Optimising the assignment of swabs and reagent for PCR testing during a viral epidemic

Alberto Santini

1Universitat Pompeu Fabra, Barcelona, Spain — alberto.santini@upf.edu

August 29, 2020

Abstract

Early large-scale swab testing is a fundamental tool for health authorities to assess the prevalence of a virus and enact appropriate mitigation measures during an epidemic. The COVID-19 pandemic has shown that the availability of chemical reagent required to carry out the tests is often a bottleneck in increasing a country’s testing capacity. Further, demand is unevenly spread between more affected regions (which require more tests they can perform) and less affected ones (which have spare capacity). These issues hint at the opportunity of increasing test capacity via the optimal allocation of swabs and reagent to laboratories. We prove that this is the case, proposing an Integer Programming formulation to maximise the number of tests a country can perform and validating our approach on both real-life data from Italy and synthetic instances. Our results show that increased inter-regional collaboration and a steadier supply of reagent (i.e., coming from local production sites rather than international shipments) can dramatically increase testing capacity. Accordingly, we propose short-term and long-term recommendations for policy makers and health authorities.

Keywords: production planning; healthcare optimization; covid-19.

1 Introduction

Viruses are pathogens that replicate after penetrating the living cells of other organisms. Outside cells, a virus is nothing more than genetic material (DNA or RNA) surrounded by protective layers of proteins and lipids. Once inside a host cell, the virus uses the cell’s structures to replicate its genetic material and assemble its protective layers, thus creating a copy of itself. At the end, it kills the cell to release both the original and the copy.

Viruses cause a host of human diseases, ranging from the common cold to AIDS. Of particular interest for this work is the recent Coronavirus disease 2019 (COVID-19) pandemic. Under particular circumstances, a virus (and the disease it causes) can spread to a large part of a population in a short time, leading to an epidemic. When the epidemic spreads across national borders and infects people worldwide, it is termed a pandemic. SARS-CoV-2, the virus spreading COVID-19, for example, possessed the right characteristics to turn its associated disease into a pandemic: people carrying the virus do not show symptoms for an average of five days, during which they can spread the virus to others [44]; its basic reproduction number has been estimated between 1.4 and 5.7, i.e., each infected person in turns infects an average of up to 5.7 people [33]; up to 44% of patients show no symptoms at all during the infection period, making their diagnosis difficult [29].

Massive testing of the population is a fundamental part of the response against epidemics and pandemics. When carriers of the virus remain asymptomatic while infecting others, as for COVID-19, testing is one of the main tools health authorities can deploy to contain the spread of the disease [37].

Viruses such as SARS-CoV-2 are RNA viruses, i.e., they contain a single strand of nucleotides rather than the “double helix” typical of DNA. For this brief introduction, we call the four nucleotides composing the RNA chain with their initials: A, C, G, U. The quickest available test for RNA viruses uses real-time reverse transcription polymerase chain reaction (rRT-PCR). A popular test consists in collecting a sample of nasal secretions on a swab [9], looking for an RNA subsequence which is unique to the virus. To identify this
subsequence, the tester uses inverse RNA: a sequence of nucleotides which pairs those they are looking for (remember that A pairs with U and C with G). If the sample contains viral RNA, it will pair with the chain injected by the tester. Once the nucleotides are paired, they effectively transform the single-strand RNA into double-strand DNA. To detect a significant presence of the viral DNA, the genetic material goes through an amplification process, the polymerase chain reaction [8]. Both transforming RNA into DNA, and amplifying it, cannot happen without certain enzymes, commonly called reagents. While this simplified description is enough to introduce our problem, we refer the reader to the reviews of Bustin [4] and Freeman, Walker, and Vrana [16] for a more accurate description of the rRT-PCR technique. Figure 1 is an example of a protocol for rRT-PCR tests used by the United Kingdom’s National Health Service.

Figure 1: United Kingdom National Health Service’s protocol for rRT-PCR analysis of samples from swabs, to detect COVID-19.

During the COVID-19 pandemic, laboratory capacities and reagent availability have become the bottleneck for increasing the number of tests in much of Europe, North America and Asia [13, 1, 43, 27]. Therefore, health authorities need to optimise the allocation of resources to their test laboratories, starting from the distribution of reagent and the assignments of testing tasks. With this work, we use tools from Operational Research to help decision makers maximise the number of tests they can conduct while limited by reagent availability and logistic constraints.

Our main conclusions are the following.

