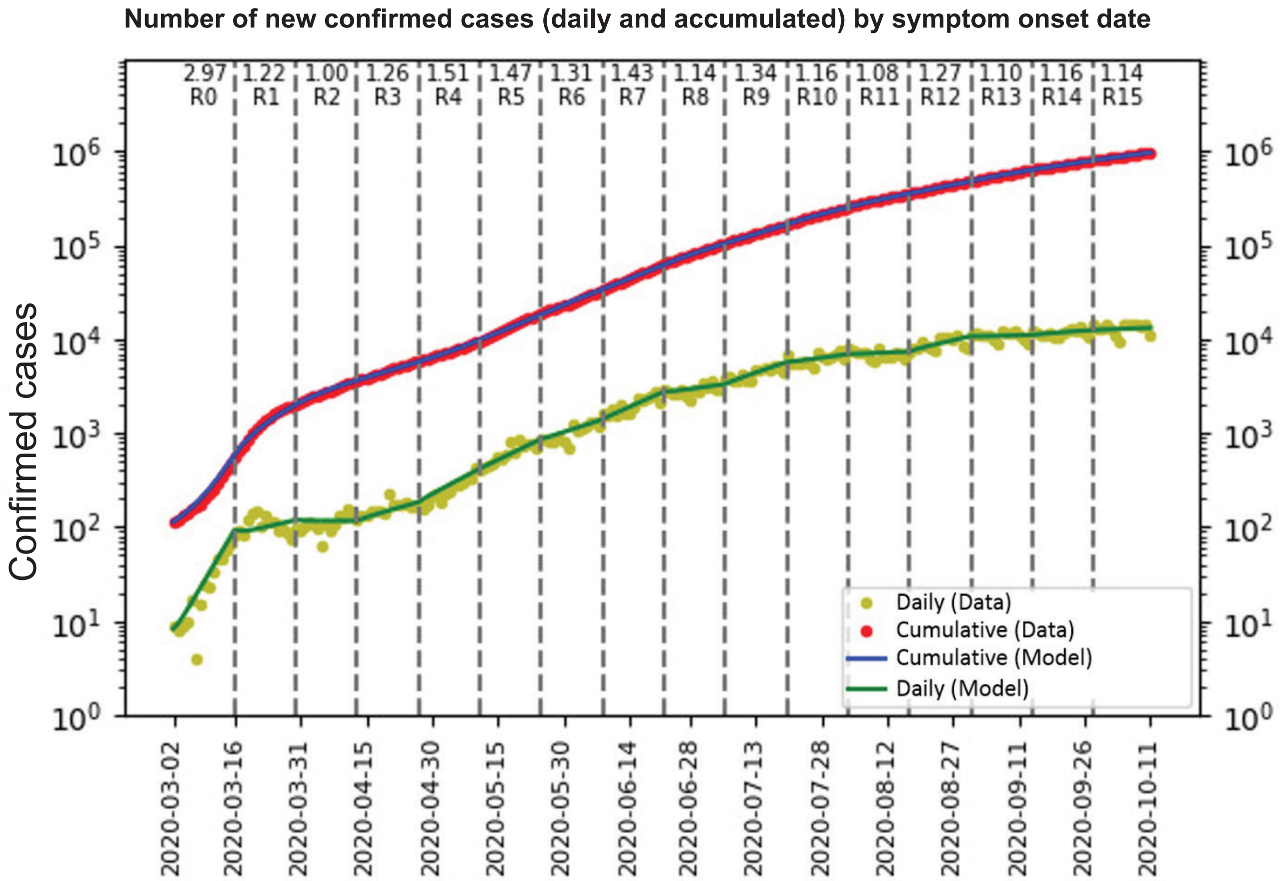
We opted for this type of model as it allows us to better understand the evolution of the system and make projections within certain time windows in the future, assuming hypothetical scenarios of change, typically through non-pharmacological interventions (i.e. public policies) reflected as changes in β(t).

We also apply an extension of this model, known as SEIR, which adds an Exposed (E) compartment to represent the infected population that is not yet contagious, to estimate $R_{ef}(t)$ as the exposed subgroup is relevant in the evolution of the COVID-19 pandemic.

Results:

Assuming a known gamma (in its average value) and having the data for the reported number of infected individuals Ir (t), we perform a constant parameter fit to time periods, in such a way that the norm of the mean square error between I (t) (model) and Ir (t) (data) is minimized. This procedure implies identifying the minimum amount of time points tk, with k = 0, ..., M in which it is necessary to subdivide the time periods. Another option is to choose fixed divisions *a priori* for the time periods. The first strategy aims to identify the points in time when there were significant changes in social dynamics whereas the second is more useful to compare identical time periods between different jurisdictions.

Figure 1 shows an example of a fit for Argentina choosing 14-day stretches (a sufficient period to capture the data changes in the number of newly infected, recovered, and deaths). Using a semilogarithmic scale, and using official data from cases reported by symptom onset date, the fit of the model (filled line) with the data (points) is observed. The new daily confirmed cases and the accumulated confirmed cases are shown, together with the successive values fit for R_{ef} (tk).



Number of new confirmed cases (daily and accumulated) by symptom onset date

Figure 1. Evolution of Ref(tk) for Argentina fitted to a SEIR-type model.

Applying the same criteria for several jurisdictions, we can compare their temporal evolutions as shown in Figure 2.

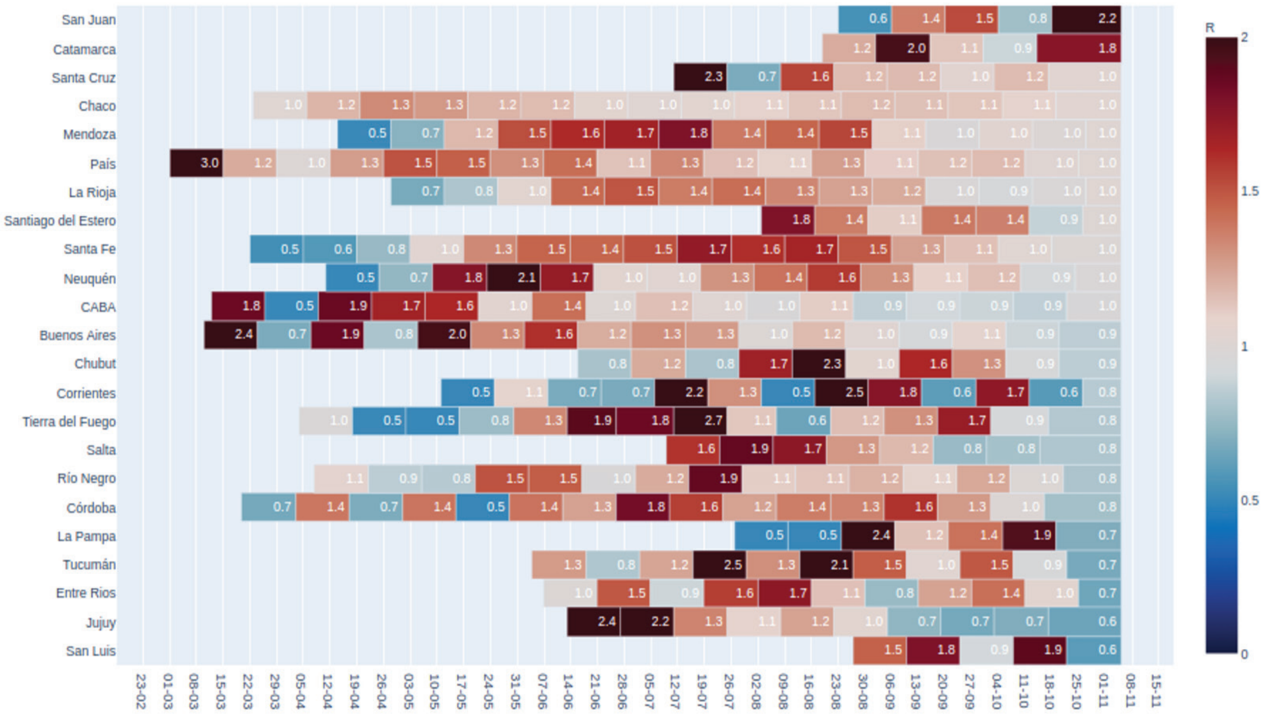


Figure 2. Temporal heat map showing evolutions of the effective reproductive number Ref(t) for the Argentine provinces (snapshot of 11/15/2020). This display is part of the automated analysis available on the website described in Section 1.4.

Conclusions:

The application of the variable-time-period and fixed-time-period methods showed consistent results. In turn, the daily updated online visual presentation proved to be informative for comprehensive analysis of the temporal progression of the stages of the epidemic in different jurisdictions of the country. The flexibility of easily defining new geographic regions of interest to be monitored through the effective reproduction number makes it easier to respond quickly to changing information demands.

1.2. Doubling times

Background:

A parameter widely used by governments around the world to know the rate at which infections are growing is the doubling time of cases (DT). In Argentina, the national government has defined phases of the pandemic also according to the DT ("Fases de Administración Del Aislamiento" 2020). According to this criterion, if in a district the doubling time exceeds 25 days, that district could change the phase in the isolation/lockdown process and, consequently, reopen activities that had been prohibited. However, uncertainty associated with inferred DT is rarely reported. This is important since, for example, a doubling time of 15.2 ± 0.1 days should be interpreted differently than one of 15.2 ± 2 days. Therefore, for a reliable diagnosis of the progress of the pandemic, it is also necessary to know the precision of the estimate.

Objective:

Our objective is to estimate the doubling time of confirmed cases and the confidence interval for this estimate. We propose to do this using two variants of the exponential growth model.

Methodology:

Let y_n be the number of confirmed cases up to day n. To estimate the doubling time on day n, data from the last 7 days are taken, $y_{n-6},\ldots,y_{n,}$ and fit to the following models:

- model 1: $\log(y_n) = \alpha + \beta n + \varepsilon_n$
- model 2: $y_n = \gamma \exp(\beta n) + \varepsilon_n$

The expected doubling time is $\log(2)/\beta$. Using the least squares method we calculate an estimator of β, which we call β^*, and we deduce that an estimator for the doubling time is $\log(2)/\beta^*$.

In model 1, the distribution of β^* is known, assuming that the errors have a normal distribution. In model 2, however, it is only possible to find the approximate or asymptotic distribution of β^* and, since we want to base our estimate on only 7 observations, this approximation may not be very good. Nevertheless, we calculated the confidence intervals under both models and did a Monte Carlo study to study their level of empirical coverage. We performed experiments with normal errors and with heavy-tailed errors (t_3 distribution).

Results:

The results of the Monte Carlo study are summarized in Table 1. Based on these results, we decided to use only the intervals calculated with model 1.

A confidence interval of level $1-\alpha$ for the doubling time under this model is

$$[\log(2)/(\beta^* + t_{n-2,\,\alpha/2}\,SE(\beta^*)),\ \log(2)/(\beta^* - t_{n-2,\,\alpha/2}\,SE(\beta^*))].$$

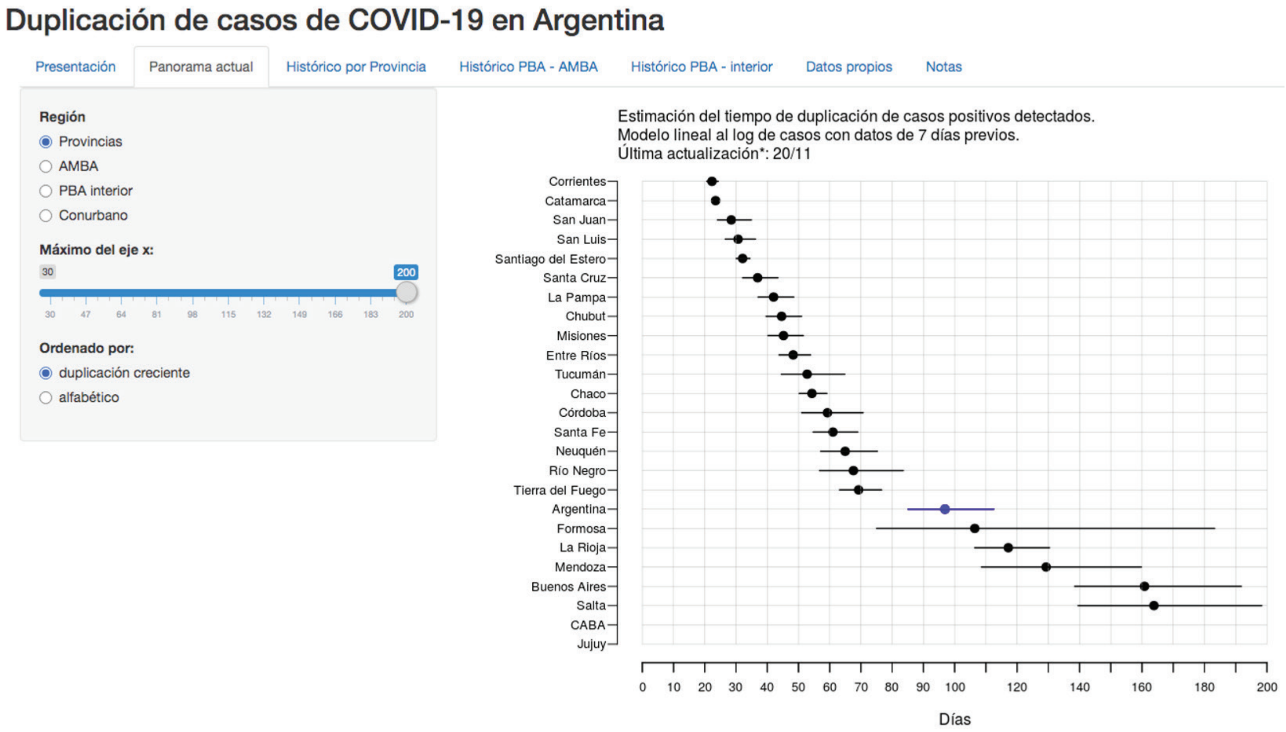
We have observed that on some days the data entry is delayed and compensated for in the following days, causing a variability in the data that does not reflect the reality of the epidemic. For this reason we decided to replace the random variables y_n with variables $z_n = (y_n + y_{n-1} + y_{n-2}) / 3$.

Preliminary simulations showed that 7 days were sufficient and that confidence intervals could not be calculated with the non-linear method. For this reason, the subsequent analysis with real data was done with the linear estimation model.

Table 1. Estimation errors obtained by Monte Carlo simulations for each of the two models

	Model 1	Model 2
Normal error	0.9499	0.8916
t_3 Error	0.9497	0.8890

This method was used to estimate the doubling time of different districts of Argentina (see Figure 3). We created an application, available to the public, in which these results can be visualized ("Duplicación de Casos de COVID-19 En Argentina" n.d.):



Duplicación de casos de COVID-19 en Argentina

Región
● Provincias
○ AMBA
○ PBA interior
○ Conurbano

Máximo del eje x: 200

Ordenado por:
● duplicación creciente
○ alfabético

Estimación del tiempo de duplicación de casos positivos detectados.
Modelo lineal al log de casos con datos de 7 días previos.
Última actualización*: 20/11

Figure 3. Screenshot of the application that can be seen at: https://institutodecalculo.shinyapps.io/tdcovid/

Conclusions:

The estimation of the doubling time was similar for the two fitting methods. It is important to highlight that the doubling time is dependent on the model and should be interpreted as a measure of the "instantaneous velocity" of growth of infections, not as a prediction.

1.3. Analysis and prediction of the evolution of cases

Forecasting systems (or prognosis models) can assist health personnel in decision-making, by identifying cases for which immediate attention or frequent follow-up is required. In turn, the implementation of these models allows us to explore which are the most important attributes in the automatic decision, and in this way to discover potential new early markers of the disease.

The vast majority of people infected with the new coronavirus SARS-CoV-2 develop mild or moderate forms of COVID-19. However, some present severe symptoms that require admission to the Intensive Care Unit (ICU). In this context, it is necessary to identify who are those patients who might require an ICU bed and for this purpose we study which comorbidities have the greatest association with the fact of needing ICU admission.

In this section we describe two studies carried out using data provided by the Ministry of Health of the Province of Buenos Aires, within the framework of a cooperation agreement with the Faculty of Exact and Natural Sciences of the UBA. In section 1.3.1, the objective is to predict the probability that an individual will develop a severe case of the disease, using age, sex, and comorbidities as covariates. In section 1.3.2, on the other hand, the objective is to predict whether an individual should be admitted to the ICU or not in the course of the disease. Although the majority of severe cases are admitted to the ICU, not all are, which is evidenced by the fact that there is a percentage of deceased individuals who did not go through an ICU. This may explain the small differences in the conclusions in both sections regarding risk factors.

1.3.1. Automatic prognosis

Background:

Given the diversity of data and experiences, the existing prognosis models within the framework of COVID-19 differ both in the variable to be predicted and in the attributes used. Regarding the variables to predict, most of the models seek to predict mortality, however models have also been developed to predict severity, for example, in the case of hospital stays longer than 10 days or the need for ICU (Wynants et al. 2020; Gupta et al. 2020). This is particularly important when resources are scarce. A recent systematic review compares 22 of the existing models for predicting deterioration throughout the hospital stay and mortality. This review uses a sample for validation of 411 cases and concludes that no model exceeds the univariate analysis of the most explanatory attribute (Gupta et al. 2020). These are the oxygen saturation upon entering the hospital and the age for deterioration and mortality, respectively. In another article, Yan, Zhang, et al. (L. Yan et al. 2020) use a model based on decision trees in a sample of approximately 500 cases where it shows that using blood studies they obtain F1 values close to 1 in the prediction of mortality (L. Yan et al. 2020).

The reported studies of automatic prognostic models generally use small samples for training and validation, between 100 and 500 cases. A recent study explored the association of patient attributes with mortality in a much larger sample ($N \sim 10^7$ people; (Williamson et al. 2020)). This study showed that age is the variable that best explains risk, increasing with age. Among the comorbidities, those with the greatest impact are diabetes, cancer, organ transplants, deficiencies in kidney or liver function, neurological diseases, previous strokes, or deficits in the immune system. Finally, other demographic or socioeconomic variables such as sex, ethnicity or socioeconomic context have a moderate incidence on risk. The findings in this study are consistent with automatic prognostic studies in more limited populations (Gupta et al. 2020; Wynants et al. 2020).

Objective:

Develop a prototype of an automatic forecasting model of the severity of the condition that a certain person will have when contracting Covid.

Methodology:

A Random Forest model was developed for the prediction of disease severity, trained on a sample of $\sim 10^5$ cases, based on attributes such as age, sex, comorbidities, and symptoms.

Data provided by the Ministry of Health of the Province of Buenos Aires as of August 28 were used, in particular:

- Age and sex of the person
- Symptoms (19 values): New-onset anosmia, new-onset dysgeusia, dyspnea, odynophagia, fever, cough, arthralgia, headache, mental confusion, seizures, diarrhea, abdominal pain, chest pain, conjunctival injection, irritability, malaise, myalgia, food refusal, and vomiting.
- Comorbidities (22 values): Diabetes, pregnancy, chronic liver disease, previous neurological disease, previous cancer disease, asthma, chronic obstructive pulmonary disease (COPD), low birth weight, previous bronchiolitis, acute dialysis, chronic dialysis, ex-smoker, smoker, high blood pressure, congenital immunosuppression or acquired, heart failure, renal failure, previous community-acquired pneumonia (CAP), obesity, prematurity, tuberculosis, and a characteristic that indicates the lack of comorbidities.
- Hospitalization and death data: An indicator of whether the patient was hospitalized and, if so, whether it was in intensive care, and whether the patient died.
- All the attributes, except for age, which is measured in years, and fever, which takes three values (no fever, fever less than 38 ºC and fever greater than 38 ºC), take binary values where 1 codes if the symptom or comorbidity is present in the patient. Only confirmed COVID-positive cases were considered.

Random Forest, a standard machine learning method, chosen for its robustness, flexibility and interpretability, was used for modeling. The model was trained with the attributes listed above as input to predict the severity of each case. Severity was defined as mild if the patient recovers with or without hospitalization but without intensive care, and as severe if he/she requires intensive care and / or dies. We want to emphasize that this definition is an approximation to the severity of the case and it is not perfect since it is possible that there are cases that were admitted to the ICU only preventively without the condition being so serious (for example, in the case of pregnant women or small children). During model training, the cost corresponding to each sample was weighed by the inverse of the frequency corresponding to its class to compensate for the imbalance between the two classes.

For the development of the models, the cases with a symptom onset date (SOD) prior to July 31, 2020 were used, while for the final evaluation we used the data with SOD from August 1 to 15, 2020, discarding the more recent cases, since it is not yet known whether they will become severe cases. Only cases with at least one symptom were included in the study. The development data included 108,249 cases, of which 4.7% were severe cases. The selection of parameters was carried out with the cross-validation method on the development data. The evaluation was carried out by training a new model on the development data and calculating the performance on the evaluation data, which included 47,014 cases, of which 2.1% were severe. The percentage of severe cases is different in the developmental and evaluation data because the data were divided by date, and, in this data, the percentage of severe cases decreased over time. In order to understand which attributes were most relevant to the task, a sequential attribute selection algorithm was used (Mustafa 2017). Starting with the model that uses all the attributes of interest, attributes are discarded one at a time, discarding at each stage the one that least degrades or most improves system performance in development data. The algorithm thus generates an ordered list of attributes according to their importance for ranking in the development data.

Results:

Figure 4 shows the recall (percentage of serious cases that are detected as such) versus precision (percentage of cases detected as serious that were actually serious) curves obtained while varying the decision threshold applied to the system's outputs for models trained with different lists of input attributes. The figure shows that age is an excellent predictor of the severity with which a case will develop and that sex adds relevant information to predict the severity of the case. These conclusions coincide with what was found in the literature (Williamson et al. 2020). On the other hand, adding comorbidities to age improves performance more than adding sex, and further including sex in addition to comorbidities makes only a marginal difference. This suggests that the information that the system extracts from the sex attribute is at least partially correlated with comorbidities. The resulting analysis of the attribute selection algorithm suggests that the most important comorbidities for classification are diabetes, obesity, kidney failure and previous oncological disease, which were also found to be relevant for prognosis in the literature. Finally, by adding the symptoms to the model, a significant gain is obtained. In this case, the selection algorithm indicates that dyspnea, odynophagia, and anosmia are the most relevant symptoms for classification. In the case of odynophagia and anosmia, the existence of the symptom occurs more frequently in mild cases than in severe cases, which had already been observed in the literature (C. H. Yan et al. 2020).