- We introduce the problem of maximising a country’s test capacity for rRT-PCR tests, during the outbreak of an epidemic. In such a situation, (i) reagent scarcity is often a bottleneck for testing, and (ii) some regions are foci of the disease, requiring more tests than others. These two conditions provide an opportunity to optimise the allocation of reagent and swabs, with the aim of maximising the number of swabs tested and reducing the wait time for results.

- We provide an Integer Programming formulation for this problem, extend it to consider the real-world case in which the health-care system is decentralised at the regional level, and consider a hierarchical multi-objective version of the problem. Because commercial software such as Gurobi solves the formulation in a few minutes, this model is suitable for direct use by public health decision makers and to run what-if analyses.
We conduct two case studies to validate our approach. In the first one, we use real-life data referring to the Italian situation during the COVID-19 pandemic. Our analysis confirms that fostering inter-regional collaboration can increase by up to +40% the number of swabs tested during a 13-day planning horizon. In the second one, we use synthetic data to determine the most important factors which impact testing capacity. We conclude that a strong industrial network able to provide a steady supply of reagent provides the greatest advantage, vs. procuring reagent via sporadic shipments, even when the total purchased quantity stays the same. Further, countries with limited lab infrastructure should prioritise acquiring more machines and training more personnel, while increasing inter-regional coordination helps those countries with a developed laboratory network.

We make available both our data and the code we used, under an open-source license [35].

2 Problem description

In our problem, a set of laboratories $L = \{1, \ldots, |L|\}$ has to perform rRT-PCR tests on swabs during a time horizon $T = \{1, \ldots, |T|\}$ (each time unit corresponds to a day). We assume, without loss of generality, that for each swab a lab needs one unit of reagent. This allows to consider different testing routines, such as group testing, by rescaling the amount of reagent needed to test one swab. Each lab $l \in L$ starts with a reserve $p_{l0}$ of reagent at the beginning of the time horizon. The labs can receive further units of reagent from factories $R = \{1, \ldots, |R|\}$, limited by their production capacity. Each factory $r \in R$ produces $f_{rt}$ units of reagent on day $t \in T$. Factories also store an initial amount $p_{r0}$ of reagent at the beginning of the planning horizon. A factory can also model other types of facilities; for example, a warehouse receiving new reagent once per week, or even a one-time shipment from a foreign country.

Each laboratory $l$ has a capacity $Q_l$ of swabs it can test during one day. This capacity applies even if the lab has a larger amount of reagent, due to limitations on available machinery and workforce. The effective testing capacity of $l$, then, is the minimum between $Q_l$ and the amount of reagent available at $l$.

A lab $l \in L$ is tasked with testing $m_{lt}$ swabs on day $t$, according to a predefined schedule to meet epidemiological needs. If a laboratory does not have enough reagent to test all the swabs, the decision-maker has three options: (i) moving some reagent from a factory to the lab, (ii) moving some swabs to another lab, or (iii) storing the swabs and schedule their testing for another day.

Remark. We assume the planner has already chosen the laboratory to which he assigns the swabs. More generally, though, a planner might need that swabs be tested in a given region, without any constraint on which specific lab performs the test, as long as they are geographically close to the point where the swabs were collected. In Section 5.1, we will extend the model to take into account this general case, making the number of swabs assigned to each lab a decision variable. We also refer the reader to Section 5.2, in which we introduce a fourth possibility for the decision maker: to move some reagent from a lab which has a surplus to one which is lacking.

The amount of reagent and swabs that a planner can move between locations each day is bounded by $q^{rg}$ and $q^{sw}$, respectively.

Remark. We assume that quantities $q^{rg}$ and $q^{sw}$ are global capacity limits on the total amount of reagent and swab movements. Such an assumption rests on the observation that, during an epidemic, a central decision-maker manages logistic resources and can reallocate them from one area to another. If this were not the case, it would not be hard to enforce local limits, e.g., on the number of swabs that the planner can move out of a single laboratory. In a model extensions presented in Section 5.1, locations are partitioned in regions and we impose region-specific capacity limits.

Given this input data, a planner must determine, for each day of the planning horizon, (i) how many units of reagent to move from the factories to the laboratories, and (ii) how many swabs to move between laboratories, with the aim of maximising the number of tests carried out.

Remark. While the main objective of the planner is to maximise the number of swabs tested, it is possible to account for secondary objectives. For example, because effective testing must be both large-scale and quick, one might want to minimise the average time swabs spend waiting at lab facilities. In Section 5.3 we propose a multi-
objective extension of the problem in which, among all solutions which maximise the number of tests carried out, we select the one with lowest average waiting time.

3 Related work

Our problem, described in Section 2, resembles superficially problems in the area of production planning, material requirement planning and supply chain optimisation (for a complete overview on these topics see, e.g., Pochet and Wolsey [30]). In our case, carrying out tests corresponds to manufacturing a product using two raw materials (swabs and reagent); for us, though, the finished product demand (tested swabs) always matches one of the raw materials’ availability (swabs), while the other (reagent) acts as a bottleneck for production. As in production planning, we deal with the flow of goods over time, including shipments of raw material and inventory carry-over. We also allow the creation of backlogs, i.e., the storage of swabs which cannot be tested on the day of collection. Other characteristics of our problem, however, are not frequently considered in production planning optimisation: the geographic position of suppliers and production facilities (in our case, laboratories) and their regional subdivision; the possibility of reassigning orders between facilities, subject to logistics constraints; the uneven demand between regions.

Most importantly, unlike problems in production planning, we disregard costs and focus exclusively on maximising production (in our case, number of tests). To this end, we do not consider transport, production, storage, raw material, order placement, personnel, nor opportunity costs, which instead have a prominent role when optimising commercial supply chains and production processes. In the standard case, in fact, the aim of the planner is pure cost minimisation [30] and satisfying the demand in full could even be sub-optimal, if the cost of producing one more unit is larger than the opportunity cost of missing a sale. Other approaches commonly used in production planning include multi-objective optimisation in which cost minimisation is tackled together with secondary objectives, such as demand satisfaction [26], steadiness of production rates [40] or environmental performance [2]. Even in the healthcare sector, if considering times when authorities are not responding to an emergency, costs are taken into account. For example, Zhao et al. [48] consider a drugs production planning problem in which they minimise both costs (avoiding over-production of drugs) and the duration of the trial. Other works (e.g., [10, 14]) have considered the production and supply optimisation problems in the clinical area, but have focused on production and distribution of drugs in “normal” times, rather than during an emergency response to a sudden epidemic event.

Literature focusing specifically on allocation of medical resources during emergencies mainly concentrates on the distribution of (life-saving) medical devices. As such, most models optimise the assignment of existing material to hospitals or individual patients, rather than the production and allocation of new material. Xiang and Zhuang [46], for example, consider the problem of assigning medical resources to patients, taking into account that the conditions of patients left untreated will deteriorate over time. Cao and Huang [5] use a simulation approach in which they consider different strategies to allocate scarce medical resources (e.g., milder patients first, critical patients first, randomly) and complement an analysis of the utility of each strategy with ethical considerations on prioritising some patients over others.

In the wake of the COVID-19 epidemics, the optimisation community has rushed to provide decision support tools for decision-makers involved in the fight against the virus. Similar to our work, researchers at MIT’s Operations Research Center [7, 6] have shown the advantages of inter-regional collaboration in sharing automatic ventilators across U.S. states. Mehrotra et al. [23] consider the same problem and use a stochastic model to show how sharing ventilators across states, even when each state is averse to the risk of not having enough machines for its own patients, can reduce the probability of shortages. The problem of allocating ventilator during influenza epidemics was already considered in the literature, e.g., by Huang et al. [19], Meltzer et al. [24], and Zaza et al. [47].

Lampariello and Sagratella [22] study a problem that, like ours, concerns diagnostic tests during the COVID-19 pandemics. The authors consider the problem of deciding how many swabs each region should collect to provide an accurate estimate of the prevalence of the disease among its population, while restricted by a nation-wide capacity limit on the number of tests performed. The problem is motivated by the practice employed in Italy during the early period of the epidemic, of double-checking at one single central laboratory
(located in Rome) the results of each positive test conducted by local labs. The capacity of the central lab limits, in this case, the number of total expected positive results coming from all other labs in the country. Note how a decision maker working under the assumptions of [22] can use the model therein to generate regional demands 𝓓ᵣ, which they can then feed to our model.

Finally, Seccia [36] considers the problem of devising balanced rosters for nurses in hospitals while, due to the virus outbreak, there is both a peak in demand for nurses’ services and fewer available health professionals, because many are COVID-19-positive and confined at home.