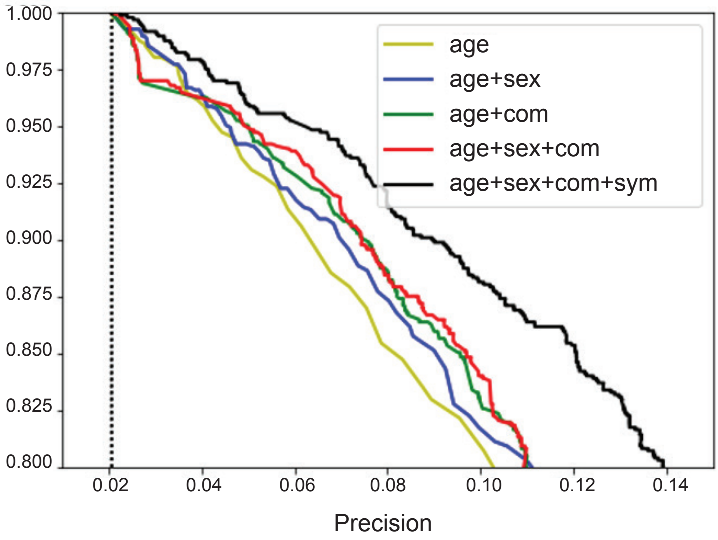


Figure 4. Recall and precision curves in evaluation data using different input attribute lists. The vertical dotted line corresponds to the percentage of severe cases. The references 'com' and 'sym' indicate comorbidities and symptoms, respectively.

Other symptoms and comorbidities in addition to those selected as important by the algorithm strongly correlate with the severity of the case, such as dialysis or heart failure when analyzed independently of the others. However, it is possible that these attributes are not chosen because they are highly correlated with others that the model does select, such as age. The next section (1.3.2) describes a work focused more specifically on the analysis of risk factors.

Conclusions:

The automatic prognosis system described in this section could be used to perform triage in cases of saturation of the health system that require postponing the care of some patients for the benefit of others with a worse prognosis. It is important to clarify that, since certain comorbidities that significantly affect prognosis occur very infrequently in the data, the model is probably not learning to correctly predict severity for those cases. For this reason, the output of the system should only be taken as a recommendation that must be confirmed by a health professional.

1.3.2. Comorbidities and risk of ICU admission due to COVID-19

Background:

In this section we aim to estimate the probability that an infected individual needs ICU admission using age, sex and the 22 comorbidities detailed in Section 1.3.1 as covariates. Our interest is focused on the analysis of risk factors.

This study is based on the data of confirmed cases collected by the Ministry of Health of the Province of Buenos Aires reported on July 7, 2020. In this section, the data analyzed corresponds to the confirmed cases of individuals under the age of 90 in the Province of Buenos Aires up until June 22, 2020, which comprises a total of 22,306 confirmed cases under study. In this analysis we discarded the confirmed cases between June 23 and July 7 since, throughout the development of the pandemic in Argentina, we have observed that, since a case is confirmed, it is necessary to allow a period of time to pass for the cases to evolve. The data indicate that for more than 99% of the individuals who were admitted to the ICU, this hospitalization occurred in 15 days or less from the confirmation of the diagnosis. An analysis of these times based on data from the entire country can be seen at http://covid.ic.fcen.uba.ar:3838/apps_ic/tiempoeventos/. The proportion of patients in this sample who required intensive care was 0.025 (2.5%).

Methodology:

In order to explore the association between the fact that a COVID-19 patient requires ICU or not and comorbidities, classification trees were used that constitute a descriptive tool based on graphs that use the branching method to represent the possible outcomes of a binary decision (James et al. 2013). A summary measure of the importance of the explanatory variables in a tree can be computed by decreasing the Gini index associated with each variable. A graphic representation of the importance of comorbidities, age and sex is presented in Figure 5. These measures were calculated from 1000 trees built using resampling of the original data and were computed as the average of the importance of each variable relative to the one of greatest importance. This descriptive study suggests age, obesity and hypertension as the variables of greatest importance.

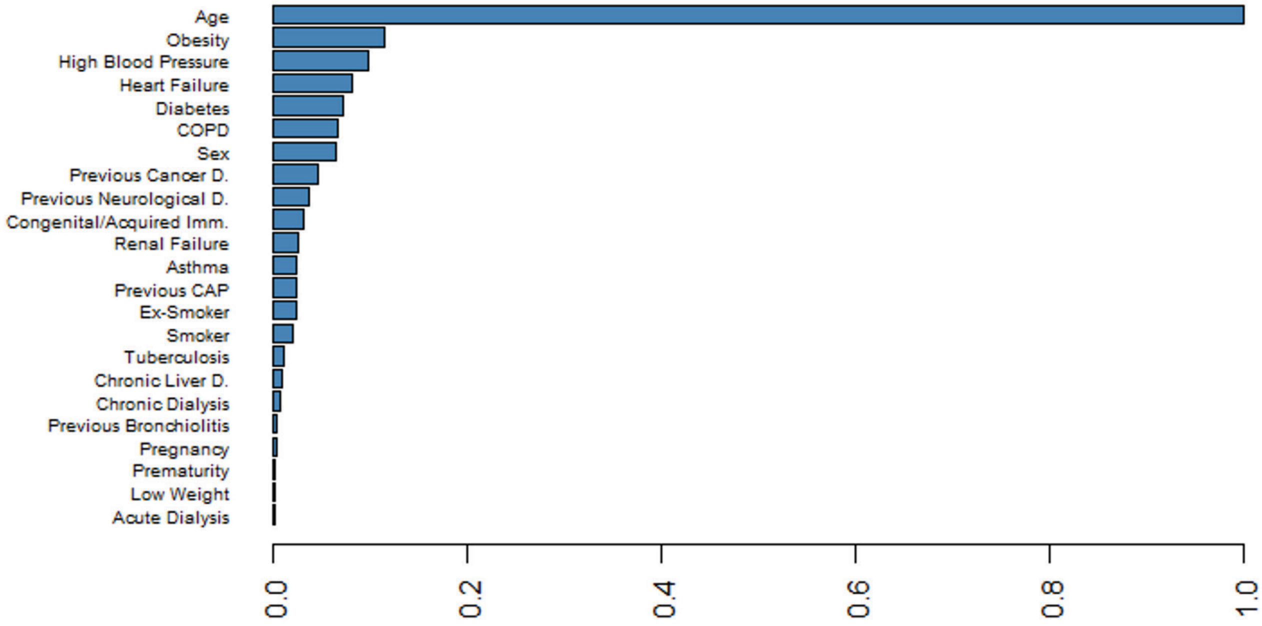


Figure 5. Importance of risk factors from 1000 trees built using resampling of the original data.

In order to make statistical inference, a fit was performed by logistic regression (Hosmer, Lemeshow, and Sturdivant 2013) categorizing the data into two classes, one representing those who required ICU and the other those who did not, and using comorbidities as explanatory variables. To that end, if $y \in \{0, 1\}$ represents the dichotomous response and x represents the vector of covariates, the logistic regression model assumes that $P(y = 1 \mid x) = \text{expit}(\beta^T x)$, where the function $\text{expit}(t) = 1 / (1 + \exp(-t))$. Its inverse is the function $\text{logit}(p) = \log(p / (1 - p))$ that involves odds. Indeed, if p is the probability of an event, the quantity $p / (1-p)$ is called the odds of the event. For example, the probability of getting 6 on a die is 1/6, and the odds are 1/5. This means that for each time a 6 is rolled, it is expected that there will be approximately 5 times when a 6 is not returned. Figure 6 shows the observed proportions of confirmed cases requiring ICU admission by age and the logit of these same proportions. Observe that the proportion of infected who require intensive care decreases until approximately 30 years of age, then increases until around 70 years and finally appears to decrease for women, while for men it seems to continue to increase, although perhaps more slowly. These trends are most clearly seen on the logit scale. These observations lead us to include two splines in the logistic model to explain the effect of being over 30 years old and being over 70 years old and possible interactions between sex and age. The inclusion of a spline to obtain a log-linear relationship between age and the probability of developing a severe form of the disease was already considered by Williamson et al.(Williamson et al. 2020). Likewise, since there are very few infected over 90 years of age, we decided to restrict our analysis to confirmed cases under 90 years of age. Using several variable selection techniques (best subset selection, forward, backward, stepwise, and lasso) and cross-validation with various fit evaluation measures (AIC, ECM, AUC, deviance), we chose the procedure to fit the model and identify the variables to include. In Table 2 we report the significant coefficients and their confidence intervals with a global level of 0.05.

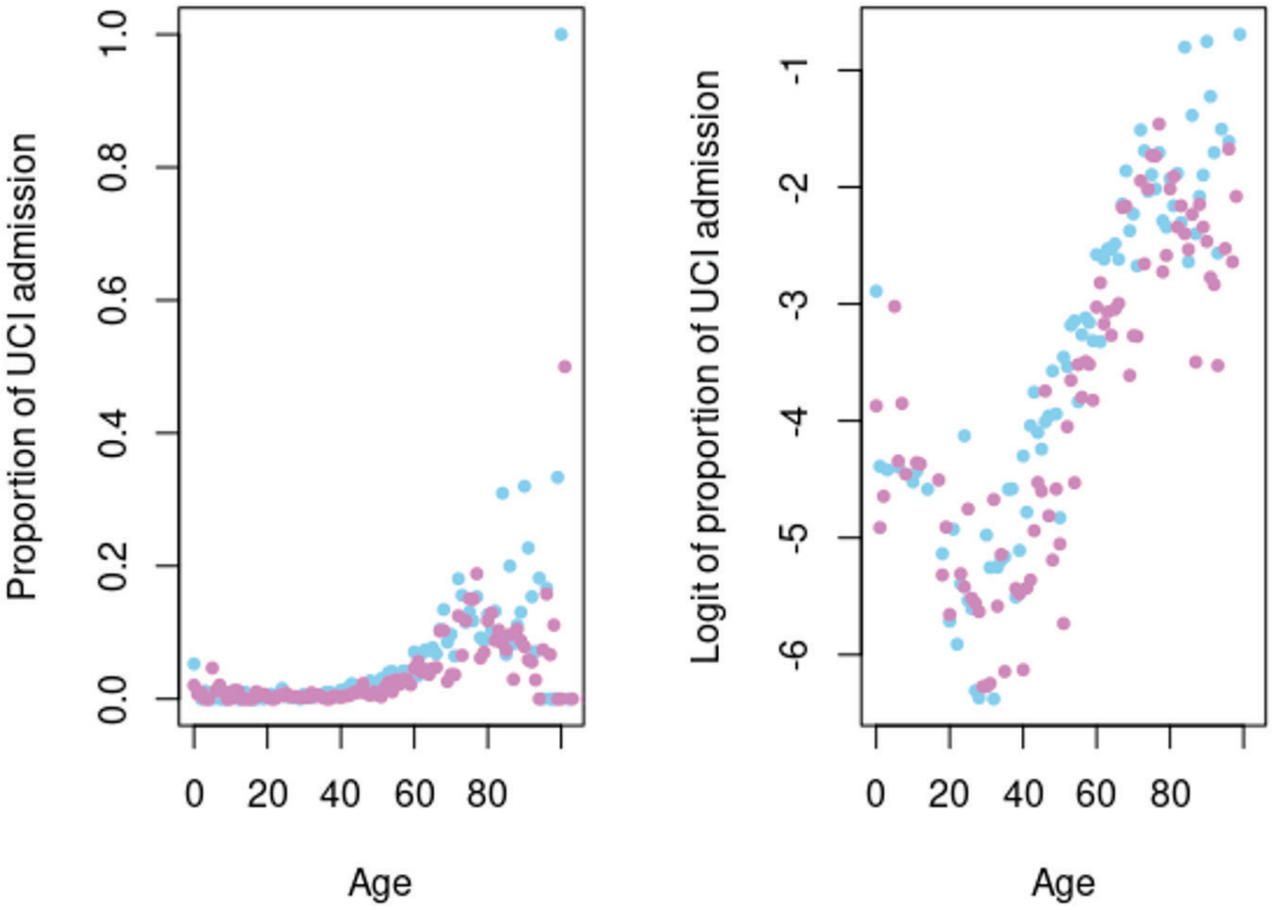


Figure 6. Proportion (left panel) and proportion in logit scale (right panel) of confirmed cases admitted to the ICU by age and sex (in light blue male and pink female).

Table 2. Odds ratios estimates and their confidence intervals of level 0.95.

Risk factor	Estimated OR	2.5%	97.5%	Risk Factor	Estimated OR	2.5%	97.5%
Diabetes	1.67	1.12	2.48	Heart Failure	1.75	1.06	2.91
Pregnancy	7.31	1.70	31.38	Obesity	3.48	2.22	5.44
Previous Oncological Disease	2.65	1.47	4.78	Over 30	1.14	1.07	1.21
COPD	1.99	1.15	3.45	Over 70	0.95	0.91	0.99
Hypertension	1.72	1.23	2.41	Age	0.95	0.90	0.99
Imm. Congenital or Acquired	3.32	1.43	7.71				

Results:

The odds of needing ICU admission among individuals with a risk factor, say obesity, is the ratio between the probability that someone infected with the risk factor needs ICU hospitalization and the probability that they do not need it. In turn, the odds ratio (OR) is defined as the ratio of the odds of needing ICU for those infected who have the risk factor of interest and the odds for those who do not, keeping all other explanatory variables constant. For example, according to Table 2, the chance that an obese infected person needs ICU admission is estimated to be approximately three times the chance for a non-obese person of the same age, of the same sex and with all other risk factors equal. The confidence intervals tell us what is the precision of the chances that we have estimated. For example, we can say that the chance of needing intensive therapy for the obese is estimated to be between 2.22 and 5.44 times that of the non-obese, while the chance of needing intensive therapy for diabetics is estimated to be between 1.12 and 2.48 times that of non-diabetics. With the fitted model it is possible to estimate the probability that a person diagnosed with COVID-19 needs ICU admission, knowing their age, sex and comorbidities.

Conclusions:

The statistical analysis presented allows us to identify risk factors that are important for developing severe forms of COVID-19 that require ICU admission. From this study, the following emerge as relevant risk factors: diabetes, pregnancy, previous oncological disease, COPD, hypertension, congenital or acquired immunosuppression, heart failure, obesity, age and its two associated splines that indicate being over 30 years old and over 70 years old. This set includes most of the comorbidities detected as important for predicting severity in Section 1.3.1.

1.4. Data-based analysis and control panel

Background:

The management of the pandemic is nourished by multiple sources of quantitative information that are constantly updated. In an emergency context, the structures of the data sources are subject to changes, errors, delays, etc. At the same time, the amount of available information sources tends to grow over time, but the quality of the data is not usually homogeneous. For these reasons, there is a need for a centralized resource that applies consistent, repeatable, and comparable methods of ingesting, cleansing, curating, persisting, and visualizing data.

Objective:

The objective is to produce a reliable and unified repository that offers information on the temporal evolution of metrics on COVID-19 in Argentina. On this platform, additionally, various data analyses can be produced with different objectives, creating new more complex metrics.

Methodology:

Every day the data that is updated by the Ministry of Health of the Nation contains new information and also updates on information from the past, due to multiple reasons such as reporting delays or updating the health status of patients who were registered for the first time in the system a while back. Also, previous information that has no updates is repeated every day, generating a great deal of redundancy. In this sense, our system preserves a complete copy of each update, for statistical purposes on the evolution of the data entry process itself. This means that we currently have more than 170 million records. However, when consolidating a database for monitoring the epidemic, we only keep operational changes in the epidemiological condition of each citizen.

The analysis of relevant indicators of the epidemic requires different territorial groupings depending on the official, civil servant or data analyst who consumes the information. For example, from a health point of view, in the Province of Buenos Aires it is important to analyze cases according to the cordons of the Buenos Aires suburbs, and also by health regions (“Regiones Sanitarias” n.d.).

Results:

The information presented on the site is built from multiple data sources, including: Open data from the Ministry of Health of the Nation (“COVID-19. Casos Registrados En La República Argentina” n.d.), COVID-19 Situation Room (Ministry of Health of the Nation) (“Información Epidemiológica” 2020) and Open data on the epidemiological situation of COVID-19 from the Government of the City of Buenos Aires (“Casos COVID-19” n.d.), among others.

At the time of writing this article, the system has complete data sets updated daily since July 15, containing epidemiological information accounting for the symptom onset as early as from January 1, 2020.

The visible face of the system is the web page, accessible at (“COVID-19 Análisis Y Control Basado En Datos” n.d.). The system architecture is designed so that any data analysis group can create new content that can be viewed on this site, taking advantage of the data infrastructure developed.

There we provide the ability to define territorial units that we call “Conglomerates” that consist of relevant groupings of localities. This can be seen in the tab Province of Buenos Aires and AMBA (city of Buenos Aires plus its suburbs) -> Sanitary Regions and also -> suburban edges (*Cordones conurbano*s in spanish). At the moment the system presents information on 550 departments nationwide and has defined 26 conglomerates of interest.

Temporal evolutions are shown in the number of daily cases and the number of accumulated cases (both by symptom onset date) and daily and accumulated deaths (by death date). Proportions of geographic concentration of confirmed cases are also shown. The information can be filtered by geographical area and consulted on a linear or logarithmic scale, in total figures or per 100,000 inhabitants and in daily values or moving averages every 7 days. The graphics are interactive, in such a way that the user can modify relevant aspects of each visualization focusing on what interests him the most.

To date, the system has been consulted by more than 850 different users with more than 3,800 visits coming 95% from Argentina and also including visits from countries in South America, North America and Europe. Likewise, the visualizations updated daily serve as input to compile reports that are provided on request and also periodically to different public administrations, such as the Ministry of Health of the Province of Buenos Aires.

The visualizations (updated daily) serve as input for compiling reports that are provided upon request as well as periodically to different public administrations, such as the Ministry of Health of the Province of Buenos Aires (see Section 1.1, Figure 1) along with an analysis of delay in reporting of deaths (see below).

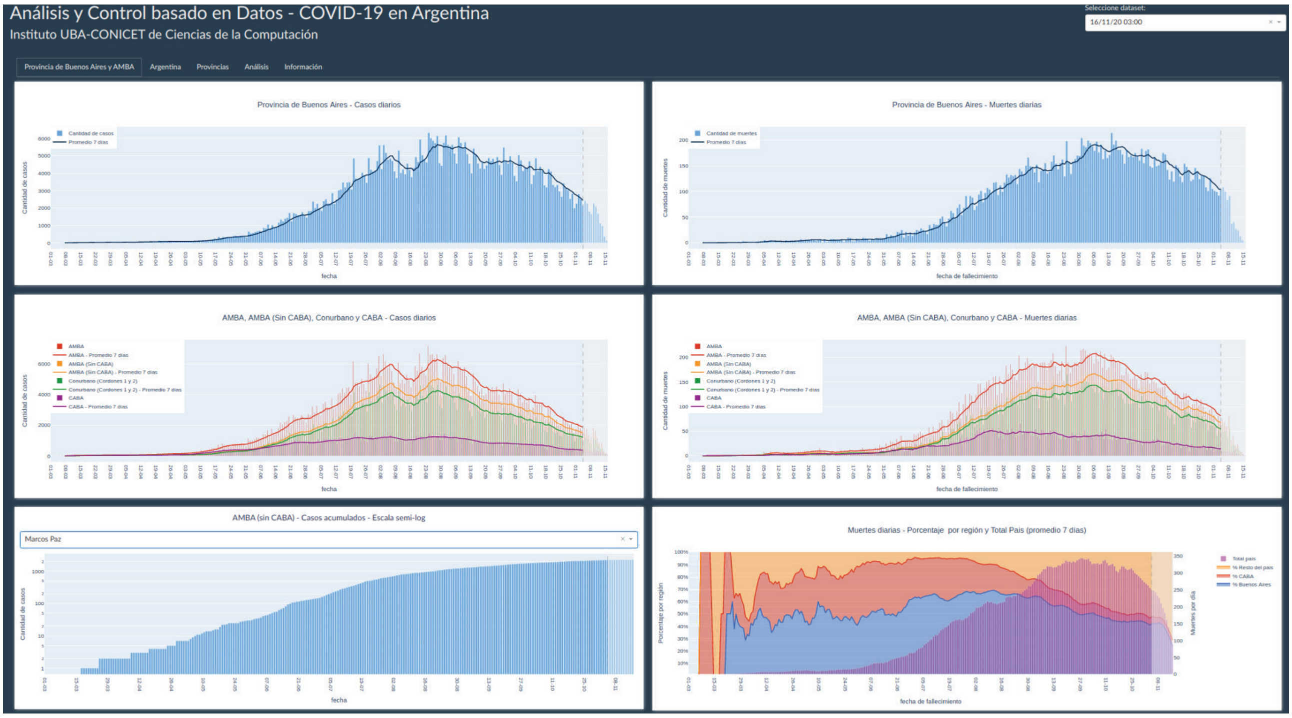


Figure 7. Screenshot of the site showing daily confirmed cases and daily deaths for the Province of Buenos Aires, AMBA, AMBA not including CABA, Cordones 1 and 2 of the Greater Buenos Aires and CABA. It includes an analysis of the proportion of deaths produced in CABA, in the Province of Buenos Aires and in the rest of the country.

It can be noted that towards the end of the graphs there is a vertical dotted line that separates the last 12 days of data from the rest of the information, and that the moving averages end there. This is due to the fact that the site uses the Date of Symptom Onset (acronym in spanish: FIS) as the date base, instead of the Report Date for cases. This is the case since the FIS provides a more stable and reliable indicator of the real dynamics of the epidemic, as it is less subject to the variability of the bureaucratic data entry process. As a counterpart, using FIS has the negative aspect that there is a delay, estimated at 12 days, until all cases that share the same FIS are reported.

Figure 8. Detail of the evolution of daily confirmed cases for Municipalities of the Province of Buenos Aires.

Finally, we show an example of data analysis that is possible exclusively if a history of all daily data updates is kept and an analysis of changes of states is applied for each individual over time (that is, it is not enough to access to "the last snapshot" of the public access databases).

Figure 9 shows the evolution of the death reporting process in CABA and PBA, focusing on the delay produced in the reporting of deaths that occur every day. The colors denote the difference between the date of death and the date of upload, for all cases that were entered into the system, day by day. This information can be found in the site's analysis section.

The comparison is relevant for highlighting the different strategies that jurisdictions use to deal with a known problem that is present in the entire data entry system, as is the case in the delay in the reporting of deaths when health administrative systems are overwhelmed.

It arises from the analysis that the Province of Buenos Aires opted for an immediate, high intensity transparency and remediation, while the Autonomous City of Buenos Aires opted for a slow and dosed remediation strategy.

Figure 9. Reporting of deaths (Screenshot of site). Left: Autonomous City of Buenos Aires, Right: Province of Buenos Aires, Above: Number of deaths reported per day, Center: Percentages (normalized to 100%) of delays suffered by deaths reported each day, Below: Cumulative evolution of deaths reported.

Conclusions:

This project has shown very satisfactory results at the moment, since it has proven to be useful to add relevant information and instantaneously to numerous debates and discussions about the evolution of COVID-19 in our country, both in academic environments and with political officials. and with public opinion. The site time series are used as input to adjust simulation models used to project possible future scenarios and suggest strategies for managing the epidemic. In this sense, having a unified and reliable panel of data such as the one presented in this section has been successfully covering a key need for advising decision-makers in public policies through mathematical-computational methods.

1.5. Estimation of prevalence at the national and district levels

Background:

We want to estimate the real percentage of infected population in each district using as input data the deaths discriminated by age range and by sex, and the lethality for each of these ranges. This also allows inferring the relationship between infected and confirmed cases.