4 Model formulation

We propose an Integer Programming model using the following sets of variables. The number of units of reagent moved from factory ᵣ to lab ˡ on day 𝒯 is denoted by variable ₓᵣˡˡ, ∈ ℤ⁺. Variables ʸᵣˡˡ, ∈ ℤ⁺ represents the number of swabs going from laboratory ˡᵢ to ˡ₂ on day 𝒯 ∈ ℤ. The quantities of reagent stored, respectively, at factory ᵣ ∈ ℤ⁺ and ˡ ∈ ℤ⁺. Variables ᶻᵣˡ ∈ ℤ⁺ represent the number of swabs stored at lab ˡ ∈ ℤ during day 𝒯 ∈ ℤ. Laboratories store swabs if they cannot carry out the tests on day 𝒯 because of a lack of reagent necessary for the rRT-PCR procedure. We assume without loss of generality that, at the beginning of the time horizon, labs do not have any stored swab (ᶻ₀ˡ = 0 for all ˡ ∈ ℤ).

Maximising the number of tests corresponds to minimising the number of swabs stored at labs at the end of the time horizon. The objective function, then, is:

\[
\min \sum_{ˡ ∈ ℤ} z_{l[T]}. \tag{1}
\]

The bounds on the quantities of reagent and swabs moved each day translate into the following two constraints:

\[
\sum_{ᵣ ∈ ℤ} \sum_{ˡ ∈ ℤ} xᵣˡˡ ≤ q^{ RG} \forall \mathcal{T} ∈ ℤ \tag{2}
\]

\[
\sum_{ˡ ∈ ℤ} \sum_{ᵣ ∈ ℤ} yᵣˡˡ ≤ q^{ SW} \forall \mathcal{T} \in ℤ. \tag{3}
\]

The next constraints limit the number of tests carried out at each lab on a given day, based on the quantity of reagent available and lab capacities. To this end, it is convenient to introduce an auxiliary variable ʷᵣˡ ∈ ℤ⁺ representing the number of tests carried out at lab ˡ on day 𝒯. Quantities ᵐᵣˡ (number of tests requested each day), \( \sum_{ˡ ∈ ℤ} yᵣˡˡ \) (swabs moving from ˡ to other labs), \( \sum_{ˡ ∈ ℤ} yᵣˡˡ \) (swabs moving from other labs to ˡ), ᶻᵣˡˡ (swabs backlog from the previous day) and ᶻᵣˡ (swabs stored at the end of the day) uniquely determine the value of ʷᵣˡ, according to the linking relation

\[
z_{l,t−1} + m_{l,l} + \sum_{ᵣ ∈ ℤ} yᵣˡˡ = z_{l,l} + w_{l,l} + \sum_{ᵣ ∈ ℤ} yᵣˡˡ \forall ℤ, \forall ℤT. \tag{4}
\]

Figure 2 shows this relation visually. In the figure, each node represents lab ˡ on a different day and arrows display the flow of swabs in and out of the laboratory.

With linking variables ʷᵣˡ, we can constrain the number of tests performed with the following inequalities:

\[
wᵣˡ ≤ rᵣ,l−1 + \sum_{ᵣ ∈ ℤ} xᵣˡˡ \quad \forall ℤ, \forall ℤT \tag{5}
\]

\[
wᵣˡ ≤ qᵣ \quad \forall ℤ, \forall ℤT. \tag{6}
\]

Next, we consider flow balance equations analogous to eq. (4), relative to the number of reagent stored at factories and laboratories:

\[
克服 rᵣ−₁ + fᵣ = ρᵣ + \sum_{ˡ ∈ ℤ} xᵣˡˡ \quad \forall ℤ, \forall ℤT \tag{7}
\]

\[
克服 l−₁ + \sum_{ᵣ ∈ ℤ} xᵣˡˡ = ρᵣ + wᵣˡ \quad \forall ℤ, \forall ℤT. \tag{8}
\]
Figure 2: Linking relation between parameter $m$ and variables $y$, $z$ and $w$ for a lab $l \in L$ at time intervals from $t - 1$ to $t + 1$.

Constraint (7) states that the quantity of reagent in the factory’s inventory at the beginning of the day, plus the quantity produced, is equal to the quantity in inventory at the end of the day, plus the quantity shipped to the labs. Constraint (8) equates, on the left-hand side, the quantity of reagent in the lab’s inventory at the start of the day and the total amount received with, on the right-hand side, the inventory at the end of the day, plus the quantity of reagent used (or, which is the same, the number of tests performed). Because all variables are non-negative, constraint (8) makes constraint (5) redundant.

Model (1)–(8), together with the following variable domain definitions, is the Base Model for our problem.

\begin{align}
 x_{rt} &\in \mathbb{Z}^+ & \forall r \in R, \forall l \in L, \forall t \in T \\
y_{lt} &\in \mathbb{Z}^+ & \forall l \in L, \forall l' \in L \setminus \{l\}, \forall t \in T \\
\rho_{rt} &\in \mathbb{Z}^+ & \forall r \in R, \forall t \in T \\
\rho_{lt}, z_{lt}, w_{lt} &\in \mathbb{Z}^+ & \forall l \in L, \forall t \in T 
\end{align}

4.1 Model strengthening

A disadvantage of the Base Model (1)–(12) is that it suffers from symmetry. For example, consider any solution with slack capacity to transport swabs between two labs $l_1, l_2 \in L$ during a day $t \in T$. One can get another solution of the same cost in which, on day $t$, $l_1$ sends one more swab to $l_2$ and $l_2$ sends one more swab to $l_1$. To reduce symmetry, we impose two restrictions on the transport of swabs between labs: (i) a lab $l$ can ship out swabs only if it is already working at full capacity, i.e., it is testing $Q_l$ swabs or it ran out of reagent; (ii) a lab can only send or receive swabs during a given day, but not both.

To tackle the first restriction, we introduce indicator variables $\gamma_{lt} \in \{0, 1\}$, which take value 1 iff lab $l \in L$ is running at full capacity during day $t \in T$. These variables link to variables $w$ and $\rho$ via the following indicator constraints (see, e.g., [3]):

\begin{align}
 \gamma_{lt} = 0 &\rightarrow w_{lt} \leq Q_l - 1 & \forall l \in L, \forall t \in T \\
 \gamma_{lt} = 0 &\rightarrow \rho_{lt} \geq 1 & \forall l \in L, \forall t \in T \\
 \gamma_{lt} = 1 &\rightarrow (w_{lt} = Q_l) \lor (\rho_{lt} = 0) & \forall l \in L, \forall t \in T. 
\end{align}

The restriction is, then, enforced with another indicator constraint:

\begin{align}
 \gamma_{lt} = 0 &\rightarrow \sum_{l' \in L \setminus \{l\}} y_{l'l} = 0 & \forall l \in L, t \in T. 
\end{align}

Constraint (16) guarantees that laboratory $l$ cannot ship out swabs unless it is at full capacity.
To address the second restriction, we can add two more sets of indicator variables and constraints, keeping track of whether a lab sends out or receives swabs. Note that the black-box commercial solver we use (Gurobi) converts all indicator constraints presented in this section into linear constraints. The new variables are \( y^l_{lt}, y^-_{lt} \in \{0, 1\} \), which assume value 1 iff, respectively, lab \( l \) sends or receives swabs on day \( t \). The necessary constraints to link these variables and enforce the restriction are:

\[
\begin{align*}
y^+_l &= \sum_{t \in T \atop t \neq l} y_{lt} \geq 1, \quad y^-_l &= \sum_{t \in T \atop t \neq l} y_{lt} = 0 \quad \forall l \in L, \forall t \in T \quad (17) \\
y^-_l &= \sum_{t \in T \atop t \neq l} y_{lt} \geq 1, \quad y^+_l &= \sum_{t \in T \atop t \neq l} y_{lt} = 0 \quad \forall l \in L, \forall t \in T \quad (18) \\
y^+_{lt} + y^-_{lt} &\leq 1 \quad \forall l \in L, \forall t \in T. \quad (19)
\end{align*}
\]

Constraint (19) enforces that each lab, on each day, cannot both send and receive swabs. This restriction also avoids that a laboratory works as a relay, receiving and instantly re-shipping (i.e., “teleporting”) swabs between otherwise distant labs.