Objective:

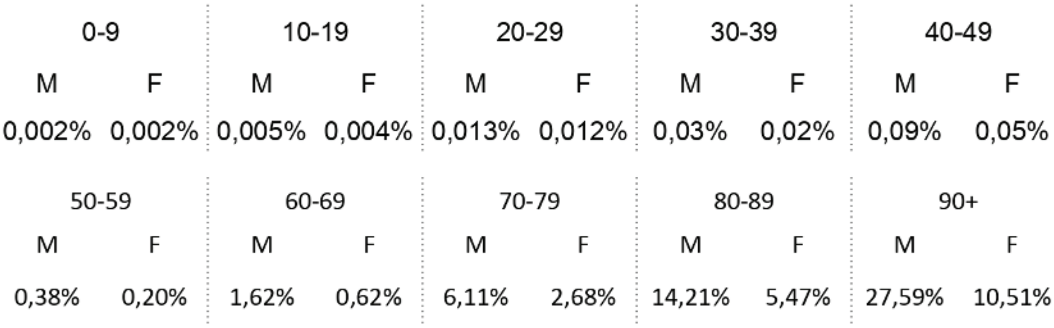
Knowing the percentage of infected population in each district is of interest when making public policy decisions, such as testing policies, tracking, restrictions on mobility, etc.

We will estimate these values for the country as a whole, for the 24 provinces of Argentina and for the 135 municipalities of the Province of Buenos Aires (PBA). The calculation for each municipality of PBA is particularly important given our close collaboration with the provincial government and the different municipalities of the Province.

Methodology:

This analysis was based on a seroprevalence study in Spain where the IFR (infection fatality rate) was obtained by age decile and by sex (Pastor-Barriuso et al. 2020). The choice of using the Spanish prevalence comes from the fact that it is the study with the highest level of detail that has been reported in the literature; however, other studies in different parts of the world show similar values. The summary of the infection fatality rate for each age decile and sex can be seen in Table 3.

Table 3. IFR by age decile and sex in Spain, estimated in (Pastor-Barriuso et al. 2020).



0-9		10-19		20-29		30-39		40-49	
M	F	M	F	M	F	M	F	M	F
0,002%	0,002%	0,005%	0,004%	0,013%	0,012%	0,03%	0,02%	0,09%	0,05%

50-59		60-69		70-79		80-89		90+	
M	F	M	F	M	F	M	F	M	F
0,38%	0,20%	1,62%	0,62%	6,11%	2,68%	14,21%	5,47%	27,59%	10,51%

We first estimate what we call "theoretical IFR" that comes from considering the Spanish IFR according to Table 3 and the age and sex distribution of the population of Argentina and its different jurisdictions. With these data we can obtain this theoretical IFR, which assumes that infections in all age ranges and in both sexes are homogeneously distributed. Once this theoretical IFR is obtained, plus the population of the district under consideration and the number of deaths to date, the percentage of the population actually infected can be estimated 2 weeks ago (we assume an average of 2 weeks between the reporting of a case and the date of possible death). As we know the reported number of confirmed cases, then we can estimate the coefficient of relationship between real and confirmed cases, which we will call the Quiros Factor (QF) (Figar et al. 2020) (Note that the QF is equal to 1 / IDR, where the IDR is the Infection to Detection Ratio, that is, the proportion of the unit of detected cases, defined in the literature). Given this QF and the confirmed cases in the last 2 weeks, we can estimate the percentage of infected population to date.

Next, we perform a fit on these calculations, using as input data the actual deaths, with their age ranges and sex, for each jurisdiction, in order to estimate what we call the current IFR.

Using these data on deaths as a starting point, plus the Spanish prevalence by age range and sex, we can estimate the number of real infected in each age range and sex, in each district, two weeks behind the moment in which we are making the estimate, which gives us the percentage of the infected population. With this percentage we can deduce the current IFR, and if we also use the number of confirmed cases, we can also estimate the adjusted QF (by age and sex of the deceased). Assuming that the QF is maintained for the last 2 weeks, we can now estimate the real percentage of infected to date, using again the data of confirmed cases in the last 2 weeks.

An additional observation: given that the number of deaths in the 0-19 age group is very low and that the prevalence taken in Spain for these groups was also calculated with very few cases, we have not used the prevalences of the 0- 9 and 10-19 deciles in the estimation of the current IFR, and to estimate the number of real cases in both groups we have taken the percentages of infected from the 60-69 group for the 10-19 group and from 70-79 for the 0-9 group (for each district), assuming that a parallel can be established between the level of care of those respective groups and therefore that the percentage of infections has been relatively even.

Results:

Table 4 shows the situation in Argentina as of 10/30/20. The data are displayed for the country as a whole, the city of Buenos Aires, the suburbs of Buenos Aires, the rest of the Province of Buenos Aires (*Interior*, in spanish) and the rest of the country. There it can be seen that by that date, taking the estimate derived from the calculation that uses ages and sexes of actual deaths, Argentina already had approximately 24% of its population infected at the national level, a value that far exceeds the percentages of infected in the countries of Europe in the first wave of March / May. In the city of Buenos Aires and the suburbs of Buenos Aires, this value rose to approximately 37% (also strongly exceeding the level of contagion in other large urban conglomerates in the world), while in the rest of the Province of Buenos Aires and in the rest of the country that value fell to values that are around 15%.

Table 4. Situation of Argentina in terms of prevalence and IFR in the main districts of the country by the end of October 2020.

Districts	**Population (2020)**	**Confirmed Cases**	**Deaths**	**Cases incidency**	**Estimated infected population (using theoretical IFR)**	**Theoretical IFR (homogeneous)**	**Quiros Factor (homogeneous)**	**Estimated infected population (using theoretical IFR)**	**Estimated infected population (with age and sex of deaths)**	**IFR (current)**	**Quiros Factor (current)**	**Estimated infected population (with age and sex of deaths)**
				16/10	16/10			30/10	16/10			30/10
Argentine	45,376,763	1,157,174	30,792	2.19%	11.16%	0.61%	5.10	13.02%	20.51%	0.33%	9.38	23.92%
Buenos Aires City	3,075,646	146,758	4,907	4.57%	17.68%	0.90%	3.87	18.47%	34.76%	0.46%	7.61	36.31%
Buenos Aires suburbs	13,630,369	458,346	14,993	3.16%	19.55%	0.56%	6.18	20.79%	35.11%	0.31%	11.11	37.34%
Province of Bs. As. (Interior)	3,910,772	86,033	2,324	1.75%	7.28%	0.82%	4.16	9.16%	12.03%	0.49%	6.88	15.13%
Rest of the country	24,759,976	463,408	8,493	1.41%	6.14%	0.56%	4.34	8.13%	11.81%	0.29%	8.35	15.62%

As can be seen in Table 4, the current IFR values are higher than the theoretical IFRs, due to the fact that infections have not been homogeneous by age range and sex. What is perceived is something to be expected; older people have taken better care of themselves and consequently in proportional terms they have been less infected, lowering the IFR values in practically all districts (for example at the national level the theoretical IFR is 0.61, whereas the current is 0.33).

As for the real QF at the national level, this number is around 9, in line with what has been seen in other parts of the world. This means that one case would be confirmed for every 9 infections.

Conclusions:

By the end of October, the number of real infected (deduced from the lethality by age range and sex, plus the reported death data) both in the city of Buenos Aires and in the Buenos Aires suburbs was already very high. This group accounts for 40% of the total population (which has implied a high number of deaths per 100,000 inhabitants, placing the country as a whole in the top-10 in the world by the end of October). This level of contagion in Buenos Aires and its surroundings resulted in a sharp drop in the weekly number of infected, mainly due to the sharp drop in the number of susceptible people, as well as the maintenance of personal care (and despite a sharp rise in the mobility). In the rest of the country, the situation still presents comparably lower values of infected (around 15% if we look at both the *interior* of the Province of Buenos Aires and the rest of the country), which forces us to take extreme measures of personal care, restriction of mobility and permanent tracing of close contacts, to keep the contagion curve (and consequently of deaths) at low levels.

1.6. Immunity: from biological to social causes

Background:

Some speculation has been raised that a certain degree of herd immunity is the main cause of the decline in infections in Europe after the initial outbreak. There were claims that between 10% and 20% of infections are enough to stop growth dynamics, instead of the 60-70% indicated by simple calculations (this value is estimated at 100 * (R0 - 1) / R0). Among the speculations about herd immunity, several possible causes are mentioned: some previous immunity in the entire population due to cross-reaction with other coronaviruses, highly inhomogeneous contagion patterns, among others. However, most of these speculations fell through when the real numbers of acquired immunity were obtained: less than 10% in most cities. The parameter known as the IFR (infection-fatality ratio), which can be estimated with a sufficient degree of precision from the death count as seen in the previous subsection, even in the absence of large seroprevalence studies, also shows low prevalence values in all European countries. Taking these facts into account, the question arises as to the main cause of the rapid decline in the infection rate in European countries.

Objective:

We have investigated an alternative explanation, in which the focus of the decline in cases shifts from biological to social causes, from herd immunity to "fright" immunity.

Results:

The term fright immunity reflects the awareness of individuals after learning about the consequences of COVID-19, including deaths and serious sequelae. When the "fright" occurs, the majority isolate themselves and begin to behave in a much more cautious manner. The effective reproduction number $R_{ef}(t)$ for the entire subpopulation of "frightful" is much less than one. However, a small fraction of individuals persist in dangerous behavior (whom we call "reckless" -mostly young- people who are not afraid and are less likely to self-isolate), which although somewhat less than before the pandemic still implies a reproduction number $R_{ef}(t)$ greater than unity. Many of these individuals shared social and physical proximity, and the virus found a fairly simple way to spread. It is important to note that these fractions are not defined institutionally, nor are they necessarily determined by their employment status. Many exempt essential workers behaved according to very strict rules and protocols of care, and vice versa. Daredevils being a smaller and much more mobile fraction of the group, achieve (as a subgroup) true herd immunity in the months after the high pandemic peak. However, the frightened majority never reach high infection numbers, although a small fraction of them were infected by the mobile population. This small fraction of a large total implies that the total number of quarantined infected is similar to that of reckless, showing one of the remarkable characteristics of the COVID-19 pandemic: some transmit, others suffer. This is accentuated due to the higher fatality among older adults, with generally less risky behavior.

It is possible to establish a simple model based on a two-population SEIR, where the beta transmission parameter is broken down into four values $beta_{11}$, $beta_{12}$, $beta_{21}$, $beta_{22}$, corresponding to transmission between populations 1 (frightened and quarantined) and 2 (reckless and mobile). Figure 10 shows a schematic model for Spain in which both populations are described as homogeneous. From the IFR, deaths and seroprevalence, actual infection figures can be derived. The fraction of reckless individuals (f_2) is a parameter adjusted to the actual data as well as the reproduction numbers $R_1 = beta_{11}/gamma$, $R_2 = beta_{22}/gamma$ and the cross transmission parameter $beta_{21}$.

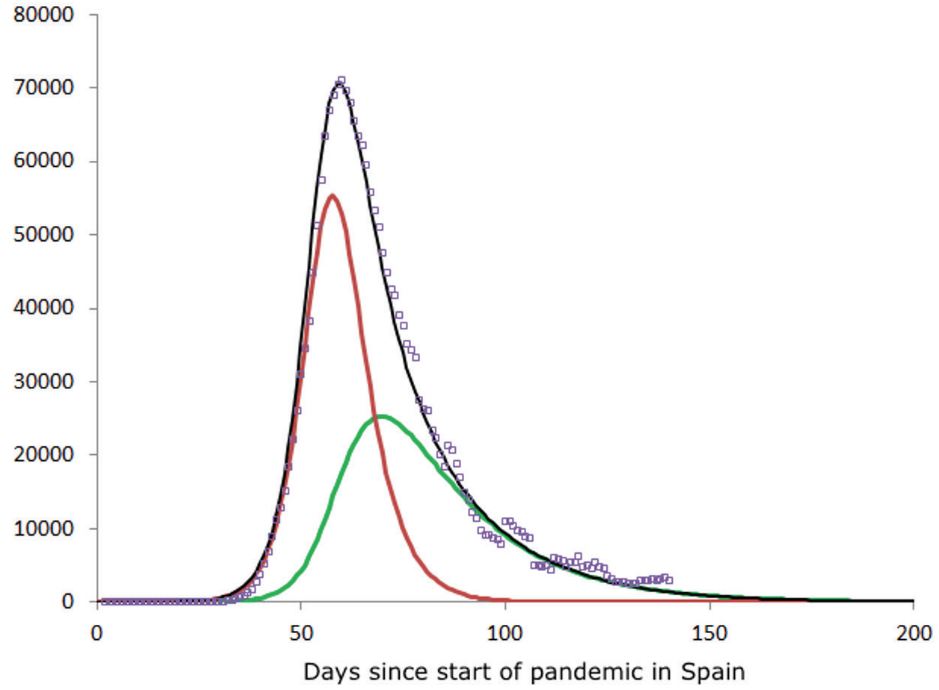


Figure 10. Fright immunity pattern for Spain. Red line: Daily infections of the mobile population. Green line: Daily infections of the quarantined population. Black line: sum of infections. Squares: real data, corrected for seroprevalence factors and IFR. Population= $47x10^6$, R_1=0.67, R_2=3,57, f_2=0,0253, $beta_{21}$=$8,3x10^{-3}$ $beta_{12}$= 0

This type of behavior should be considered as a cartoon, to the extent that there is undoubtedly a whole gradient of behaviors associated with more or less the extreme positions of the model. When the fitting procedure is applied to various countries or cities in Europe (United Kingdom, Germany, Belgium, France, Madrid, London) it can be seen that in all cases the fitting yields two well differentiated values of R corresponding to a behavioral situation. On the contrary, when the data of the Autonomous City of Buenos Aires is treated with this same model, the adjusted Rs are 1.6 and 1.13 respectively, both greater than 1 and close to each other, which shows that there aren't two clearly differentiated populations as in European countries. This low level of "fright" is very compatible with the success in the objective sought to "flatten the curve" of infections. The gradual arrival of cases and deaths, avoiding collapse of the health system significantly reduced the perception of danger, and encouraged a behavior closer to the norm, only partially inhibited by restrictions on circulation and certain activities. The analysis must be differentiated with respect to popular/urban slum neighborhoods, where barriers to transmission are minimal and R values are much higher, than the expected level for the natural epidemic curve of COVID. The strategies used to mitigate this situation were ineffective, reaching a cumulative prevalence of more than 50% in a few weeks (eg Barrio Carlos Mugica) (Figar et al. 2020). A very different situation occurs in the traditional neighborhoods of the City. As seen in the previous chapter, at the end of October the accumulated prevalence of CABA can already be estimated to be above 35%. However, the fraction of infected is very different for the various age ranges.

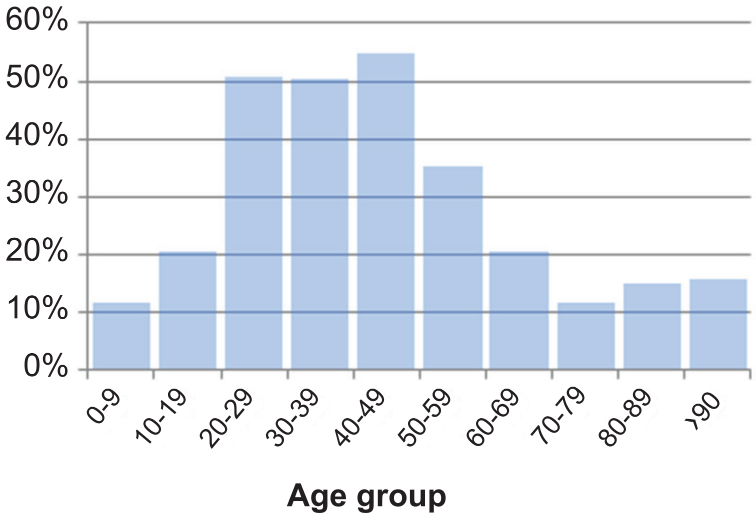


Figure 11. Estimated prevalence (as of October 16, 2020) for the different age ranges in the City of Buenos Aires.

Figure 11 shows that younger individuals have behaved more recklessly and reached important prevalence levels, although much more slowly than the reckless fraction in Spain and other European countries, whereas older adults have had much less contact with the virus, having understood the associated risk, particularly high at these ages. This low prevalence of the elderly further indicates that for the high number of deaths in the city of Buenos Aires and its suburbs, the number of real cases among young individuals has been several times higher than those that occurred in Spain and other European countries.

Conclusions:

A mathematical model is exhibited that could explain the reason for the decline in the contagion curve in European countries as of May, despite the low prevalence number. The causes of this behavior move from biological to social reasons. While in Europe there is a sharp peak with saturation of the health system, followed by an abrupt drop in cases as a result of the measures taken and the awareness of the population, the epidemic curve in the city of Buenos Aires and its suburbs is completely different; there is no generalized fear despite the high number of cases and deaths, since these appear gradually. While young people present a prevalence close to "herd" conditions, a sign of a relatively more mobile behavior during the isolation imposed by the government, which helps to lower the number of daily cases (surely added to the seasonal factor that may have accentuated the descent), the elderly have remained with little contact. The low prevalence among the age ranges most prone to severe COVID-19 represents an additional risk if transmission is not stopped at very low levels before weather conditions become more favorable for the spread of the virus.

2. Rapid response projects

2.1. Hospitals in networks: the case of Del Cruce Hospital in Florencio Varela, Province of Buenos Aires

Background:

Because of the advance of the pandemic, hospitals in the southeastern region of Greater Buenos Aires, which includes the municipalities of Almirante Brown, Quilmes, Florencio Varela and Berazategui, began to be restructured to work as a single functional unit (Figure 12). One of the objectives of forming this network was to organize and thus optimize the occupation of hospital beds in the context of the looming pandemic. The network has 7 hospitals, the central hospital of which is the Hospital del Cruce de Florencio Varela. During the pandemic, 3 Modular hospitals and 4 Early Care Units (ECU) were also built, some of which operate within the same hospitals.

Figure 12. Location of the hospitals in the southeastern region of Greater Buenos Aires, whose central hospital is the Del Cruce Hospital in Florencio Varela. The modular hospitals and ECUs that were built during the development of this project are included.

Objective:

Our aim was to build a software to organize patients and hospital beds for this hospital's network which could be used by any other hospital network in the country. We also aim that the software offer the capacity to collect relevant information for future analysis.

Methodology:

Due to the urgency to start using the tool, we performed a first version (minimum viable product) in a few days, and then the development was expanded as more requirements were defined. Figure 13 shows the main screen of the system. The image corresponds to a panel with fictitious data.

Although the system responded to the requirements of this hospital network, it was developed so that any hospital network can use it ("Software as a service"). The system is located in a place accessible from any computer. Everything is customizable and self-manageable.

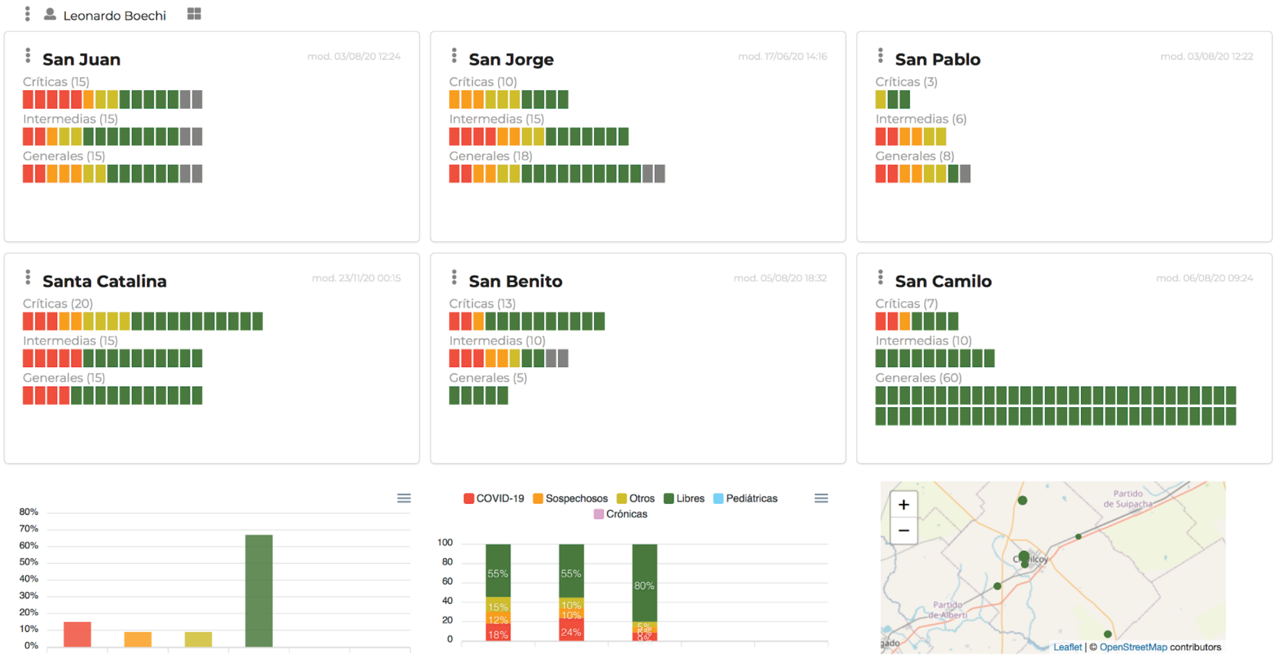


Figure 13. Main Panel of Hospital Beds. The hospitals in the network (7 hospitals plus 7 Modular / ECU). The hospital beds organized in Types are observed: General, Intermediate or Critical. The color of each bed corresponds to its Status: COVID-19, Suspect, Others, Free, Disabled. Each circle on the map corresponds to a hospital / Modular / UPA in the network, the area and the color of the circle correspond to the number of beds in that state. The panel allows users to download data. The image corresponds to a panel with fictitious data.

The system also allows entering patient data, which can be seen in the Patient Panel (Figure 14). This panel is used to organize the patients and beds of each unit. It should be clarified that the system does not collect sensitive information from patients such as name, ID, address, telephone, etc.

Figure 14. Panel of Patients. The dial shows the level of social risk, the person has a different color according to the clinical risk, the number of days of hospitalization, the age of the patient, and whether he has been referred from another hospital (ambulance) is also included.

Conclusions:

The tool is widely used within the hospital network (Yacobitti et al., n.d.). It is making it possible to organize beds and patients within hospitals, as well as transfers between them. The project continues being improved to include new functionalities and requirements. The link to the system is available at ("Hospitales" n.d.).

2.2. COVID-19 Prevention Center at the Faculty of Exact and Natural Sciences (UBA)

Background:

One of the prevention programs developed by the Province of Buenos Aires is the identification and monitoring of close contacts (CC[1]) of confirmed or suspected cases of COVID-19 to ensure that they do not continue to spread, and at the same time that they can receive the assistance they need.

On July 22 of this year, a University Telemedicine Center was put into operation in the Faculty of Exact and Natural Sciences of the University of Buenos Aires, in order to carry out the identification and monitoring of CCs. This prevention center provides support to the Municipality of Vicente López and the Municipality of Tigre. This Center is part of a program of the Province that involves another 34 university centers with the same dynamics (all of them created within national public universities).