The solutions produced by the Base Model can have another property which is undesirable in real-life applications: consider two laboratories \( l_1, l_2 \in L \) and two reagent factories \( r_1, r_2 \in R \) such that \( l_1 \) is closer to \( r_1 \) than to \( r_2 \) and \( l_2 \) is closer to \( r_2 \) than to \( r_1 \). The model allows \( l_1 \) to buy reagent from the more distant factory \( r_2 \) while \( l_2 \) buys reagent from \( r_1 \). To avoid this counter-intuitive situation which represents a sub-optimal use of logistic resources, we introduce parameter \( d_{lt} \in \mathbb{R}^+ \) representing the distance between lab \( l \in L \) and factory \( r \in R \) and add the following constraint to the model:

\[
(x_{r_1l_1t} = 0) \lor (x_{r_1l_2t} = 0) \quad \forall t \in T, \forall l_1, l_2 \in L, \forall r_1, r_2 \in R: \ d_{l_1r_1} < d_{l_1r_2} \quad \text{and} \quad d_{l_2r_2} < d_{l_2r_1}. \quad (20)
\]

While (20) refers to reagent, we add an analogous constraint to prevent moving swabs inefficiently:

\[
(y^+_{l_1l_1t} = 0) \lor (y^-_{l_1l_2t} = 0) \quad \forall t \in T, \forall l_1, l_1', l_2, l_2' \in L: \ d_{l_1l_1'} < d_{l_1l_2'} \quad \text{and} \quad d_{l_2l_2'} < d_{l_2l_1'}, \quad (21)
\]

where \( d_{ll'} \in \mathbb{R}^+ \) is the distance between labs \( l \in L \) and \( l' \in L \).

5 Model extensions

In this section we propose extensions to the model presented above, to account for more realistic scenarios. We present an Extended Model, taking into account these considerations, in Appendix A.

5.1 Regional partitioning

In Section 2 we mentioned two limitations of the Base Model which might not hold true in real-life applications. The first limitation is that there is a fixed quantity of swabs assigned to each laboratory. In most countries, central or regional health authorities oversee and plan testing. Swabs collected in a region, for example, might be sent for testing to any laboratory in the same region. The second limitation is that tests (and reagent) can move overnight from any location to any other location subject to global capacity limits. When applying the model to large countries, though, logistic constraints might impose that movement of material only happen between geographically close locations.

To address these aspects, we extend the initial problem formulation as follows. The set of laboratories is partitioned as \( L = L_1 \cup \ldots \cup L_n \) (with \( L_i \cap L_j = \emptyset \) for any two \( i, j \in \{1, \ldots, n\}, \ i \neq j \)). We call each set composing the partition a region.

The planner receives as input the number of swabs to test each day in each region \( L_i \), denoted as \( m_{it} \), but must determine how to split the swabs between the region’s labs. A parameter \( \delta_{lt} \in \{0, 1\} \) determines whether a factory \( r \in R \) can ship reagent overnight to laboratory \( l \in L \). Analogously, parameter \( \mu_{lt} \in \{0, 1\} \) determines whether two labs \( l_1, l_2 \in L \) can send each other swabs overnight.
In our model, logistic capacities refer to each region. We consider, respectively, quantities \( q_i^{sw} \) and \( q_i^{rg} \) of reagent and swabs that can be shipped to region \( i \in \{1, \ldots, n\} \) in a day. Depending on the logistic infrastructure, a planner could specify bounds at different levels of aggregation and for both inbound and outbound movements. For example, one might consider a maximum quantity of reagent shipped out of a factory, or a group of factories. Although we present here regional-level inbound capacities, our formulation is flexible enough to allow a wide array of modelling options.

5.2 Transshipments of reagent

A limitation of the Base Model is that it allows to move swabs but not reagent between laboratories. This reflects the conservative stance of laboratories, which are often unwilling to part with a scarce resource even if they have it in surplus, out of fear of running out later and not seeing their “favour” returned.

Without capacities, moving swabs and reagent would be equivalent. For example, a lab we are indirectly introducing distance-based lead times in the reagent supply chain. At the same time, it uses some units of reagent to perform tests and eventually sends some to other labs; the

**Base Model**

A limitation of the Base Model is that it allows to move swabs but not reagent between laboratories. This reflects the conservative stance of laboratories, which are often unwilling to part with a scarce resource even if they have it in surplus, out of fear of running out later and not seeing their “favour” returned.

Without capacities, moving swabs and reagent would be equivalent. For example, a lab \( l_1 \) could send a swab it cannot test to another lab \( l_2 \), because \( l_1 \) does not have enough reagent to test it; or, which is the same, \( l_2 \) could send one unit of reagent to \( l_1 \), which could test their swab by themselves. In the presence of capacities determined by equipment and personnel availability, though, the two options are not interchangeable. In the example above, \( l_2 \) might not have the capacity to test one more swab, while \( l_1 \) could have it if provided with the necessary reagent.

In light of the above considerations, we propose an extension of the model in which reagent can be forwarded between labs: a practice which goes by the name of transshipment. A planner can also use transshipments as a way to extend the radius at which factories send their products, by first storing them at a lab to which the factory ships and, when the need arises, forwarding them to other labs — even if these labs were not originally reachable from the factory with an overnight shipment. Because reagent take one day to move between labs, we are indirectly introducing distance-based lead times in the reagent supply chain.

To account for transshipment of reagent, we extend the model introducing a new set of variables \( x_{l,l,t} \in \mathbb{Z}^+ \) indicating the amount of reagent moved from lab \( l_1 \in L \) to lab \( l_2 \in L \) on day \( t \in T \). We assume that transshipments happen between laboratories which are close enough (\( \mu_{l_1, l_2} = 1 \)) and that a unit sent on day \( t \) is available on day \( t + 1 \).

\[
\begin{align*}
\sum_{r \in R} x_{r,l,t-1} + \sum_{l' \in L, l' \neq l} x_{l',l,t-1} & = \rho_{l,t-2} \\
\sum_{r \in R} x_{r,l,t} + \sum_{l' \in L, l' \neq l} x_{l',l,t-1} & = \rho_{l,t-1} \\
\sum_{r \in R} x_{r,l,t+1} + \sum_{l' \in L, l' \neq l} x_{l',l,t} & = \rho_{l,t+1}
\end{align*}
\]

\[
\begin{align*}
\mu_{l,t-1} & = \sum_{l' \in L, l' \neq l} x_{l',l,t-1} \\
\mu_{l,t} & = \sum_{l' \in L, l' \neq l} x_{l',l,t} \\
\mu_{l,t+1} & = \sum_{l' \in L, l' \neq l} x_{l',l,t+1}
\end{align*}
\]

**Figure 3:** Flow of reagent for a lab \( l \in L \) at time intervals from \( t-1 \) to \( t+1 \), when accounting for transshipments.

**Figure 3** shows the flow of reagent for a lab \( l \in L \) when accounting for transshipments. Each day, the lab has in its availability the reagent already stored there, plus that received from factories and from other laboratories. At the same time, it uses some units of reagent to perform tests and eventually sends some to other labs; the remaining units go to the inventory for next day. In the model, we replace constraint (8) with the following:

\[
\rho_{l,t-1} + \sum_{r \in R} x_{r,l,t} + \sum_{l' \in L, l' \neq l} x_{l',l,t-1} = \rho_{l,t} + \mu_{l,t} + \sum_{l' \in L, l' \neq l} x_{l',l,t} \quad \forall l \in L, \forall t \in T,
\]  

(22)
where $x_{p,l,t-1} = 0$ if $t = 0$, and constraint (2) with:

$$
\sum_{r \in R} \sum_{l \in L} x_{rlt} + \sum_{r \notin L} x_{rlt} \leq q^g \quad \forall t \in T. \tag{23}
$$

Lastly, we add a constraint analogous to (20) and (21) to ensure that reagent move efficiently between labs:

$$(x_{l',l,t} = 0) \lor (x_{l,t} = 0) \quad \forall t \in T, \forall l, l', l_2 \in L : d_{l,l_1} < d_{l,l_1'} \text{ and } d_{l,l_2} < d_{l,l_2'}. \tag{24}
$$

### 5.3 Multi-objective model

As discussed in Section 2, the number of tests carried out is not the only important parameter for an effective testing campaign. Having fast results also helps assessing the disease’s spread and enact appropriate and well-timed confinement and mitigation measures.

We propose, then, a Multi-Objective Optimisation (MOO) approach to the problem. For a recent review of MOO applications in optimisation, we refer the reader to Gunantara [17] and Kalyanmoy [20]. In the following, we consider the hierarchical objective approach, in which the planner first optimises for their main objective (number of tests performed) and then, among all the solutions which give an optimal objective value, looks for the one optimising a secondary objective (mean test waiting time).

Let $Z^*$ be the optimal objective value of model (27)–(53). We can obtain the solution to the hierarchical problem by solving the following Hierarchical Model:

$$
\min \sum_{l \in L} \sum_{t \in T} z_{lt} \tag{25}
$$

s.t. \hspace{1cm} \sum_{t \in T} z_{l[T]} = Z^* \tag{26}

where eq. (25) minimises the sum of all amounts of swabs stored at labs waiting to be tested and eq. (26) imposes that the considered solutions match the optimal value with respect to objective function (27). Objective function (25) is equivalent to minimising the average swab waiting time because (26) fixes the total number of swabs tested. When minimising the primary objective, two untested swabs collected, respectively, on the first and last day of the time horizon would contribute the same penalty. Adding the secondary objective avoids solution where old swabs remain untested for a long time, while newer ones get tested quickly.