---

[1] A CC (as defined by the PBA) is considered to be any person who has remained within a distance of less than 2 meters with a confirmed or suspected case of COVID-19, while exhibiting symptoms or within 48 hours prior to the onset of symptoms, for a period of at least 15 minutes.

Objective:

End the chain of infections by ensuring that individuals who have tested positive for COVID-19 as well as their close contacts remain isolated in their homes. Early detection of potential infected cases. Assist people who are confined during the days of their confinement.

Methodology:

The Center of Exactas is made up of a group of more than 40 students and alumni of our Faculty. It operates from Monday to Saturday in 4 fixed shifts of 4 hours, each of which is coordinated by a person responsible for daily tasks. All work is registered on a digital platform developed by the PBA. Both the physical facilities and the individual behaviors comply with the sanitary measures established by a strict Hygiene and Safety protocol provided by the academic unit of the institution and approved by the Province.

The daily work methodology at the center consists of:

Identification:

i. Telephone communication with the suspected or confirmed COVID-19 person.
ii. Determination of the date of symptom onset.
iii. Survey of close contacts
iv. Entry of CC information into the PBA application.

Accompaniment:

i. Telephone communication with the CC.
ii. Verification of the date of last contact with the epidemiological link.
iii. Ask about the appearance of one or more symptoms compatible with COVID-19.
iv. Survey on the needs for food assistance and / or hygiene supplies.
v. Classification of the evolution of the follow-up according to: continuation in isolation, discharged from follow-up or becoming a suspicious case due to the presence of symptoms.

In addition to the routine questions, the calls provide a space for advice and containment to people during the strict isolation they have to face, as well as an opportunity to relieve needs in the provision of food and hygiene supplies, among others. This information is duly compiled and transmitted daily to those responsible for the municipalities; who in turn use it as a source of information in their internal management.

Since the start of operation of the CeTeU and for the period of 16 weeks, 31070 effective calls have been made in total, of which 6560 correspond to suspected or confirmed people to identify their CCs (an average of 71 people per day), obtaining 3931 people with the CC classification (an average of 42 people per day). This shows us that approximately every 10 suspects contacted by the center identifies 6 CCs. Although the data are not conclusive, a possible explanation for this relationship is that a good part of the suspected cases are discarded because their swabs turned out to be negative. On the other hand, a total of 20,579 follow-up calls were accumulated, with an average of 224 calls per day for this task. Figure 15 shows the total per week of the suspects contacted along with the total per week of identified CCs, whereas Figure 16 shows the total per week of follow-up calls made.

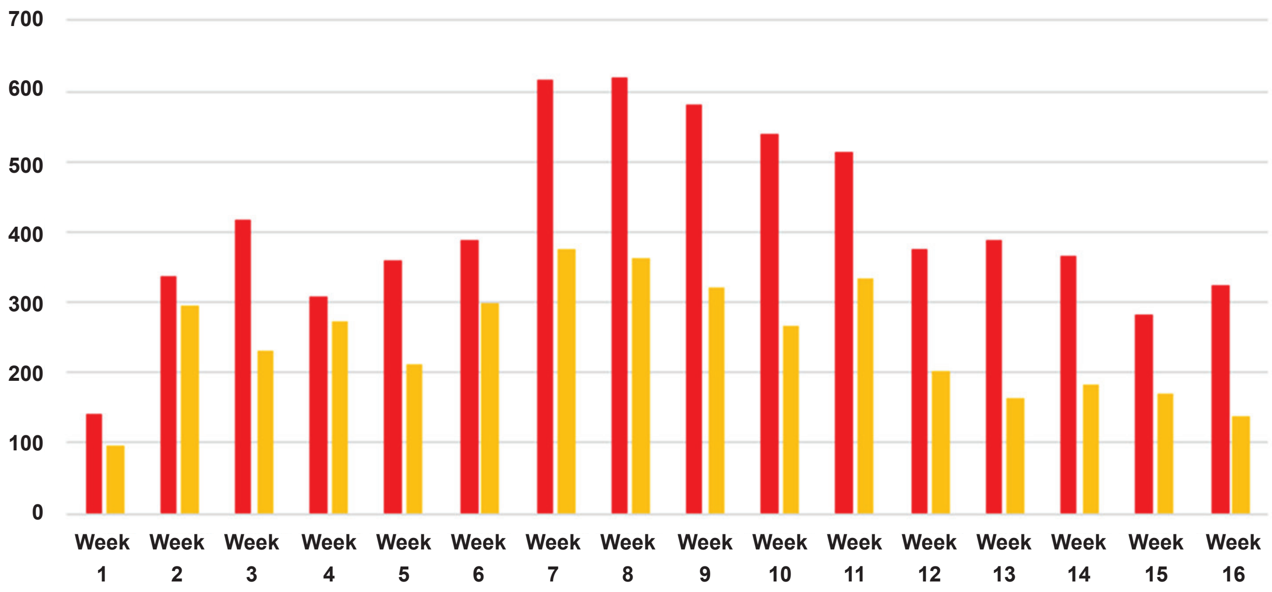


Figure 15. Number of suspected/confirmed (red) cases contacted and close contacts (orange) identified per week.

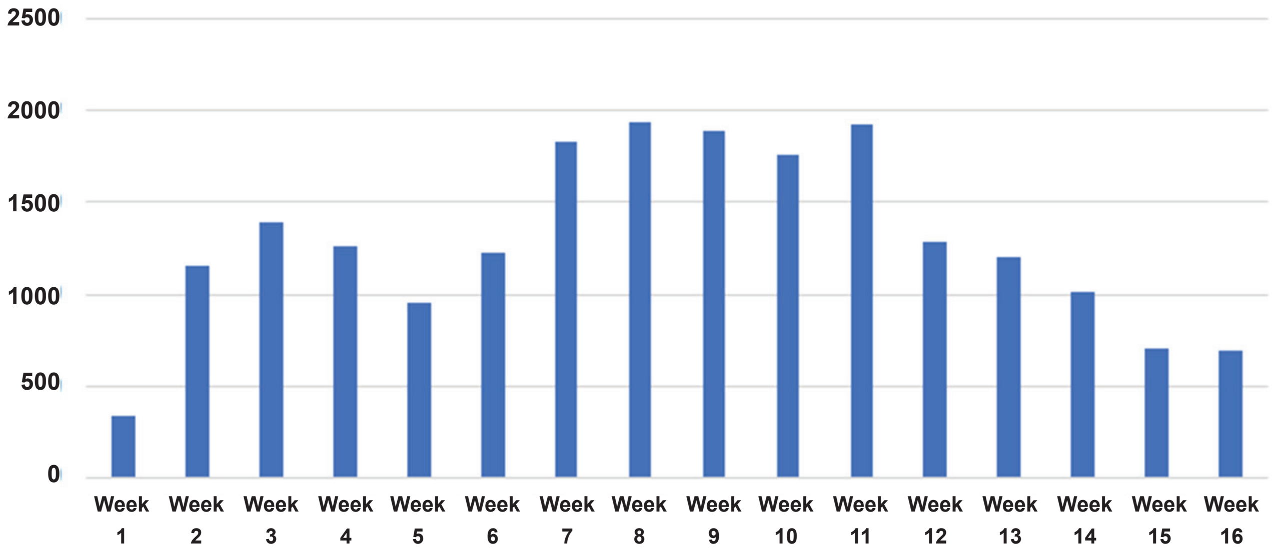


Figure 16. Number of follow-ups with close contacts per week.

Although it is still premature to measure the real impact of the work carried out by the Center, there are certain positive indicators. A first indication can be seen in Figure 17, the confirmed cases in Vicente López show a stabilization towards the end of July (when the center began its activities), while the accumulated cases of the other 12 municipalities in same Sanitary Region (administrative organization that uses PBA to group municipalities) continued to grow steadily.

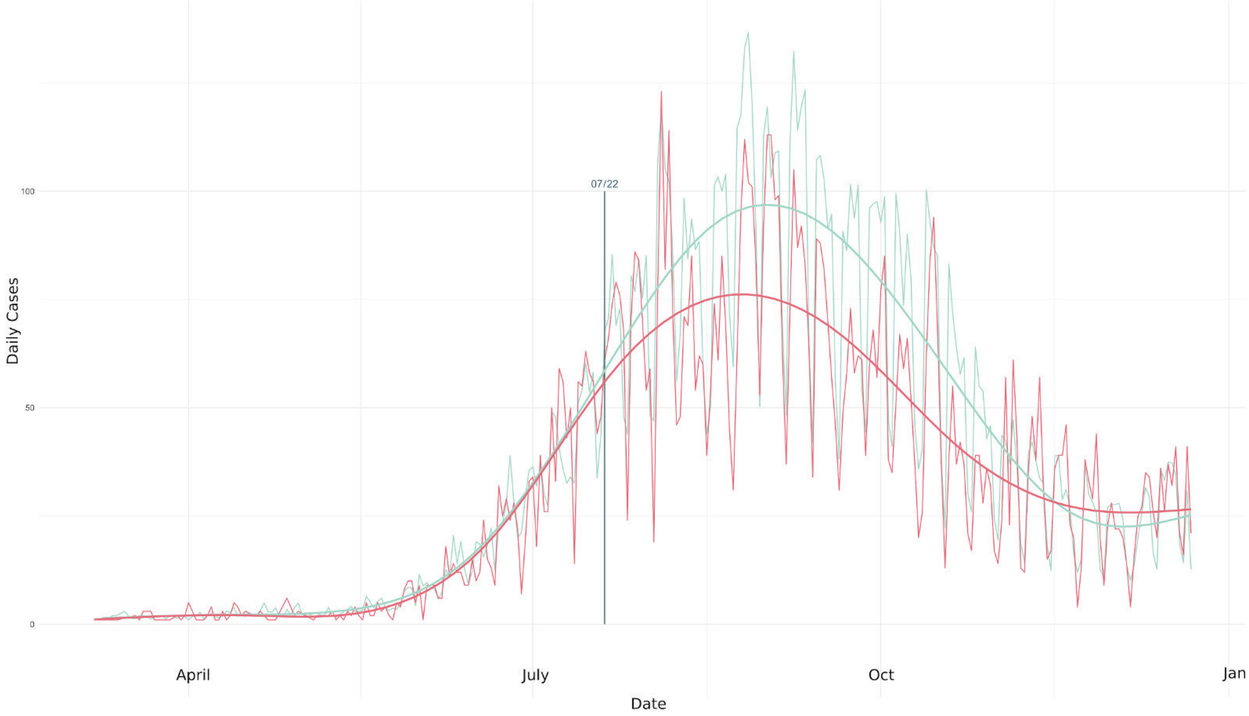
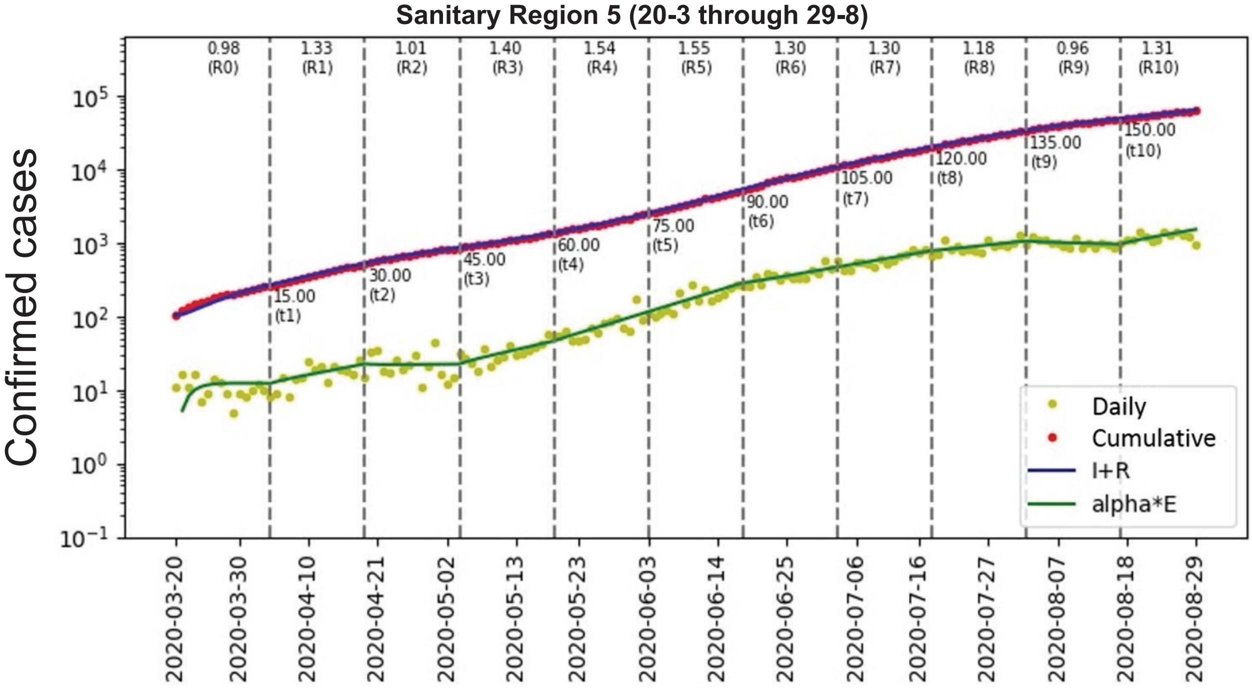


Figure 17. Evolution of daily confirmed cases over time for the municipality of Vicente López (in red) and for the group of other municipalities of Sanitary Region V (in green). Vertical line indicates start of contact tracing (Source: Antonio Milanese of Municipality of Vicente López).

On the other hand, the evolution of R also shows a greater decrease in Vicente López than in the group of parties in the same Sanitary Region, as can be seen in Figure 18.


Sanitary Region 5 (20-3 through 29-8)

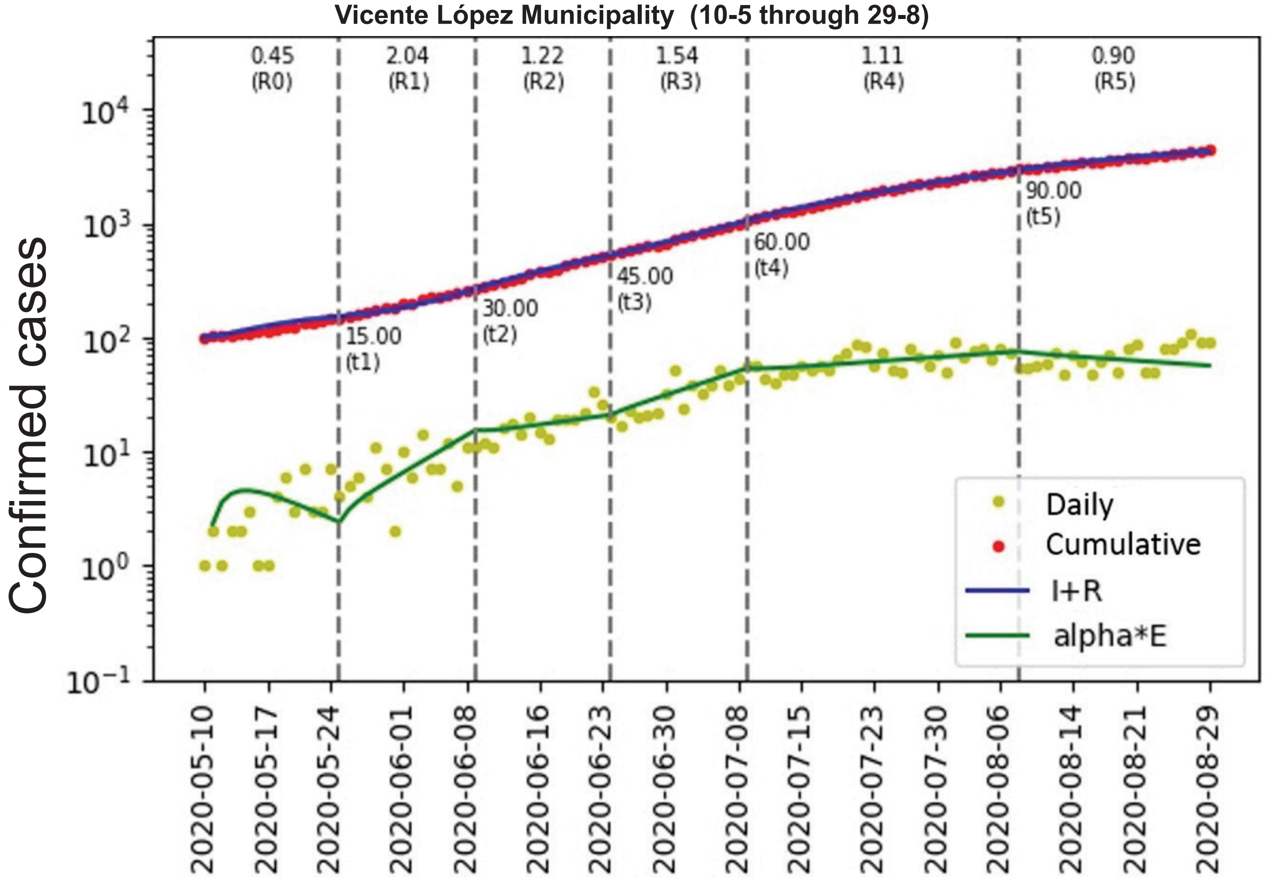

Vicente López Municipality (10-5 through 29-8)

Figure 18. Evolution of the R in Vicente López (bottom), and in the set of municipalities of Sanitary Region V to which Vicente López belongs (top). These indicators come from the parameter adjustment method based on simulation models described in Section 1.1 which is automatically calculated daily by the analysis, control and visualization system described in Section 1.4.

Other indications are of a qualitative nature, such as the strong awareness that is made through each call, either explaining the need for strict isolation of the CCs and the appropriate prevention measures that they must take. The social response to these calls is in the vast majority of cases positive and grateful. On the other hand, the Municipality of Vicente López has also expressed great appreciation for the work carried out in the center and the contribution of the information provided. In the case of Tigre, the beginning is very recent, but without a doubt, the participation of the Center allows the Municipality to identify and monitor a greater number of cases and, above all, earlier contact.

Conclusions:

In conclusion, it can be said that the experience of the University Telemedicine Center at the Faculty of Exact and Natural Sciences of the University of Buenos Aires is satisfactory and is generating an impact that may soon be even more quantifiable. Meanwhile, it is a useful tool in prevention tasks, being the first line of data acquisition used for management in the containment and mitigation of the disease.

2.3. Towards an agent-based simulation platform for the municipality of Marcos Paz, in the Province of Buenos Aires

Background:

The compartmental models described in Section 1.1 incur a level of abstraction that does not allow the study of detailed dynamics at the territorial level. The compartments, the reproductive number R, and the structure of the model itself are a macroscopic abstraction, representing the emergent behavior of networks of complex microscopic interactions between people of different ages who participate in very different ways in the epidemic process. These interaction networks are manifested in the territory according to the socioeconomic structure of each jurisdiction. At this level of detail, aspects such as the location of public buildings, shops and manufacturing plants, communication routes with other municipalities, location and capacity of hospitals, population density in residential areas and capacity dynamics towards central areas of greater activity, to name a few.

Analyzing social dynamics at this level is of particular interest to municipalities that must make decisions about regulations to manage the epidemic, which must contemplate the specific reality of the jurisdiction (a municipality supported by summer tourism activity, as in the case of Mar del Plata, will develop strategies different from a Mediterranean municipality focused on the food industry, as in the case of Marcos Paz).

Objective:

We therefore propose the development of a comprehensive platform based on simulation models that allows assisting decision-making with an explicit territorial level of detail, with a particular focus on the Municipality of Marcos Paz in the second cordon of the suburbs of the Province of Buenos Aires .

Methodology:

Spatially explicit simulation based on individual agents (Taillandier et al. 2019; Vynnycky and White 2010) allows to exercise hypothetical scenarios where interventions on reality take the form, for example, of restrictions on the number of vehicles that are allowed to cross the municipality from other jurisdictions, number of resident excepted personnel who are allowed to carry out working hours outside the municipality, quantity and size of public buildings with authorization for the capacity of citizens and their limitations on the number of people simultaneously sharing certain spaces, etc. The agents, in addition to forming family groups located according to the density of residence indicated by the censuses, run the epidemic through the susceptible, exposed, asymptomatic infected, symptomatic infected (possibly mild or severe), recovered, and deceased states.

Results:

We have developed a simulation model, together with its visualization infrastructure, for the Municipality of Marcos Paz ("Inicio - Municipio de Marcos Paz" n.d.), with the capacity to explicitly simulate the 65,000 inhabitants of the municipality and flows of travelers from / to other municipalities. Through close interaction with Marcos Paz's municipal administration and its Ministry of Health, we carried out incremental phases of fitting of the model to capture the main socioeconomic dynamics that describe the mobility of people and the evolution of the metrics of the epidemic.

An example of a "snapshot" of activity simulated using our spatially explicit agent model is shown in Figure 19.

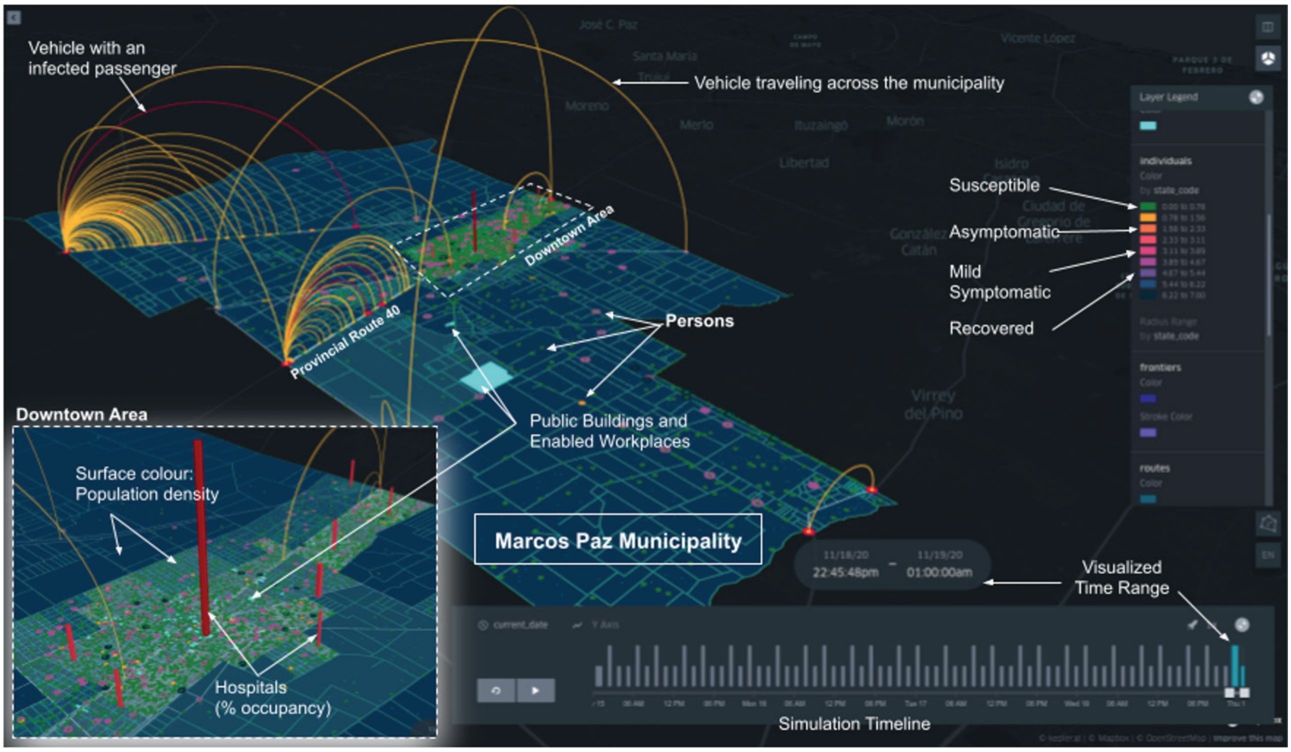


Figure 19. Agent-based simulation visualization for the Municipality of Marcos Paz in the province of Buenos Aires.

Conclusions:

This tool shows a great wealth for communication, interaction and discussion of possible results and scenarios with decision makers of public policies. Although it is in the initial stages of development, its structure opens the doors to provide a framework for future analysis of socio-economic problems that require explicit representation at the urban scale in the Municipality of Marcos Paz.

3. General conclusions and discussion

The pandemic has put the entire society on edge, mainly due to the rapid spread of the virus, influenced by the high mobility that exists due to commercial and tourist flows. That is why it took most nations by surprise, wreaking havoc on the global health and economic system.

The international scientific community reacted quickly by looking for palliative methods and vaccines, and developing models to be able to anticipate or at least keep up with the development of the pandemic. At the local level, the scientific community responded very quickly and proactively, generating solutions of all kinds, from detection systems, different types and mechanisms of tests, treatments and different proposals to contain contagions and to understand the behavior of the pandemic, to the development of the official application for prevention and public health care (*Cuidar* app ("Sistema Y Aplicación Cuidar" 2020)), in which scientists from our groups actively participated.

In this work we have presented the main contributions that a group of scientists from the Applied Math and Computer Science Institutes of the Faculty of Exact and Natural Sciences of the University of Buenos Aires and CONICET, together with colleagues from other institutions, in the fight against the pandemic in Argentina. Quantitative tools were developed that are helping to understand the state of the pandemic, as well as testing scenarios to evaluate different containment policies. Our work also included policies in the territory, such as the developments for hospitals and the tracking and contact identification centers (the latter an essential task but one that was not implemented on the scale that would have been necessary at the national level). We have also provided consulting and scientific-technological support to municipalities and provincial states, as well as organized a weekly seminar (29 to date) to highlight and debate these initiatives with experts of diverse disciplines ("Instituto de Cálculo" n.d.).

If we were to analyze Argentina's behavior in response to the pandemic, we could say that the early quarantine dictated by the national government on March 20 allowed a slow growth of the contagion curve (see Figure 1), which in turn gave rise to a strengthening of the health system. As a favorable result and unlike what happened in most European countries, the healthcare system was not saturated and potential deaths due to sanitary collapse were avoided. In contrast, this “flattening” of the curve, where the slow but permanent growth of infections is seen, led to the accumulated number of infected at the national level throughout the months being very high (see Section 1.5), similar to what has been observed in the number of deaths, turning Argentina into a country with a very high death toll per 100,000 inhabitants. These figures could have been reduced with a national strategy aimed at the suppression of the virus, based on the tracing of close contacts of confirmed and suspected cases, and the search for asymptomatic infected persons using intelligent testing techniques. Some of these actions had good local results (as can be seen, for example, in Section 2.2), but the strategy was not expanded nationally and in time to achieve a drastic decrease in infections.

We hope to be near the end of the pandemic, either because it is suppressed through efficient tracing and testing, because immunity is obtained through a vaccination campaign, because effective treatments are developed, or through a combination of these efforts; nevertheless, accurate information is essential to make the best decisions to control the spread of the virus. It is possible that these modeling techniques could also be useful for the organization of vaccination campaigns and drug delivery, as well as for dealing with other types of diseases with similar characteristics.

In summary, we consider that the contributions made by our working group have been very valuable in the fight against COVID-19 in Argentina. This is evidenced in the receptivity that the proposals had in different local and provincial governments, and in the strong impact that a good part of these developments had in the regional and national media.

Author contributions

The main participation of each of the co-authors of the publication in the different sections of the work is detailed below.

Editing and general compilation: Mehrnoosh Arrar, Guillermo Durán y Diego Garbervetsky.

Section 1.1: Rodrigo Castro, Esteban Lanzarotti y Ezequiel Pecker-Marcosig.

Section 1.2: Ana M. Bianco, Guillermo Solovey y Marina Valdora.

Section 1.3.1: Laouen Belloli, Leonardo Pepino, Juan E. Kamienkowski y Luciana Ferrer.

Section 1.3.2: Ana M. Bianco y Marina Valdora.

Section 1.4: Rodrigo Castro, Pablo Laciana, Esteban Lanzarotti, Ezequiel Pecker-Marcosig y Mariano Zapatero.

Section 1.5: Guillermo Durán, Roberto Etchenique, Mario Lozano, Rodrigo Maidana y Rodrigo Quiroga.

Section 1.6: Roberto Etchenique.

Section 2.1: Leonardo Boechi.

Section 2.2: Guillermo Durán, Natalia Fernández, Diego Garbervetsky, Rodrigo Goldsmit, Carolina Grillo Vidal, Mario Lozano, Mauricio Mendiluce y Sol Minoldo.

Section 2.3: Rodrigo Castro y Ezequiel Puerta.

Acknowledgements

We want to thank the governments of the Province of Buenos Aires (collaboration formalized through Institutional Cooperation Agreement between the FCEyN-UBA and the Ministry of Health of PBA), Marcos Paz (collaboration formalized through Institutional Cooperation Agreement between the FCEyN-UBA and the Municipality) and Vicente López, to the authorities of the Hospital del Cruce and its network of hospitals, to the authorities of the Faculty of Exact and Natural Sciences of the UBA and to the Huesped Foundation and Fundar, for their strong support in different stages of this work. A Lauriano Alimenti, Carlos Bianco, Leandro Cahn, Inés Caridi, Jorge Cassará, Sebastián Ceria, Sebastián Crespo, Ricardo Curutchet, Miguel Fernández, Julián Fernández Bonder, Juan Flo, Daniel Foguelman, Matías García Fallit, Marcelo Ferreyra, Mariano Goldman, Rafael Grimson, Daniela Hozbor, Matthieu Jonckehere, Silvia Kochen, Ernesto Kofman, Nicolás Kreplak, Mauro Lizama, Mar Lucas, Mónica Lucas, Daniel Maceira, Franco Märsico, Antonio Milanese, Gabriel Mindlin, Axel Molina, Nora Mores, Juan Pablo Pinasco, Ana Quaglino, Fernando Rago, Mauro Ramos, Juan Carlos Reboreda, Martín Reydó, Daniela Rodríguez, Mateo Sagardoy, Lucio Santi, Soledad

Santini, Fernando Schapachnik, Soledad Suárez Mazzone, Omar Sued, Pablo Terlisky, Gustavo Tieffenberg, Sebastián Uchitel, Federico Vázquez y Agustín Wydler, for their contributions to the projects presented here. We also especially thank Jorge Aliaga, Martín Barrionuevo, Daniel Feierstein, Axel Kicillof y Tomás Pueyo for the valuable discussions on data analysis and social issues related to COVID-19 in Argentina and the world, which helped us to understand more fully the problems of the situations posed by the pandemic. This initiative has been partially supported by the projects "Mathematical-computational tools for the intelligent control of insulation levels in each municipality of the suburbs of the Province of Buenos Aires" and "Projection of trends and evaluation of intervention scenarios for the COVID-19 epidemic in Argentina through Modeling and Computational Simulation", both financed by the Ministry of Science, Technology and Productive Innovation of the Argentine Republic, and by the project "Development of optimization tools, statistics and data science for the management, monitoring and evaluation of public policies", of the PIDAE program at the University of Buenos Aires. Partial funding for the preparation of this article was provided by the Instituto Sistemas Complejos de Ingeniería (ISCI) in Santiago, Chile (CONICYT PIA/BASAL AFB180003) and grant nos. UBACyT 20020170100495BA (UBA, Argentina).

References

[1] "Casos COVID-19." n.d. Accessed November 17, 2020. https://data.buenosaires.gob.ar/dataset/casos-covid-19.

[2] "COVID-19 Análisis Y Control Basado En Datos." n.d. Accessed November 17, 2020. http://sedcovid.exp.dc.uba.ar/.

[3] "COVID-19. Casos Registrados En La República Argentina." n.d. Accessed November 17, 2020. http://datos.salud.gob.ar/dataset/covid-19-casos-registrados-en-la-republica-argentina.

[4] "Duplicación de Casos de COVID-19 En Argentina." n.d. Accessed November 23, 2020. https://institutodecalculo.shinyapps.io/tdcovid/.

[5] "Fases de Administración Del Aislamiento." 2020. May 13, 2020. https://www.argentina.gob.ar/coronavirus/aislamiento/fases.

[6] Figar, Silvana, Vanina Pagotto, Lorena Luna, Julieta Salto, Magdalena Wagner Manslau, Alicia Mistchenko, Andrea Gamarnik, Ana Maria Gomez Saldano, and Fernan Quiros. 2020. "Community-Level SARS-CoV-2 Seroprevalence Survey in Urban Slum Dwellers of Buenos Aires City, Argentina: A Participatory Research." *medRxiv*, July, 2020.07.14.20153858.

[7] Gupta, Rishi K., Michael Marks, Thomas H. A. Samuels, Akish Luintel, Tommy Rampling, Humayra Chowdhury, Matteo Quartagno, et al. 2020. "Systematic Evaluation and External Validation of 22 Prognostic Models among Hospitalised Adults with COVID-19: An Observational Cohort Study." *The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology*, September. https://doi.org/10.1183/13993003.03498-2020.

[8] Hosmer, D. W., Jr, S. Lemeshow, and R. X. Sturdivant. 2013. "Applied Logistic Regression." https://books.google.com.ar/books?hl=en&lr=&id=64JYAwAAQBAJ&oi=fnd&pg=PR13&dq=applied+logistic+regression+hosmer+lemeshow&ots=DtbP3V4ljO&sig=Rvs5dEoh-tltolBSkx3gBS66Pb8.

[9] "Hospitales." n.d. Accessed November 17, 2020. https://hospitales.puntodev.com.ar.

[10] "Información Epidemiológica." 2020. April 12, 2020. https://www.argentina.gob.ar/salud/coronavirus-COVID-19/sala-situacion.

[11] "Inicio - Municipio de Marcos Paz." n.d. Accessed November 17, 2020. https://www.marcospaz.gov.ar/.

[12] "Instituto de Cálculo." n.d. Accessed December 15, 2020. http://www.ic.fcen.uba.ar/.

[13] James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer, New York, NY.

[14] Mustafa, S. 2017. "Feature Selection Using Sequential Backward Method in Melanoma Recognition." In *2017 13th International Conference on Electronics, Computer and Computation (ICECCO)*, 1–4. ieeexplore.ieee.org.

[15] Pastor-Barriuso, Roberto, Beatriz Perez-Gomez, Miguel A. Hernan, Mayte Perez-Olmeda, Raquel Yotti, Jesus Oteo, Jose Luis Sanmartin, et al. 2020. "SARS-CoV-2 Infection Fatality Risk in a Nationwide Seroepidemiological Study." *Medrxiv*. https://www.medrxiv.org/content/10.1101/2020.08.06.20169722v1.abstract.

[16] Pueyo, Tomas. 2020. “Coronavirus: The Hammer and the Dance - Tomas Pueyo - Medium.” Medium. March 19, 2020. https://tomaspueyo.medium.com/coronavirus-the-hammer-and-the-dance-be9337092b56.

[17] “Regiones Sanitarias.” n.d. Accessed November 17, 2020. https://www.gba.gob.ar/saludprovincia/regiones_sanitarias.

[18] “Report 1 - Estimating the Potential Total Number of Novel Coronavirus (2019-nCoV) Cases in Wuhan City, China.” n.d. Accessed December 4, 2020. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-1-case-estimates-of-covid-19/.

[19] “Sistema Y Aplicación Cuidar.” 2020. May 12, 2020. https://www.argentina.gob.ar/jefatura/innovacion-publica/acciones-coronavirus/aplicacion-y-tableros-de-gestion.

[20] Taillandier, Patrick, Benoit Gaudou, Arnaud Grignard, Quang-Nghi Huynh, Nicolas Marilleau, Philippe Caillou, Damien Philippon, and Alexis Drogoul. 2019. “Building, Composing and Experimenting Complex Spatial Models with the GAMA Platform.” *GeoInformatica* 23 (2): 299–322.

[21] Vynnycky, Emilia, and Richard White. 2010. *An Introduction to Infectious Disease Modelling*. OUP Oxford.

[22] Williamson, Elizabeth J., Alex J. Walker, Krishnan Bhaskaran, Seb Bacon, Chris Bates, Caroline E. Morton, Helen J. Curtis, et al. 2020. “Factors Associated with COVID-19-Related Death Using OpenSAFELY.” *Nature* 584 (7821): 430–36.

[23] World Health Organization. n.d. “Coronavirus Disease (COVID-19).” https://www.who.int/docs/default-source/coronaviruse/situation-reports/20201005-weekly-epi-update-8.pdf.

[24] Yacobitti, Alejandro, Administrative Lisandro Otero, Adminsitrative Valeria Doldan Arruabarrena, Juan Arano, Sergio Lage, Martin Silberman, Martin Zubieta, et al. n.d. “Clinical Characteristics of Vulnerable Population Hospitalized with Diagnosed of COVID-19 in the Southeast Region of Greater Buenos Aires, Argentina.” *Journal of General Internal Medicine*.

[25] Yan, Carol H., Farhoud Faraji, Divya P. Prajapati, Christine E. Boone, and Adam S. DeConde. 2020. “Association of Chemosensory Dysfunction and COVID-19 in Patients Presenting with Influenza-like Symptoms.” In *International Forum of Allergy & Rhinology*. Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1002/alr.22579.

[26] Yan, Li, Hai-Tao Zhang, Jorge Goncalves, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun, et al. 2020. “An Interpretable Mortality Prediction Model for COVID-19 Patients.” *Nature Machine Intelligence* 2 (5): 283–88.

[27] Imai N., Dorigatti I., Cori A., Riley S., Ferguson N., Report 1 - Estimating the potential total number of novel Coronavirus (2019-nCoV) cases in Wuhan City, China, MRC Centre for Global Infectious Disease Analysis; Imperial College London, UK [cited 24 Nov 2020]. Available https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-1-case-estimates-of-covid-19/

[28] Pueyo T., Coronavirus:The Hammer and the Dance [cited 24 Nov 2020]. Available https://tomaspueyo.medium.com/coronavirus-the-hammer-and-the-dance-be9337092b56

Bios

Mehrnoosh Arrar

Mehrnoosh Arrar is a research scientist at the Instituto de Cálculo of the University of Buenos Aires. She obtained her PhD in Physical Chemistry from the University of California San Diego and her Bachelors in Chemistry from the University of Florida.

Laouen Belloli

Laouen Belloli is a graduate student in Computer Science from the University of Buenos Aires and he is currently finishing his second year as a Ph.D. student at the Laboratorio de Inteligencia Artificial (LIAA) of the University of Buenos Aires under the supervision of Ph.D. Diego fernández Slezak. Laouen research mainly focuses on computational psychiatry.

Ana M. Bianco

Ana M. Bianco holds a PhD and a degree in Mathematics from the Universidad de Buenos Aires. She is full professor at the Instituto de Cálculo and independent researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas of Argentina. Her reserach interests are in Mathematical Statistics, mainly focused on Robust Statistics.

Leonardo Boechi

Leonardo Boechi received his PhD at the University of Buenos Aires, and completed his postdoctorates at the University of California San Diego, where he was awarded by the Pew Foundation and by the Bunge & Born foundation. He has a permanent position at the Instituto de Calculo of the University of Buenos Aires. His principal area of interest involves statistical methods applied to biological systems.

Rodrigo Castro

Rodrigo Castro holds a degree and a PhD in Engineering from the National University of Rosario. He is a researcher in the UBA-CONICET Institute of Computer Science where he heads the Simulation Laboratory, and a professor in the Computer Science Department, School of Exact and Natural Sciences, University of Buenos Aires. He is a modeling and simulation consultant for a variety of public and private organizations in Argentina, USA and Europe.

Guillermo Durán

Guillermo Durán holds a PhD in computer science and a degree in mathematics from the University of Buenos Aires. He is head of the Instituto de Cálculo, associate professor at the University of Buenos Aires, adjunct professor at the University of Chile, and Senior Investigator at the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina. He is an operations research consultant for a variety of public and private organizations in Chile and Argentina.

Roberto Etchenique

Roberto Etchenique: obtained his degree in Chemistry (1994) and his PhD in Inorganic, Analytical and Physical Chemistry (1998) from the School of Exact and Natural Sciences at the University of Buenos Aires (FCEN-UBA), Argentina. Currently, he is Senior Investigator of CONICET, Head of the Molecular Devices Lab at the Institute of Physical Chemistry of Materials, Environment and Energy (INQUIMAE), CONICET and Full Professor at the Inorganic, Analytical and Physical Chemistry Department, FCEN-UBA.

Natalia Brenda Fernández

Natalia Brenda Fernández holds a PhD in biological science and a degree in biology from the the University of Buenos Aires. She is a postdoctoral fellow at the Consejo

Nacional de Investigaciones Científicas y Técnicas de Argentina. She carry out her research at the Instituto de Biociencias, Biotecnología y Biología Traslacional (iB3) at the University of Buenos Aires.

Luciana Ferrer

Luciana Ferrer is a researcher at the Computer Science Institute (ICC, for its acronym in Spanish), affiliated to the University of Buenos Aires (UBA) and the National Scientific and Technical Research Council (CONICET), Argentina. Luciana received her Ph.D. degree in Electronic Engineering from Stanford University, USA, in 2009, and her Electronic Engineering degree from the University of Buenos Aires, Argentina, in 2001. Her primary research focus is machine learning applied to speech processing tasks.

Diego Garbervetsky

Diego Garbervetsky is an Associated Professor in the University of Buenos Aires and Deputy director of the Institute of Research in Computer Science (ICC/CONICET). His research focuses on automatic program analysis techniques aimed at program understanding, verification and validation. He regularly publishes and participes in program committees on the main conferences in Software Engineering. He is highly interested in technology transfer and has been participating in several collaboration projects with the software industry. Since the beginning of the covid pandemic he was devoting his time in forming interdisciplinary groups that worked intensively in computational and mathematical models aiming at assisting policy makers with evidence based on data and analytics.

Rodrigo Goldsmit

Rodrigo Goldsmit is a med student from the University of Buenos Aires. He is a teaching assistant in Pharmacology at the University of Buenos Aires. He was one of the coordinators of the CETEC-U FCEyN during 2020.

Carolina Grillo Vidal

Carolina Grillo Vidal is a PhD student in the University of Buenos Aires, working at the Geochronology and Isotopic Geology Institute and have a doctoral scholarship of the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina. Is coordinate the Covid Prevention Center in the Faculty of Exact and Natural Sciences of UBA.

Juan E. Kamienkowski

Juan E. Kamienkowski holds a PhD and a degree in physics from the University of Buenos Aires. He is head of the Maestría en Data Mining (Universidad de Buenos Aires), and researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) from Argentina. He is a member of the Laboratorio de Inteligencia Artificial Aplicada (Instituto de Ciencias de la Computación, Universidad de Buenos Aires).

Pablo Laciana

Pablo Laciana is a senior software developer. He has a degree in Computer Science from the Faculty of Exact and Natural Sciences of the University of Buenos Aires. He currently works as a Technical Assistant at the UBA-CONICET Institute of Computer Science.

Esteban Lanzarotti

Esteban Lanzarotti holds a PhD in Biochemestry from the University of Buenos Aires and a degree in Computer Science.

Mario E. Lozano

Mario E. Lozano holds a PhD in Biochemistry from the National University of La Plata. Titular Professor in Biochemistry, National University of Quilmes (UNQ). Assistant Researcher, CONICET. Formerly: General Director of the National University Council for Argentina (CIN, 2017-2020). Rector of UNQ (2012-2016), Vice-Rector of UNQ (2008-2012); Dean of Department of Science and Technology, UNQ (2004-2008); General Secretary of UNQ, (2003-2004); Secretary of Student Affairs, Faculty of Exact Sciences of the National University of La Plata (1986-1988); Faculty and University Adviser, National University of La Plata, 1990-1994.

Rodrigo Maidana

Rodrigo Maidana is pursuing a degree in Economics at the National University of La Plata. Currently is an independent analyst and consultant of COVID-19.

Mauricio Mendiluce

Mauricio Mendiluce has a degree in Mathematics from the University of Buenos Aires. He is a PhD student in Mathematical Sciences at the Instituto de Investigaciones Matemáticas Luis A. Santaló (IMAS) belonging to the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET).

Sol Minoldo

Sol Minoldo holds a PhD in social sciences from the University of Buenos Aires and a degree in sociology. She is researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina.

Ezequiel Pecker-Marcosig

Ezequiel Pecker-Marcosig holds a Ph.D. in Engineering and a degree in Electrical Engineering from the University of Buenos Aires (UBA). Since 2015, he has been a Head Teaching Assistant with the Electronics Department (FI-UBA). He is a postdoctoral fellow with the Discrete Event Simulation Lab at the UBA-CONICET Institute of Computer Science (ICC).

Leonardo Pepino

Leonardo Pepino is pursuing a PhD in computer science at University of Buenos Aires and holds a degree in Acoustic engineering from the University of Tres de Febrero. He is a member of the speech processing group of the Applied Artificial Intelligence Laboratory, which is part of the Instituto de Ciencias de la Computación. His main interests are in deep learning and its applications to acoustic signal processing.

Armando Ezequiel Puerta

Armando Ezequiel Puerta is an advanced Computer Science student at the University of Buenos Aires. He is currently doing his Master's thesis at the Discrete Event Simulation Laboratory under the direction of Rodrigo Castro.

Rodrigo Quiroga

Rodrigo Quiroga holds a PhD in Molecular Biology and a degree in biochemistry from the University of Córdoba. He is an associate professor at the University of Córdoba in the School of Chemical Sciences, Department of Theoretical and Computational Chemistry, and an assistant researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas in Argentina.

Guillermo Solovey

Guillermo Solovey is a Professor at the Universidad de Buenos Aires and researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina. His research ranges different topics in quantitative cognitive science. He applies online experiments, computational and statistical models to understand human decision making. In particular, how people use evidence in their environment to update their beliefs and judge the veracity of information.

Marina Valdora

Marina Valdora holds a PhD in mathematics from the University of Buenos Aires. She is a professor at the University of Buenos Aires in the area of statistics and a researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina.

Mariano Zapatero

Mariano Zapatero is a senior software architect and developer. He is a consultant for several data-centric projects in varied institutions both in industry and academia. He is also the lead technical adviser and applications architect for the Discrete Event Simulation Lab, UBA-CONICET Institute of Computer Science.

Science Reviews from the end of the world

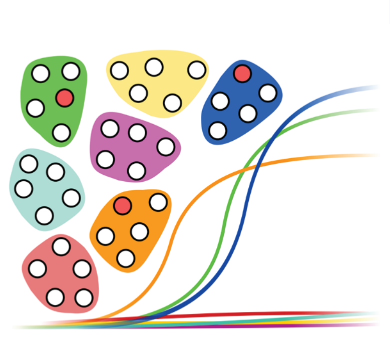
Vol. 2, No. 2, April 2021 // pp. 41-54 /

Pool Strategy for Surveillance Testing of SARS-CoV-2

Felipe Marceca[1,2], Luciana Rocha-Viegas[1,3], Nicolás Pregi[1,4], Gabriela Barbas[5], Daniela Hozbor[6], Adali Pecci[1,3] and Roberto Etchenique[1,7,*]

[1]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Buenos Aires, Argentina
[2]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Matemática, Instituto de Investigaciones Matemáticas "Luis Santaló", CONICET, Buenos Aires, Argentina
[3]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Instituto de Fisiología, Biología Molecular y Neurociencias, CONICET, Buenos Aires, Argentina
[4]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales, CONICET, Buenos Aires, Argentina
[5]Secretaria de Prevención y Promoción de la Salud, Ministerio de Salud de la Provincia de Córdoba
[6]Instituto de Biotecnología y Biología Molecular, Facultad de Ciencias Exactas, Universidad Nacional de La Plata y CCT La Plata, CONICET, La Plata, Argentina
[7]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Química Inorgánica, Analítica y Química Física, Instituto de Química Física de los Materiales, Medio Ambiente y Energía, CONICET, Buenos Aires, Argentina

** Corresponding author. E-mail: rober@qi.fcen.uba.ar*



Abstract

Due to the great morbidity and mortality in the risk groups of the pandemic COVID-19 caused by the emerging coronavirus SARS-CoV-2 and in the absence of effective therapeutic or preventive measures, quarantines, social distancing and the use of masks were the measures most used by health systems to reduce infections. The social, economic and health impact caused by these measures have begun to be evaluated in the different countries. These analyses lead to underestimations because in general they evaluate disease confirmed by a laboratory test and in some cases by epidemiological link without considering asymptomatic or oligosymptomatic infection. Therefore, mitigating fast circulation of the virus requires continuous tracking, detection, and isolation of cases, for which active surveillance able to address asymptomatic cases can make a valuable contribution over the dynamics of the disease in a given society, and to allocate adequate health resources and evaluate the effectiveness of control measures. Mathematical models such as the Susceptible-Exposed-Infectious-Removed (SEIR) allow not only to improve the estimates of the evolution of the pandemic at the local level, but also to evaluate health strategies. In the context of large testing requirements and the expansion of such testing capacity, it is also essential to develop approaches that improve the efficient use of these resources. Active surveillance undoubtedly contributes to improving estimates of virus circulation and it is of particular importance in vulnerable groups of high population density that have one or more risk factors, difficult access to the health system, and inhabit semi-closed facilities such as residential care homes, mental hospitals, prison houses, police stations housing prisoners, etc. Group testing strategies are especially useful for routine community survey and for monitoring of cohesive groups. While the frequency of infection in a population, who have only some symptoms compatible with the disease or do not have any symptoms, may be low, diagnosing even a single positive person typically requires quarantine of the entire group to prevent further spread in the community. In these surveillance strategies, pooling may allow more routine monitoring and detection of low frequency of carriage, thereby improving estimates, informing policy makers, reducing transmission, and alleviating the strain on healthcare services. By means of molecular tests based on RT-qPCR, the pooling strategy has been assayed with different algorithms also for COVID-19, particularly in the asymptomatic population, since a low prevalence of the disease is expected there. This has increased SARS-CoV-2 testing throughput while maintaining high sensitivity. Here, we discuss the relevance of some active surveillance strategies to determine key facts about COVID-19 pandemics and review different testing strategies that different countries have applied for tracking SARS-CoV-2.

Keywords: SARS-CoV-2, COVID-19, coronavirus, pooling, active surveillance

Introduction

At the end of last year, China informed the United Nations (UN) health agency of the existence of several cases of atypical pneumonia in the city of Wuhan caused by a novel coronavirus (SARS-CoV-2). On January 30, when there were only 80 cases outside of China, the UN declared COVID-19 an international public health emergency, the highest alert that the UN health agency can issue to warn governments about the need to take measures to prevent its spread.

At the beginning of February 2020, the World Health Organization (WHO) convenes a research and innovation forum on COVID-19, attended by more than 400 experts and funders from around the world. By March 11, due to the alarming levels of spread and severity of the disease, the WHO determined in its assessment that COVID-19 is a pandemic. The recently emerged coronavirus pandemic is critically challenging healthcare systems around the world. As of March 17, the WHO reported more than 170,000 cases with more than 10,000 new diagnoses added in those last 24 hours; and till September 2020 SARS-CoV-2 has infected more than 25 million people and provoked more than 840,000 deaths all over the world. Nowadays (March, the 24th 2021), nearly 12902242 cases of COVID-19 have been reported, including more than 2,5 million deaths with several cities going through their second wave of increased cases (https://covid19.who.int/). In Argentina, particularly in the Province of Buenos Aires, the first confirmed case was detected on March 8, three days later than the first detected case in the country. From this date on, the rate of increase in the weekly average of total number of cases as well as in the daily reported cases was fast until the end of May, although few cases were still reported.

The complex COVID-19 symptomatology was recently classified in six groups that might correlate with illness severity [1,2]. Elderly, and those with underlying medical conditions like heart and lung problems, high blood pressure, diabetes, or cancer, are at higher risk of developing serious illness. Meanwhile, other age groups can become infected and develop moderate symptoms or even carry the SARS-CoV-2 infection asymptomatically. These groups of oligosymptomatic, presymptomatic and asymptomatic people represent a great concern for the health system since they may go unnoticed while contributing to the circulation of SARS-CoV-2 [3–5].

The pandemic context more clearly shows that urgent need for systematic collection, analysis and interpretation of health data are fundamental for the planning, implementation and evaluation of public health policies. It is important to remember at this point that epidemiologic surveillance in general, and for COVID-19 in particular, is supported by passive surveillance systems, active surveillance and specialized or sentinel systems. The first begins when the case consults the health personnel and is registered through the usual case notification system. In passive surveillance, data is collected from as many sources as possible. In active surveillance, the personnel in charge actively seek information about the disease under investigation. Medical personnel are contacted, health care centers are visited, health records are analyzed for signs of the disease, and cases are searched in the field. If suspicious cases are detected, samples are taken and sent to laboratories for analysis. All actors who can offer information on possible cases, with which they have been in contact with, are contacted. As part of active surveillance, the competent authorities are also quickly informed through previously established channels. The sentinel surveillance system uses high-quality data, collected in specialized centers that are carefully selected. Centers and professionals specialized in the disease under surveillance and high-quality diagnostic laboratories generally participate.

The first population study collecting information on the demography, clinical presentation, hospitalization, contact network and the presence of SARS-CoV-2 infection, was carried out during February 2020 in Padua (Italy). In this study, two surveys were conducted, with a time difference of two weeks. A prevalence of infection of 2.6% was found in the first survey, which was carried out at the beginning of the town lockdown. In the second survey, which was carried out at the end of the confinement, prevalence down to 1.2% [6], supporting the efficacy of the control measures. A second study, performed in Iceland, during March 2020 compared two sampling strategies, one using an open invitation method to the entire population of Iceland, and a second sampling on a random population. [7]. Then, another study performed by the Public Health Agency of Sweden, in collaboration with Swedish Armed Forces, conducted a prevalence survey in Stockholm. In this study, 2.5% tested positive for SARS-CoV-2 infection, estimating 59364 infected people in Stockholm in the last week of March. This protocol was repeated at end of April and resulted in a prevalence of 2.3%. However, a nation-wide study revealed a prevalence of 0.9% (https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/). Finally, the Office for National Statistics, the University of Oxford, the University of Manchester, Public Health England, and the Wellcome Trust conducted a COVID-19 infection survey at the UK from April 27 to May 10 that estimated an average of 0.27% to be infected in the community population (https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/england21may2020).

All these studies were based on individual testing, which makes them expensive and time consuming to reproduce on large scales. In this way, conducting large-scale population RT-qPCR studies through pooling samples seems to be the most efficient strategy to perform.

Regardless of the surveillance system used, the novelty of the disease made it necessary from the beginning to have a specific laboratory tool for detecting SARS-CoV-2 infection in order to confirm the clinical case. This methodology refers to molecular techniques based fundamentally on reverse transcription techniques and quantitative polymerase chain reaction (PCR). Depending on the objective, this strategy can be applied to a sample from a single individual or to a pool of samples from different individuals. This last procedure will be our main focus, since it is used to gain information at a population level and is essential for the continuous tracking, detection and isolation of cases.

The pooling of individual samples as a cost-saving method for diagnosis of infectious diseases is presented as a possible strategy to confront the COVID-19 pandemic. In the past, this group testing strategy was successfully used for identifying individuals with syphilis, HIV, hepatitis B virus, hepatitis C virus, *Chlamydia trachomatis* and *Neisseria gonorrhoeae* using different biological samples [8–15]. As for COVID-19, the pooling strategy has been applied for the general population in different countries *i.e* Australia [16], China [17], Ecuador [18], Germany [19], Ghana [20], India [21], Israel [22], Malaysia [23], Malawi [20], South Africa [20], Tunisia [24], Uruguay [25], US [26,27] and Zimbabwe [20] among others.

In accordance to the US Centers for Disease Control and Prevention (CDC) guidance, laboratories certified under the Clinical Laboratory Improvement Amendments (CLIA) can use a specimen pooling strategy to expand SARS-CoV-2 nucleic acid diagnostics or screening testing capacity when using a test authorized for such use by the U.S. Food and Drug Administration (FDA) [Policy for Coronavirus Disease-2019 Tests During the Public Health Emergency (Revised) ("Policy for COVID-19 Tests")]. In other countries, this strategy is carried out in experienced laboratories that work in coordination with the ministries or secretaries of health. In its simplest form, this strategy consists of two-stages: first, mixing several samples together in a single tube and then testing the pooled samples in only one reaction. If the pooled sample is negative, it can be presumed that all individuals were negative. However, if the pooled sample turns out positive and individual identification is requested, a second stage is needed, where each sample will be separately tested to find out which individual sample was responsible for that result. In low prevalence scenarios, because pooled samples are often negative, ultimately fewer tests are run overall, saving time and testing supplies, and therefore, the results can be returned to patients more quickly in most cases. The advantages of this two-stage specimen pooling strategy include preserving testing reagents and resources, reducing the amount of time required to test large numbers of specimens, and lowering the overall cost of testing.

By implementing the eight-sample Dorfman pooling to test 26,576 samples from asymptomatic employees from the Hadassah Medical Centre (HMC) of Israel, 31 (0.12%) SARS-CoV-2 positives were identified, thus achieving a 7.3-fold increase in throughput [28]. In another report, by using a Shiny application (https://www.chrisbilder.com/shiny), it was proposed that the most efficient pool size is five specimens [27]. In Argentina, this strategy was performed successfully in semi-closed facilities in the Province of Buenos Aires at different pandemic periods [29]. With this strategy, the efficiency, measured as people tested per determination, ranged from 2 to 7.4, depending on the place of origin, with an average of 3. Remarkably, it allowed the early detection of outbreaks, and in addition, has evidenced that infected individuals, either asymptomatic or presymptomatic, may present cycle threshold (Ct) values as low as those of symptomatics.

Here we review those group testing strategies that have been applied for tracking COVID-19, the different algorithms used for saving resources compared to testing samples individually and the assessment of pooling performance. We also discuss the importance of group testing to carry on an active surveillance able to address a valuable contribution for the detection of asymptomatic cases.

Molecular diagnostic tests based on RT-qPCR

Since the beginning of the pandemic, there was a desperate need to develop diagnostic tests for COVID-19 to limit the spread of the virus. The whole genome sequencing of SARS-CoV-2 performed on January 10 led scientists to design testing protocols to detect the pathogen in the affected people. Though alternative diagnostic techniques to use at the point of care are also required, real time reverse transcription polymerase chain reaction (RT-qPCR)-based methodologies have been widely recommended. In fact, the WHO proposed this technique as the frontline diagnostic approach to detect SARS-CoV-2 infection in suspected patients. The assay is carried out by isolating RNA from a clinical sample (nasopharyngeal or oropharyngeal swabs, saliva) and adding it to a qPCR master mix containing specific forward and reverse primers.

Most countries, the US and Argentina among them, mainly use a one-step PCR format to diagnose COVID-19. The list of the different kits approved for use in these countries are detailed on these web pages https://www.fda.gov/media/134922/download, https://www.argentina.gob.ar/anmat/regulados/productos-medicos/reactivos-covid-19, respectively.

Group testing strategies

As mentioned before, the COVID-19 pandemic has attracted renewed attention to group testing strategies. By using a single test for a group of samples that are pooled together, these strategies allow the identification of positive samples efficiently, saving time and resources.

Group testing is any procedure that allows identifying certain specimens by testing groups rather than each specimen individually. It was first introduced by Dorfman to dramatically reduce the number of tests needed to screen future soldiers for syphilis [30]. Since then, it has grown into a rich subject that can be studied from different perspectives such as statistics, combinatorial mathematics and information theory. Applications include the identification of positive samples in medicine which will be our main focus, physical mapping of genes in molecular biology, multiple access communications, data storage and compression, and cybersecurity [31–33]. Let us describe the main idea behind group testing. For a set of samples that can either be positive or negative one can group any number of them together in a pool and perform a single test. If the test is negative every sample in the pool is presumed negative, whereas if the test is positive there is at least one positive sample in the pool. For example, Dorfman's original algorithm consists of two stages. First the samples are divided in groups of a certain size and each group of samples is tested in a single pool. During the second stage, samples in positive pools are tested individually and all positive samples are identified (Figure 1).

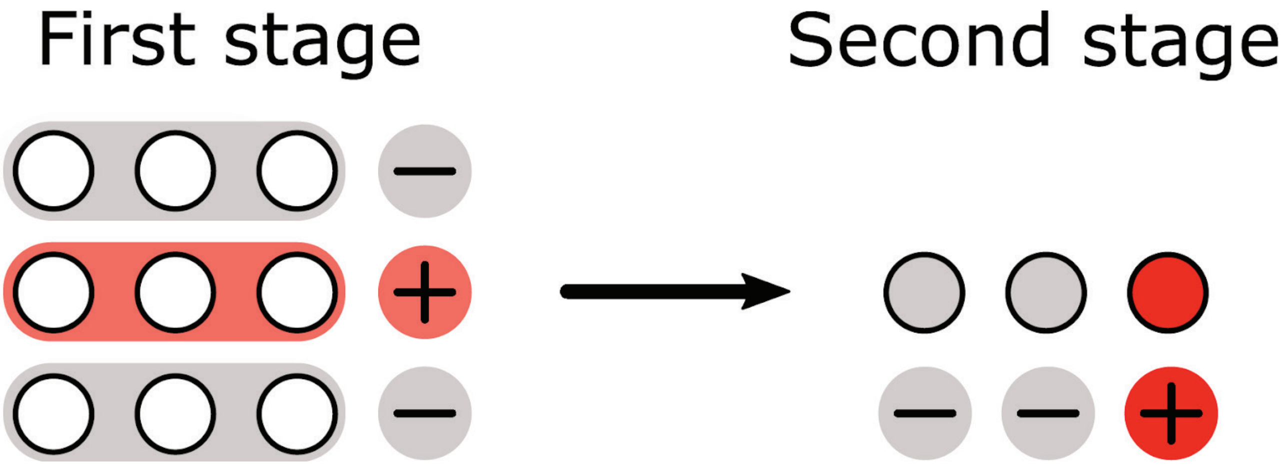
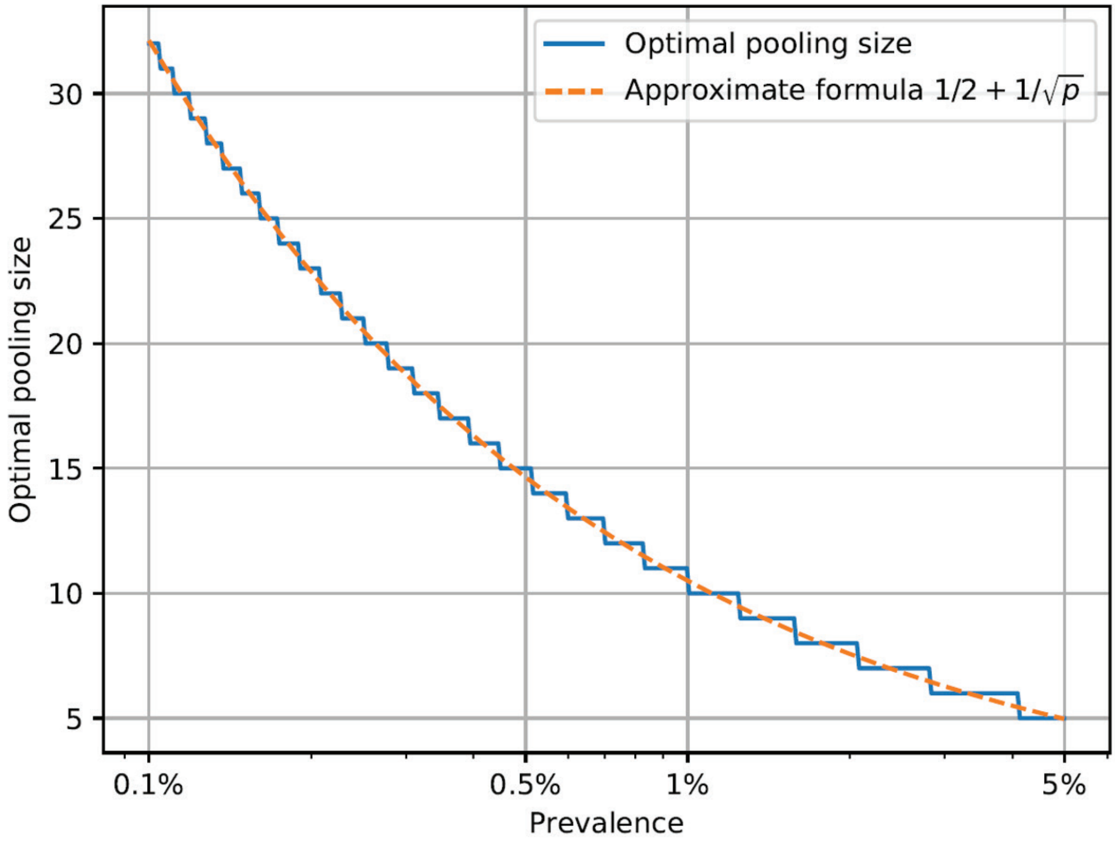


Figure 1. Scheme of Dorfman´s algorithm: Two stages are carried out. Samples are divided into groups and for each group, one test is made pooling its samples together. Then, samples from positive pools are tested individually.

When the number of positive samples is small, group testing allows to quickly rule out large numbers of negative samples using few tests. Group testing strategies delve into the optimal design of pooling schemes according to different criteria such as minimizing number of tests, time or other resources. They are also subject to constraints such as maximal pool size due to dilution effects [22]overwhelmed healthcare systems are already experiencing shortages of reagents associated with this test, calling for a lean immediately applicable protocol. Methods: RNA extracts of positive samples were tested for the presence of SARS-CoV-2 using reverse transcription quantitative polymerase chain reaction, alone or in pools of different sizes (2-, 4-, 8-, 16-, 32-, and 64-sample pools, maximum number of simultaneous tests due to equipment capacity, etc. For a given group-testing method, parameters such as the sizes of the pools are optimized depending on the presumed positivity (the proportion of infected individuals in the whole sample). If the individuals are randomly chosen from the population, it equals the prevalence. The lower the prevalence, the bigger the pool-sizes and the bigger the savings on tests will be. For example, to achieve the minimum number of expected tests in Dorfman's method for a prevalence p one has to use a pool of size roughly $1/\sqrt{p}+1/2$ and one should expect to reduce the number of tests by a factor of $2\sqrt{p}$. Therefore, it is important to know how sensitive the method is to changes in prevalence and to have a good estimation of this parameter (Figure 2). Also notice that pooling strategies are more effective than individual testing only when the prevalence is sufficiently small. For example, Dorfman's method offers no improvement whenever $p\approx 0.25$ (that is, $p\approx 25\%$) so that $2\sqrt{p}\approx 1$. It should be mentioned that usually these models assume samples are independent. This is not the case when testing closed facilities such as nursing homes, where samples have a high correlation that should be taken into account in the design of the pooling scheme.



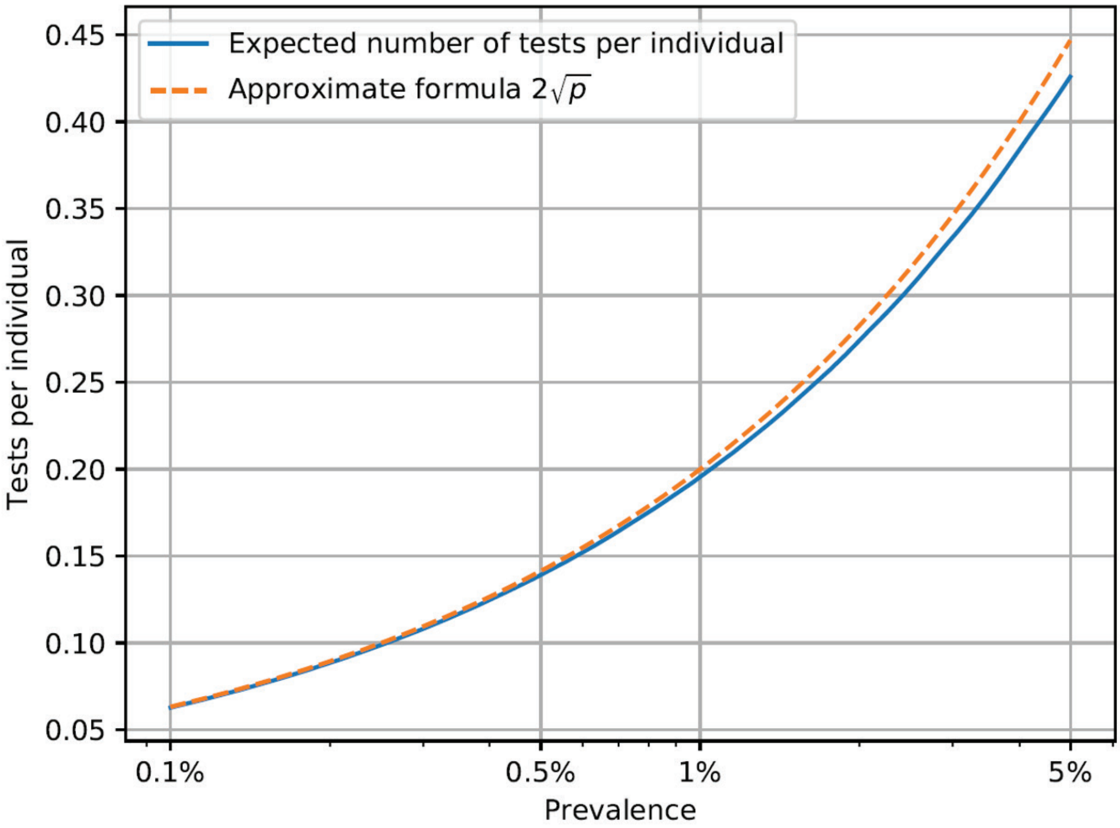


Figure 2. *Upper panel:* optimal pooling size for Dorfman's algorithm in terms of the prevalence that minimizes the expected number of tests per individual. *Lower panel:* expected number of tests per individual for Dorfman's algorithm in terms of the prevalence when using the optimal pooling size.

There are two general types of group testing schemes known as adaptive and non-adaptive. Adaptive strategies perform tests sequentially and use the results of previous tests to determine which test will follow. This makes them more effective in saving tests and easier to tailor but may consume more time and be more difficult to scale up. For non-adaptive strategies on the other hand, tests are predetermined and conducted simultaneously. This allows streamlining test-flow and avoiding waiting for previous results to prepare the next batch of tests. However, non-adaptive approaches may sometimes leave samples unclassified if the proportion of positive samples is unusually high.

Dorfman's original algorithm is a simple example of an adaptive scheme that is being implemented in several countries to identify SARS-CoV-2 among asymptomatic individuals (see for example [28]). There are also several variants of Dorfman's method which have been discussed in the context of the COVID-19 pandemic. One approach known as nested pooling, is to increase the number of stages by dividing positive pools into smaller sub-pools and iterate this process a specified number of times before reaching individual testing [34,35] (Figure 3). Increasing the number of stages can notably reduce the number of tests required (even among other group-testing strategies), but also adds time and complexity to the procedure.

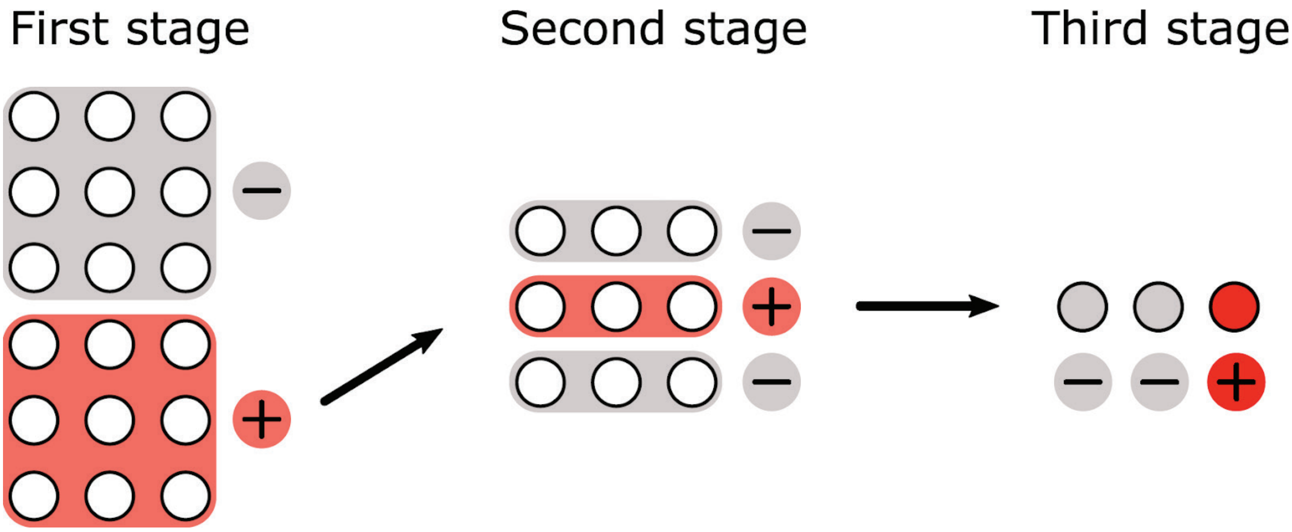


Figure 3. Nested pooling. Iteration process based on dividing positive pools into smaller sub-pools a specified number of times before reaching individual testing.

Another approach is to maintain only one group-testing stage and one individual-testing stage as in Dorfman's method, but to improve it by implementing a more sophisticated group-testing design. This can be done for example by arranging samples in a square matrix and testing each row and each column in a different pool (Figure 4). In other words, each sample is included in two pools, one for its row and one for its column. Since samples on negative pools are considered to be negative, only samples whose row and column are positive have to be tested in the individual stage. In particular, if there is only one positive sample, it can be identified directly in the first stage and no further testing is needed. A suboptimal pool-size can be tuned so that this happens quite frequently prioritizing saving time over tests. This method was implemented for surveillance in closed facilities such as nursing homes [29]. Generalizations replacing matrices with higher dimensional arrays were proposed [36].

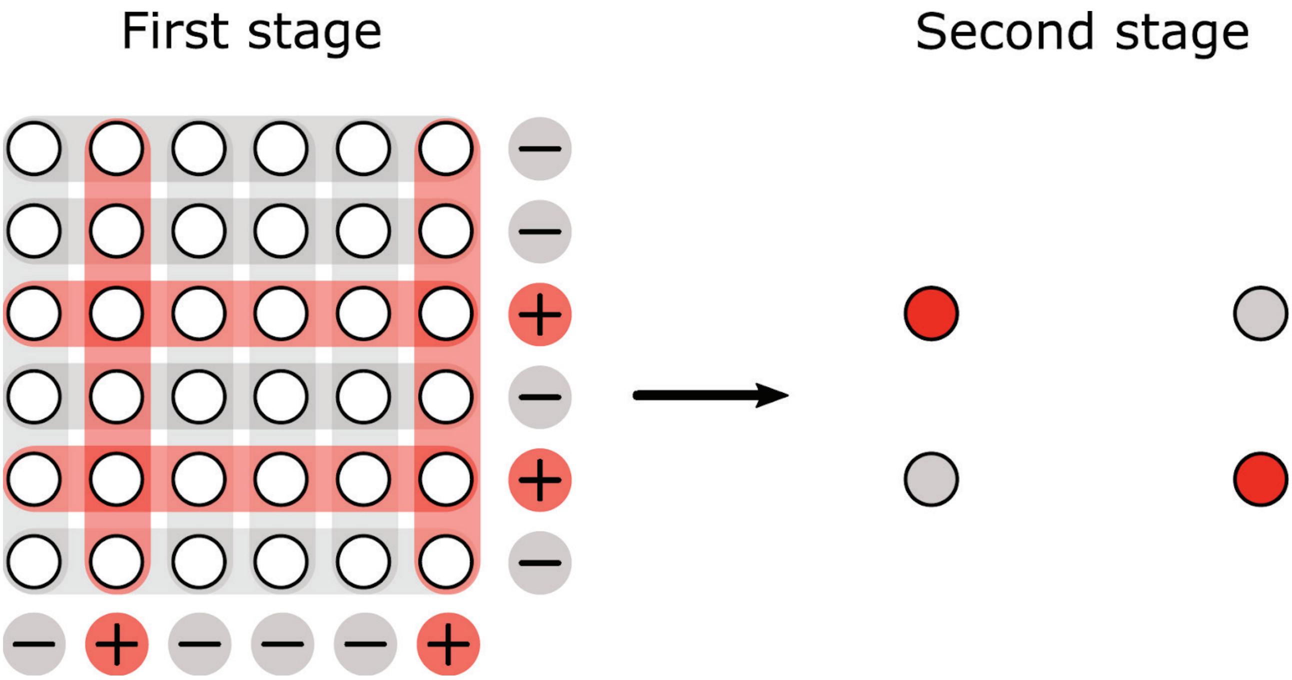


Figure 4. Matrix pooling. Square-matrix arrangement where each sample is included in two pools, one for its row and one for its column. Only samples whose row and column are positive are then tested in the individual stage.

Regarding non-adaptive methods, a single group-testing stage is designed including each sample in several pools. The assignment is done in such a way that the maximum number of positive samples can be identified with the least number of pools in one go. These designs are usually inspired on objects well-known in the context of combinatorial mathematics as well as on error correcting codes from computer sciences. For example, Steiner systems [37] and an error correcting code [38] were employed. Furthermore, in the latter, their non-adaptive pooling design was implemented to screen 1115 health care workers using 144 tests. The success of non-adaptive methods relies heavily on the proportion of positive samples not exceeding a threshold value. For this reason, prevalence has to be carefully estimated leaving a safe margin. Another drawback is that pool designs of non-adaptive methods are very intricate by nature, which may make them difficult to implement without automation.

Assessment of pooling strategy performance

Generally, the FDA recommends validating the test with either pooling approach (sample/media pooling or swab pooling) in a way that preserves the sensitivity of the test as much as possible. That is, it is preferable to use an approach where all specimens identified as positive when tested individually are also identified as positive when tested using the pooled testing approach. However, a decrease in performance is likely with pooling strategies, due to dilution of the primary clinical sample. Nevertheless, since sample pooling will greatly increase the number of individuals that can be tested using existing resources, a small reduction in sensitivity may be acceptable depending on the pooling efficiency and other mitigations in place. Therefore, the FDA generally recommends that, after pooling, test performance includes ≥85% percent positive agreement when compared with the same test performed on individual samples (https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/pooled-sample-testing-and-screening-testing-covid-19). An additional limitation is that samples from a negative pool are considered presumptive negatives but never tested individually. Therefore, the pool size must be chosen so that the amount of false negatives generated by this procedure remains low. In this sense, key principles for successful application of group testing involve knowledge of the limit-of-detection, sensitivity, and specificity of the assay, and the prevalence of disease in the population. The goal of the process is to determine a pool size that provides the greatest conservation of resources while maintaining the reliable performance of testing.

The University of Stanford in the US performed a retrospective study with pools of 9 or 10 individual samples that had not been tested for SARS-CoV-2 early in the pandemic. This pooled screening facilitated detection of early community transmission and enabled timely implementation of appropriate infection control measures to reduce spread [26]. In the meantime, within the community of Nebraska with a specimen positive rate around 5%, a pooling ratio of 1 to 5 was expected to retain accuracy and result in greater efficiency of test resources. Results indicated that all positive samples by the non-pooled method were detected in pools with 4 other negative samples. Determination of the optimal pool size prior to testing is capable of improving the overall efficiency of pooled testing [27].

Within a 6.8% positive rate, Spanish clinical microbiology laboratories obtained excellent results in terms of sensitivity, specificity and positive and negative predictive values in pools of 9 and 10 samples [39]. However, web-based applications for pooling estimate that a pool size with 5 specimens may conserve more resources than a pool size of 10 as the prevalence rate increases [23].

In Germany, two pooling techniques were compared, a “routine high throughput” approach where random samples are pooled together for testing, or a “door to door” approach where cohesive groups (ie, families, neighbors, etc.) are pooled together for testing. While both approaches save substantial resources, the “door to door” approach was found to carry more benefit, reducing tests by 56% to 93%, whereas the “routine high throughput” resulted in 24% to 86% fewer tests. In low to moderate infection levels, even a pool size of 5 would reduce the number of tests needed by 5-fold (78%) [40].

In countries with infection levels over 20%, a pool size of 10 would still result in a considerable reduction in the number of tests required (up to 50%). These studies showed that pooled testing is able to detect positive samples with sufficient accuracy. It is best used for population-wide screening, contact tracing, and the monitoring of essential workers and asymptomatic individuals with unidentified risk, such as in airports, versus being much less effective if used in settings with high clinical suspicion, such as patients showing symptoms. While concerns exist that low positive samples, such as those found in convalescent patients, could escape detection with increasing pool size, additional amplification cycles could be employed to allow better detection of larger pools. Nevertheless, group testing shows great potential in increasing testing capacity with existing resources with minimal loss of accuracy [40].

With a positive rate (even in asymptomatic persons) within 5%, the Indian Council of Medical Research observed that pooling of 5 or 10 specimens did not affect the sensitivity of detecting SARS-CoV-2 when the PCR cycle threshold (Ct) of the original specimen was lower than 35. However, in specimens with low viral load (Ct > 35), 13.3% were false negative. Thus, in pooled samples, graphs should be analyzed for the sigmoid curve even beyond Ct value 40, and in case of the appearance of any graph, the RT-qPCR should be repeated with deconvoluted samples [41].

When prevalence is low, 5-sample pooling versus 10-sample pooling showed Ct values differences of less than 5 as compared with individual tests [23]. Even more, a single positive sample can be detected in pools of up to 16-32 samples, with an estimated false negative rate of 10%. Despite a relatively restrictive cutoff, pooling would reach sensitivity of 96% for a pool size of 16 samples. Single positive samples can be detected when pooling either after or prior to RNA extraction [22]. Consistently with this, a comparison of Ct value differences between pool sizes from 4 up to 30 samples and individual positive samples ($Ct_{pool} – Ct_{positive}$ sample) demonstrated to be in the range of up to 5 [19].

An important concept to take into account when testing by pooling is the validation of the selected analytic method. This will allow us total control over the testing. FDA suggests a plan for ongoing monitoring of the positivity rate and of the performance of a test with a pooling strategy. In this sense, laboratories must perform their own validation pool studies for kits used for each RNA extraction and DNA amplification, based on the prevalence rate of COVID-19 in their own region. Finally, pooling is an effective approach to expand the impact of limited test resources and reagents during specific stages of an infectious disease outbreak.

In the province of Buenos Aires (Argentina) pooling strategies were successfully carried out in semi-closed facilities. By comparative analysis with single tests, we detected that clusters of 5 samples for RNA extraction and 10 (two sub-pools containing RNA extracted from 5 samples) is adequate since only 2 Cts are increased on average [29]. Concordance between the 5 or 10 sample pool and individual sample testing was 100% in the Ct value ≤36 cycles. As reported by other authors, the effectiveness of sample pooling is dependent on the prevalence of positive samples within the pool [42]. The sample pooling strategy worked best in settings with low prevalence and individual subgroups with low clinical suspicion such as asymptomatic individuals. Individual groups for whom there is a higher pre-test probability and those with serious manifestations should not be included in pooled testing but should be tested individually. To date and since the end of May we have analyzed more than 11,000 samples using the pooling test strategy and the obtained data suggests that pooling of up to 5-10 samples *per* pool can increase test capacity with existing equipment and test kits and detects positive samples with sufficient diagnostic accuracy.

This strategy was also implemented by the Ministry of Health of the province of Córdoba at the Faculty of Chemical Sciences and Ceprocor Ministry of Science and Technology. From the end of July until the end of October, when the positive rate ranged from 0.7% to 6.0%, the number of people tested per test used ranged from 5 to 2.

Group testing also evidenced that asymptomatic/presymptomatic/or oligosymptomatic infected individuals may present Ct values as low as those of symptomatic individuals. Furthermore, early stages of COVID-19 outbreaks were detected in different localities of Buenos Aires, increasing the likelihood of saving lives especially in places where risk groups were concentrated.

Importance of active surveillance to identify key facts about COVID-19 pandemics

Asymptomatic transmission of SARS-CoV-2 has been one of the most widely extended controversies since COVID-19 pandemics started. The initial evidence suggested that this new coronavirus was transmitted by respiratory droplets, mainly through coughing or sneezing as many other respiratory diseases (https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html). Additionally, surface-related contagion was considered key. As coughing is a typical symptom of COVID-19 infection, symptomatic contagion was seen as the main route by which SARS-CoV-2 enters and diffuses into societies.

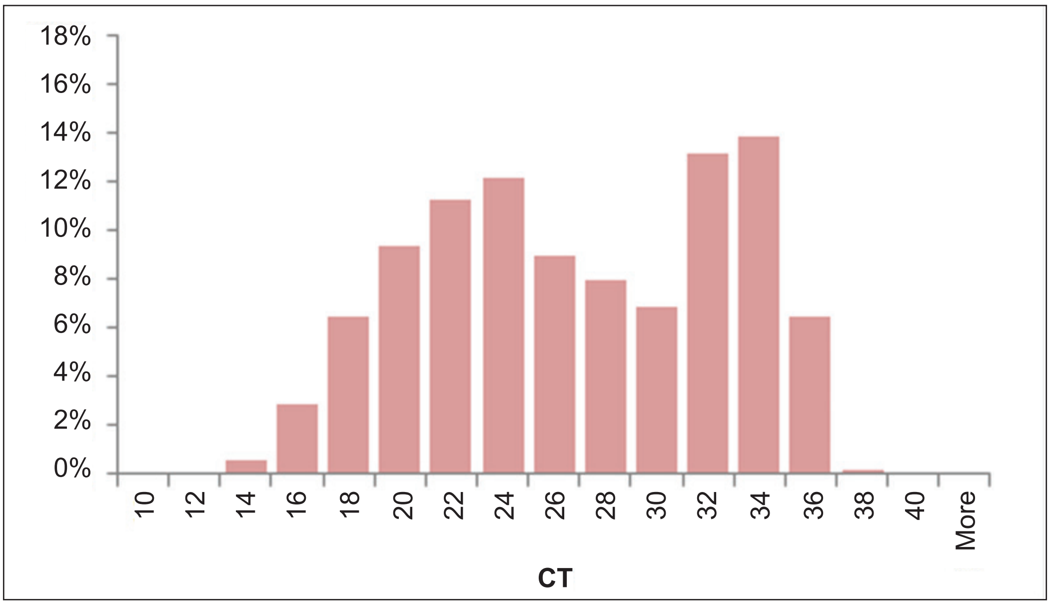
However, the course of the pandemics clearly showed that other transmission ways could be responsible for most of the contagions, especially when a majority of the people assessed the risks of COVID-19 transmission and preempted any contact with other people after having compatible symptoms. Perhaps the most important of these ways is the transmission by aerosols, which acquired notoriety after the Guangzhou restaurant episode [43]. Aerosols are tiny droplets, much smaller than the ones involved in cough-related emissions. While these droplets fall down in seconds timescale, nanodrops constituents of aerosols can remain suspended for minutes [44]. This characteristic, together with the possibility of air stirring and active dissemination through convection, increases the risk of being in close spaces with an infected person, even if no symptoms are present. The mere act of talking, singing, or even breathing can transmit the virus through aerosolized particles, which are then inhaled by a susceptible person. The CDC site, after several months denying the "airborne" (aka. aerosolized) transmission, changed its view by September 2020, and stated that this mechanism constitutes a way of contagion in closed environments.

A key question to understand SARS-CoV-2 transmission is the fraction of the contagions that are due to exposure to individuals that are asymptomatic at the very moment of the contact. This definition implies that the somewhat usual differentiation between a- and pre- symptomatic must be left aside in order to focus on the practicality of the issue.

"Infodemic" about COVID-19 has profusely touched the asymptomatic issue. Its relationship with commercial activity lockdowns and travel barriers, school opening, and many other social-related situations have put this controversy in the center of political agendas. Moreover, reports of indisputable goals in the fight against COVID-19, as the massive testing after complete remission of local epidemics in Wuhan, China, have been misquoted as proof that asymptomatic contagion practically does not exist [17]. On the other side, authoritative institutions such as US CDC have put a warning on this danger, indicating that the absolute majority of the contagions are due to asymptomatic individual transmission (https://www.cdc.gov/coronavirus/2019-ncov/more/masking-science-sars-cov2.html). The transmission potential of any individual arises from two main factors: the number of close encounters with other subjects and the probability of transferring SARS-CoV-2 to the susceptible one. While the first depends mainly on social factors, the latter is closely related to the viral load in the upper respiratory tract of the infected person.

In our experience, around ten thousand of asymptomatic individuals were tested within the framework of active surveillance strategies designed by the Ministry of Health of the Province of Buenos Aires [29]. That means that we have detected hundreds of asymptomatic/presymtomatic individuals who were positive for SARS-CoV-2 RT-qPCR, and their Ct values can be used as a proxy of their viral loads and compared with symptomatic persons from the same regions.

Figure 5 shows the histograms for the N gene Ct values in two sets of clinical samples (nasopharyngeal swab). At the top, the samples corresponding to symptomatic individuals show a bimodal distribution which is characteristic of COVID-19 [29,45]. A preliminary inner analysis of this data suggests that no correlation exists between Ct value (value that estimates the viral load) and symptomatic severity, according to other findings [46] although a high correlation was described for IgG and IgM serological tests. On the bottom panel, the asymptomatic samples show a similar distribution, also bimodal. The average Ct is indeed slightly lower in the asymptomatic cases, although no significant differences are apparent. The log dependence of Ct values obscures the fact that the viral loads comprise a span of 4 million-fold among both symptomatic and asymptomatic individuals. As a rule of thumb, derived from different tests based in culture infectivity, Ct below 30 are considered infectious. Both distributions have two thirds of the positive samples in this zone, while the very high infective, superspreader-like viral loads of Ct<20 (1000-fold higher or more) constitute the 20% of the symptomatic individuals and more than 30% of the asymptomatic ones.



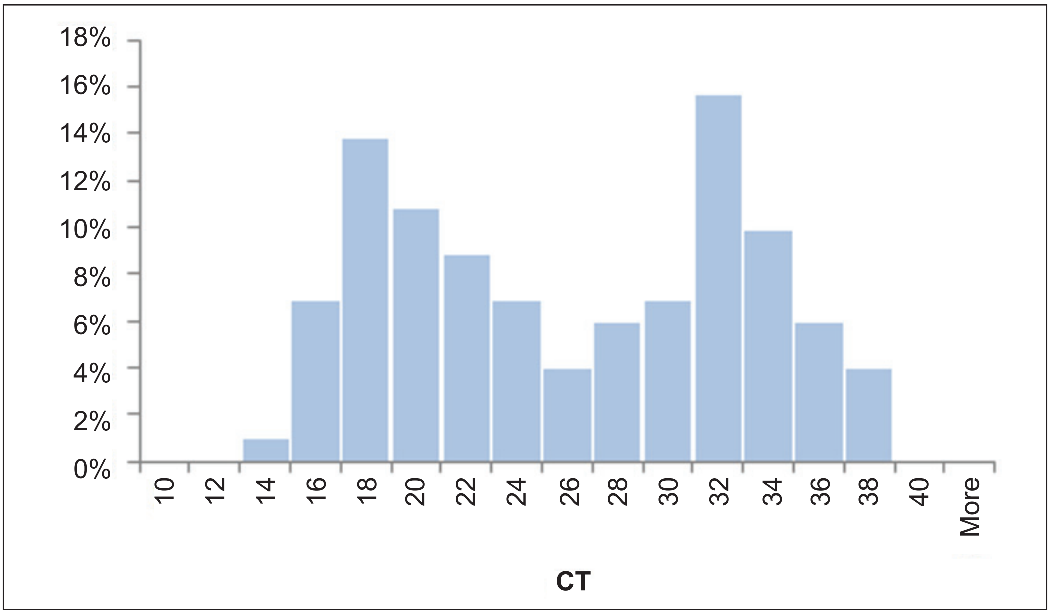


Figure 5. Distribution of cycle threshold (Ct) for N gene in Top: symptomatic individuals (N=526) and Bottom: asymptomatic individuals (N = 102).

Given this insight, and the now recognized fact that asymptomatic manifestation of COVID-19 constitutes the main fraction of all cases [47], it immediately follows that active surveillance of asymptomatic infected individuals becomes mandatory in order to be effective in the task of killing the contagion dynamics. For a sensitive and very selective, but rather slow and expensive test, as RT-qPCR is, pool strategy can be a good choice, especially in low prevalence conditions.

From the comparison of the detection probability and the complete histogram of Ct values obtained from the clinical samples, we estimated the robustness of the pool testing method. In this sense, the probability of positive detection of a single sample in a 20-samples pool was assessed. The recovery value was 95.3%. The 5% that was lost was not evenly distributed among the samples but corresponded to the lower viral loads of the specimens [29]. The very low Limit of Detection (LoD) of this technique allows even very high numbers of individual samples per pool. Together with a fast reaction from the health system, pooled testing of asymptomatic individuals can be the more cost-effective choice.

Concluding remarks

The implementation of tests in pooled samples for COVID-19 allow the expansion of detection capabilities in the community, as well as in groups of close contacts, such as hospital institutions, or institutions for the care of people in general. This methodology not only saves consumables but also time, which is a key aspect for making timely decisions of health systems. Our experience has shown that the use of pooling strategy in an active surveillance allows for the early detection COVID-19 outbreaks, increasing the likelihood of saving lives especially in places where risk groups are concentrated. As the capacity of the health systems for follow-up the pandemic is limited because of its magnitude, trials in most countries generally focus on patients with severe, or at least obvious symptoms, while potentially infectious carriers in the community remain undetected with the consequent spread of the virus. The detection of SARS-CoV-2 infected individuals including the asymptomatic/presymptomatic or oligosymptomatic (APO) can be approached from the laboratory with strategies such as those described here. As expected, the pooling strategy works best in settings with low prevalence and cohesive individual subgroups with low clinical suspicion such as APO individuals. The use of this strategy is not only of crucial importance during the pandemic in the absence of vaccination against COVID-19, but also upon vaccination in order to be able to monitor the infection particularly, in vaccinated populations who may not undergo the disease but do present infection.

Acknowledgements

We thank the Ministry of Health of the province of Buenos Aires and the Central Laboratory of the Ministry of Health of the province of Córdoba for the coordination and design of the active surveillance of COVID-19, and for providing the necessary inputs. We also thank the School of Natural and Exact Sciences (UBA), the School of Exact Sciences (UNLP) and the Institute of Virology “Dr. J. M. Vanella” (Faculty of Medicine, UNC), for their contributions of inputs. We especially thank Axel Kicillof and Carlos Bianco for their support of active surveillance as a strategic tool, which made possible the early start of this work. This work was made with a grant from the National Agency of Scientific and Technology Promotion (COVID-19 GRANT #843). LRV, NP, DH, AP and RE are staff of CONICET.

References

[1] Sudre, C.H., Lee, K.A., Lochlainn, M.N., Varsavsky, T., et al., Symptom clusters in COVID-19: A potential clinical prediction tool from the COVID Symptom Study app. *Sci. Adv.* 2021, 7, eabd4177.

[2] Yang, J., Zhao, X., Liu, X., Sun, W., et al., Clinical Characteristics and Eosinophils in Young SARS-CoV-2-Positive Chinese Travelers Returning to Shanghai. *Front. Public Heal.* 2020, 8.

[3] Gao, M., Yang, L., Chen, X., Deng, Y., et al., A study on infectivity of asymptomatic SARS-CoV-2 carriers. *Respir. Med.* 2020, 169.

[4] Li, H., Wang, Y.Y., Ji, M., Pei, F., et al., Transmission Routes Analysis of SARS-CoV-2: A Systematic Review and Case Report. *Front. Cell Dev. Biol.* 2020, 8, 618.

[5] Zhou, R., Li, F., Chen, F., Liu, H., et al., Viral dynamics in asymptomatic patients with COVID-19. *Int. J. Infect. Dis.* 2020, 96, 288–290.

[6] Lavezzo, E., Franchin, E., Ciavarella, C., Cuomo-Dannenburg, G., et al., Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo’. *Nature* 2020, 584, 425–429.

[7] Gudbjartsson, D.F., Helgason, A., Jonsson, H., Magnusson, O.T., et al., Spread of SARS-CoV-2 in the Icelandic Population. *N. Engl. J. Med.* 2020, 382, 2302–2315.

[8] García, Z., Taylor, L., Ruano, A., Pavón, L., et al., Evaluation of a pooling method for routine anti-HCV screening of blood donors to lower the cost burden on blood banks in countries under development. *J. Med. Virol.* 1996, 49, 218–222.

[9] Cardoso, M.S., Koerner, K., Kubanek, B., Mini-pool screening by nucleic acid testing for hepatitis B virus, hepatitis C virus, and HIV: Preliminary results. *Transfusion* 1998, 38, 905–907.

[10] Kacena, K.A., Quinn, S.B., Howell, M.R., Madico, G.E., et al., Pooling urine samples for ligase chain reaction screening for genital Chlamydia trachomatis infection in asymptomatic women. *J. Clin. Microbiol.* 1998, 36, 481–485.

[11] Kacena, K.A., Quinn, S.B., Hartman, S.C., Quinn, T.C., Gaydos, C.A., Pooling of urine samples for screening for Neisseria gonorrhoeae by ligase chain reaction: Accuracy and application. *J. Clin. Microbiol.* 1998, 36, 3624–3628.

[12] Lefrère, J.J., Coste, J., Defer, C., Mercier, B., et al., Screening blood donations for viral genomes: Multicenter study of real- time simulation using pooled samples on the model of hepatitis C virus RNA detection. *Transfusion* 1998, 38, 915–923.

[13] Morandi, P.A., Schockmel, G.A., Yerly, S., Burgisser, P., et al., Detection of human immunodeficiency virus type 1 (HIV-1) RNA in pools of sera negative for antibodies to HIV-1 and HIV-2. *J. Clin. Microbiol.* 1998, 36, 1534–1538.

[14] Roth, W.K., Weber, M., Seifried, E., Feasibility and efficacy of routine PCR screening of blood donations for hepatitis C virus, hepatitis B virus, and HIV-1 in a blood-bank setting. *Lancet* 1999, 353, 359–363.

[15] Yerly, S., Pedrocchi, M., Perrin, L., The use of polymerase chain reaction in plasma pools for the concomitant detection of hepatitis C virus and HIV type 1 RNA. *Transfusion* 38, 908–914.

[16] Chong, B.S.W., Tran, T., Druce, J., Ballard, S.A., et al., Sample pooling is a viable strategy for SARS-CoV-2 detection in low-prevalence settings. *Pathology* 2020, 52, 796–800.

[17] Cao, S., Gan, Y., Wang, C., Bachmann, M., et al., Post-lockdown SARS-CoV-2 nucleic acid screening in nearly ten million residents of Wuhan, China. *Nat. Commun.* 2020, 11, 5917.

[18] Freire-Paspuel, B., Vega-Mariño, P., Velez, A., Cruz, M., Garcia-Bereguiain, M.A., “Sample pooling of RNA extracts to speed up SARS-CoV-2 diagnosis using CDC FDA EUA RT-qPCR kit.” *Virus Res.* 2020, 290, 198173.

[19] Lohse, S., Pfuhl, T., Berkó-Göttel, B., Rissland, J., et al., Pooling of samples for testing for SARS-CoV-2 in asymptomatic people. *Lancet Infect. Dis.* 2020, 20, 1231–1232.

[20] Nyazika, T.K., Kaela, R., Mugoni, M., Musomekwa, K., et al., Implementation of Antibody Rapid Diagnostic Testing versus Real-Time Reverse Transcription-PCR Sample Pooling in the Screening of COVID-19: a Case of Different Testing Strategies in Africa. *mSphere* 2020, 5.

[21] Gupta, E., Padhi, A., Khodare, A., Agarwal, R., et al., Pooled RNA sample reverse transcriptase real time PCR assay for SARS CoV-2 infection: A reliable, faster and economical method. *PLoS One* 2020, 15, e0236859.

[22] Yelin, I., Aharony, N., Tamar, E.S., Argoetti, A., et al., Evaluation of COVID-19 RT-qPCR Test in Multi sample Pools. *Clin. Infect. Dis.* 2020, 71, 2073–2078.

[23] Lim, K.L., Johari, N.A., Wong, S.T., Khaw, L.T., et al., A novel strategy for community screening of SARS-CoV-2 (COVID-19): Sample pooling method. *PLoS One* 2020, 15.

[24] Abid, S., Ferjani, S., El Moussi, A., Ferjani, A., et al., Assessment of sample pooling for SARS-CoV-2 molecular testing for screening of asymptomatic persons in Tunisia. *Diagn. Microbiol. Infect. Dis.* 2020, 98, 115125.

[25] Taylor, L., Uruguay is winning against covid-19. This is how. *BMJ* 2020, 370.

[26] Hogan, C.A., Sahoo, M.K., Pinsky, B.A., Sample Pooling as a Strategy to Detect Community Transmission of SARS-CoV-2. *JAMA* 2020, 323, 1967–1969.

[27] Abdalhamid, B., Bilder, C.R., McCutchen, E.L., Hinrichs, S.H., et al., Assessment of Specimen Pooling to Conserve SARS CoV-2 Testing Resources. *Am. J. Clin. Pathol.* 2020, 153, 715–718.

[28] Ben-Ami, R., Klochendler, A., Seidel, M., Sido, T., et al., Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. *Clin. Microbiol. Infect.* 2020, 26, 1248–1253.

[29] Ambrosis, N., Martin Aispuro, P., Belhart, K., Bottero, D., et al., Active Surveillance of Asymptomatic, Presymptomatic, and Oligosymptomatic SARS-CoV-2-Infected Individuals in Communities Inhabiting Closed or Semi-closed Institutions. *Front. Med.* 2021, 8.

[30] Dorfman, R., The Detection of Defective Members of Large Populations. *Ann. Math. Stat.* 1943, 14, 436–440.

[31] Wolf, J.K., Born Again Group Testing: Multiaccess Communications. *IEEE Trans. Inf. Theory* 1985, 31, 185–191.

[32] Du, D.-Z., Hwang, F.K., *Pooling Designs and Nonadaptive Group Testing*, vol. Volume 18, WORLD SCIENTIFIC, 2006.

[33] Aldridge, M., Johnson, O., Scarlett, J., Group testing: An information theory perspective. *Found. Trends Commun. Inf. Theory* 2019, 15, 196–392.

[34] Armendáriz, I., Ferrari, P.A., Fraiman, D., Dawson, S.P., Group testing with nested pools. *arXiv* 2020.

[35] Eberhardt, J.N., Breuckmann, N.P., Eberhardt, C.S., Multi-Stage Group Testing Improves Efficiency of Large-Scale COVID-19 Screening. *J. Clin. Virol.* 2020, 128, 104382.

[36] Mutesa, L., Ndishimye, P., Butera, Y., Souopgui, J., et al., A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature* 2020.

[37] Ghosh, S., Rajwade, A., Krishna, S., Gopalkrishnan, N., et al., Tapestry: A Single-Round Smart Pooling Technique for COVID-19 Testing. *medRxiv* 2020, 2020.04.23.20077727.

[38] Shental, N., Levy, S., Wuvshet, V., Skorniakov, S., et al., Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers. *Sci. Adv.* 2020, 6, 5961–5972.

[39] de Salazar, A., Aguilera, A., Trastoy, R., Fuentes, A., et al., Sample pooling for SARS-CoV-2 RT-PCR screening. *Clin. Microbiol. Infect.* 2020, 26, 1687.e1-1687.e5.

[40] Sunjaya, A.F., Sunjaya, A.P., Pooled Testing for Expanding COVID-19 Mass Surveillance. *Disaster Med. Public Health Prep.* 2020, 14, e42–e43.

[41] Garg, J., Singh, V., Pandey, P., Verma, A., et al., Evaluation of sample pooling for diagnosis of COVID-19 by real time-PCR: A resource-saving combat strategy. *J. Med. Virol.* 2020, n/a.

[42] Regen, F., Eren, N., Heuser, I., Hellmann-Regen, J., A simple approach to optimum pool size for pooled SARS-CoV-2 testing. *Int. J. Infect. Dis.* 2020, 100, 324–326.

[43] Lu, J., Gu, J., Gu, J., Li, K., et al., COVID-19 Outbreak Associated with Air Conditioning in Restaurant, Guangzhou, China, 2020. *Emerg. Infect. Dis.* 2020, 26, 1628–1631.

[44] Milton, D.K., A Rosetta Stone for Understanding Infectious Drops and Aerosols. *J. Pediatric Infect. Dis. Soc.* 2020, 9, 413–415.

[45] Barak, N., Ben-Ami, R., Sido, T., Perri, A., et al., Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *Sci. Transl. Med.* 2021, eabf2823.

[46] Lee, S., Kim, T., Lee, E., Lee, C., et al., Clinical Course and Molecular Viral Shedding among Asymptomatic and Symptomatic Patients with SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. *JAMA Intern. Med.* 2020, 180, 1447–1452.

[47] Day, M., Covid-19: four fifths of cases are asymptomatic, China figures indicate. *BMJ* 2020, 369, m1375.

Bios

Felipe Marceca

Obtained his degree (2016) and his PhD (2021) in Mathematics from the School of Exact and Natural Sciences at the University of Buenos Aires (FCEN-UBA), Argentina. Currently, he has a postdoctoral position at the University of Vienna, Austria, where he is conducting research in the fields of functional and harmonic analysis.

Luciana Rocha-Viegas

Obtained her degree in Biology (2000) and her PhD in Biological Chemistry (2005) from the School of Exact and Natural Sciences at the University of Buenos Aires (FCEN-UBA), Argentina. Currently, she is Associate Investigator of CONICET at the Laboratory of Nuclear Receptors and Chromatin at the Institute of Physiology, Molecular Biology and Neurosciences (IFIBYNE), CONICET and teaching assistant of molecular and cellular biology at FCEN-UBA.

Nicolás Pregi

Obtained his degree in Biology (2001, University of Buenos Aires), and has a PhD from the University of Buenos Aires (2007). Nowadays, he works as a researcher at CONICET in the field of neuroepigenetics of stress at the IQuiBiCEN (CONICET, Argentina), and as a professor at the National University of San Martín (Buenos Aires, Argentina).

María Gabriela Barbás

Obtained her degree in Biochemistry (1999) and her specialization in virology (2003) from the Chemistry school at the National University of Córdoba (FCQ-UNC), Argentina. She is currently, secretary of prevention and health promotion at the Health Ministry of Córdoba, Argentina. Also she was Head of the Central Laboratory of Córdoba, Argentina (since 2007 until 2019); teacher at the chair of microbiology of the Catholic University of Córdoba (since 2014 until 2019); Management consultant of AIDS and STD of the National Health Ministry of Argentina (since 2014 until 2019).

Daniela Hozbor

Obtained her degree in Biochemistry (1987) and her PhD in Biochemical Sciences (1993) from La Plata National University (UNLP), Argentina. Currently, she is Senior Investigator of CONICET, Head of the Laboratory of Vaccines (VacSal) at the Institute of Biotechnology and Molecular Biology (IBBM), UNLP-CONICET and Full Professor of the Biotechnology and Molecular Biology Area of the Faculty of Sciences. UNLP. She is in charge of the Pertussis Reference Laboratory and is a member of the steering committee of the Global Pertussis Initiative.

Adali Pecci

Obtained her degree in Chemistry (1988) and her PhD in Biological Chemistry (1997) from the School of Exact and Natural Sciences at the University of Buenos Aires (FCEN-UBA), Argentina. Currently, she is Senior Investigator of CONICET, Head of the Laboratory of Nuclear Receptors and Chromatin at the Institute of Physiology, Molecular Biology and Neurosciences (IFIBYNE), CONICET and Associated Professor at Biological Chemistry Department, FCEN-UBA. She is Vice-Director of the IFIBYNE and Academic Officer of FCEN-UBA.

Roberto Etchenique

Obtained his degree in Chemistry (1994) and his PhD in Inorganic, Analytical and Physical Chemistry (1998) from the School of Exact and Natural Sciences at the University of Buenos Aires (FCEN-UBA), Argentina. Currently, he is Senior Investigator of CONICET, Head of the Molecular Devices Lab at the Institute of Physical Chemistry of Materials, Environment and Energy (INQUIMAE), CONICET and Full Professor at the Inorganic, Analytical and Physical Chemistry Department, FCEN-UBA.

INSTRUCTIONS FOR AUTHORS

GUIDELINES

1. Length

Research reviews should not be less than 7000 words or more than 10,000 words (plus up to 15 figures and up to 50 references).

2. Structure

Whenever possible articles should adjust to the standard structure comprising:
(a) Graphical abstract,
(b) Abstract,
(c) Introduction describing the focus of the review,
(d) Article main body including assessment and discussion of available information (may be further subdivided),
(e) Conclusions,
(f) Bibliography.

3. Format

Authors must submit their articles in a Microsoft Word archive. Figures must be embedded in the article and also submitted in a separate .zip or .rar file.

4. References

References must be numbered in the text ([1] [2] [3]) and identified with the same numbers in the *References* section. Up to 50 references are allowed. The following are examples to take into account in each case.

BOOKS:
- *Single author*

Chung, R. *General Chemistry: Fundamental Knowledge,* 2nd ed.; McGuffin-Hill: Kansas City, 2003.
- *More than one author*

Chung, R.; Williamson, M. *General Chemistry: Fundamental Knowledge*, 2nd ed.; McGuffin-Hill: Kansas City, 2003.
- *Edited Book*

Kurti, F. Photodissociation and Reactive Scattering. In *The Rise of Chemical Physics*; White, AD, Ed.; Wilson: New Jersey, 2007; Vol. 128; p. 257.
- *Book in Series*

Goth, V. Polymer Chemistry. In *The Foundational Course in Organic Chemistry*; ACDC Symposium Series 1151; American Chemical Fraternity: Seattle, 2014; pp 123-149.
- *Article from a reference book*

Powder and Metallurgy. *Dictionary of Chemical Technology*, 3rd ed.; Wilson: New Jersey, 1971; Vol. 12, pp 68-82.

ARTICLES:
- *Article in a scientific journal*

Evans, A.; Stitch, M.; Smithers, ET; Nope, JJ Complex Aldol Reactions to the Total Synthesis of

Phorboxazole B. *J. Am. Chem.* Soc. 2012, 122, 10033-10046.
- *Article in a popular/non-scientific magazine*
Tatum, CJ Super Organics. *Wireless* , June 2001, pp 76-93.
- *Article from an online journal*
Turkey-Lopez, E. Inexact Solutions of the Quantum Double Square-Well Potential. *Chem. Ed.* [Online] 2007, 11, pp. 838-847. http://chemeducator.org/bibs/0011006/11060380lb.htm (accessed Aug 5, 2019).

PUBLICATION ETHICS

The journal considers that the primary objective of all submissions must be a contribution of relevant and appropriate content, and that all review processes must be structured based on that general criterion. Therefore, there is an emphasis on the concern to maintain the highest quality and ethics standards in the reception, evaluation and publication of articles. These standards include the three participants of the process: author, reviewer and editor.

1. Author's responsibilities

- Submitted manuscripts should maintain rigorous scientific criteria for data validation and conclusions.
- All data (Figures, Tables, etc.) reproduced from previous published articles must give the appropriate recognition to the source. Plagiarism is cause enough to reject the submission.
- Authorship must include all individuals who have contributed in a substantial way to the composition, prior investigation, and execution of the paper. Minor contributions must be acknowledged, but these contributors should not be listed as authors. The main author or authors of the article will make sure that all participants of the paper have approved the final version of the document submitted.
- All authors must reveal in their final manuscript any financial or other type of conflict of interests that might interfere with the results and interpretations in their research. All funding received to carry out the project must be acknowledged.
- After the article is published, in the event an author notices a crucial fault or inaccuracy, he or she should immediately report that fault or inaccuracy, so that an Erratum can be issued as soon as possible.

2. Reviewer's responsibilities

Reviewing is a time-consuming process that is carried out *ad honorem* by *bona fide* scientists conversant with the subject of the reviewed paper. The quality and the ethical standards of the journal depend critically on the quality of the reviewing process, and the following guidelines are established:

- All documents sent to the journal for review will be considered confidential documents and will not be discussed with external third parties.
- When invited, a potential reviewer should decline if: (a) the subject of the article is not within his/her area of expertise; (b) there is any kind of conflict of interest; (c) if the review cannot be finished within the period established by the journal.
- Any criticism or objections to the paper should be done in a neutral tone and based on reasonable grounds, not limited to simple opinions or purely subjective expressions.

3. Editor's responsibilities

- The editors are responsible for selecting the papers that will be published in the journal. The Editorial Management must comply with the ethical standards of the journal, as well as with all legal guidelines, including the prohibition of plagiarism and any other form of copyright infringement.
- The editors will evaluate and make decisions on the articles sent to the journal regardless of the

gender, sexual orientation, religious beliefs, ethnic origin, nationality or political ideology of the authors.
- Revealing information identifying reviewers is forbidden.
- The final version of all materials can be published only with the prior approval of their author.
- The editors will refrain from publishing manuscripts that imply a conflict of interests because of any possible connection with other institutions, companies and authors.
- Before deciding to send an article to a peer review, the editors are committed to thoroughly read all texts received and determine their appropriateness to the thematic universe of the journal.
- If a misbehavior or unethical action by an author or reviewer is identified, the editors must request the informer of such conduct or action to provide the evidence that may justify a possible investigation. All accusations will be handled seriously until reliable results are obtained regarding its truthfulness or falseness. If an investigation takes place, the editors are responsible for choosing the appropriate way in which it will be carried out. They can also request the advice and assistance from the Editorial Board, as well as from reviewers and authors.
- In the event a serious non-malicious mistake or a dishonest conduct by an author or a reviewer is proved, the editors shall act according to the nature and seriousness of the case. The actions the editors may take include, but are not limited to: notifying the author or reviewer of the existence of a serious mistake or misapplication of the ethical standards of the journal; writing a strong statement that reports and warns about a bad practice or unethical behavior; publishing that statement; unilaterally withdrawing the reported paper from the review or publication process; revoking the paper if it has already been published; communicating the journal's decision and the reasons behind it to the general public; and banning paper submissions by the people involved for a certain period of time.

PRIVACY STATEMENT

The names and email addresses entered in this journal site will be used exclusively for the stated purposes of this journal and will not be made available for any other purpose or to any other party.

Science Reviews
from the end of the world

ISSN 2683-9288

Centro de Estudios sobre Ciencia, Desarrollo
y Educación Superior
538 Pueyrredón Av. - 2° C – Second building
Buenos Aires, Argentina - C1032ABS
(54 11) 4963-7878/8811
sciencereviews@centroredes.org.ar
www.scirevfew.net