A planner could also add further objectives. For example, if he wants to guarantee a more equitable distribution of swabs to laboratories, he might want to minimise their average workload. To this end, we introduce a new variable $W \in \mathbb{R}^+$ such that $W \geq \frac{1}{Q_l} \sum_{t \in T} w_{lt}$ for all labs $l \in L$. Minimising $W$ corresponds to minimising the highest average workload across all labs. Alternatively, if $W \geq \frac{w_{lt}}{Q_l}$ for all $l \in L$ and $t \in T$, then minimising $W$ amounts to minimising the peak workload over all labs and all days of the planning horizon.

### 6 Computational analysis

We want to assess the impact of optimising the allocation of reagent and swabs to laboratories on the testing capacity of a country. We present two case studies. In the first, we perform an analysis on a realistic dataset relative to the Italian response to the COVID-19 pandemic, during the period from April 1st to April 13th, 2020. In the second, we use synthetic data to do a sensitivity analysis highlighting the decisions a planner can take to most increase testing capacity and lower waiting times. The datasets and the code used are available at Santini [35]. We use the Hierarchical Model, which we solve using commercial solver Gurobi [18] with a time limit of 15 minutes for the primary and 5 minutes for the secondary objective, on a laptop equipped with a 4-core Intel i7 processor running at 1.6GHz. This time limit proved enough to solve to optimality the majority of instances, while those that remain open all show gaps smaller than 5% and, often, smaller than 1%.
6.1 Italy, April 1–13, 2020

We create a set of instances based on data sources relative to the COVID-19 pandemic in Italy [32, 25, 28, 38, 12]. We use data for the period 01–13 April, 2020 for the entire national territory. Figure 4a shows the distribution of laboratories and reagent factories in Italy [25]: dots represent labs which are officially authorised to execute COVID-19 swab tests by the Italian Health Ministry, while stars represent factories whose testing kits (including reagent, extraction kits, probes, negative controls) have been certified for use in COVID-19 swab testing. Figure 4b reports the number of swabs analysed in Italy during the period 01–13 April, 2020, on a regional basis. We note that there is no publicly available data on the average wait times between swab collection and analysis.

(a) Map of rRT-PCR laboratories (dots) and reagent factories (stars).

(b) Number of swabs analysed per region, over the period from April, 1 to April, 13, 2020.

Figure 4: Regional-level data for Italy.

We analyse the number of swabs tested, day by day and region by region, using five different scenarios. The first uses the official numbers released by Italy’s “Dipartimento della Protezione Civile”, which we refer to as Real Data. Next, we consider the output obtained by the Hierarchical Model, under the following hypotheses. Swabs collected in a region cannot be sent to other regions, but can be allocated to any laboratory within the collection region (for this reason we label this data as Model - Regional). In modelling terms, this corresponds to setting $\mu_{l_1, l_2} = 1$ if and only if, for two labs $l_1, l_2 \in L$, there is a region $L_k$ such that $l_1, l_2 \in L_k$. We use this hypothesis to model how testing happens in Italy, with local laboratories and regional “reference laboratories” collecting swabs from all-over the region and providing double checks before non-urgent hospital admissions (emergency patients are admitted even before test results are ready) and discharges. Also remind that the Italian healthcare system is managed on a regional basis, with ample autonomy given to the local health authorities but little inter-regional interaction [15]. We assume that labs can procure reagent from their geographically closest factory, and that procurement is optimised by a planner deciding the quantities of reagent sent to each laboratory, to maximise the number of tested swabs. This assumption reflects how Italian regions were procuring supplies autonomously, before the central government created and rendered operational a COVID-19 Commissariat which also acted as a national procurement centre [11, Art. 122].

Because news sources and press releases by regional health authorities (we found statements by authorities of Basilicata, Campania, Emilia-Romagna, Marche, Lombardia, Piemonte, Puglia, Sardegna, Sicilia, Toscana, Umbria and Veneto, i.e., 12 out of the 20 Italian regions) hint that reagent was a major bottleneck in expanding test capabilities, we expect this model to be able to increase the number of swabs tested even if it does not allow inter-regional swab reassignment, just by optimising reagent assignment. Figure 5a shows how laboratories cluster within each region: an edge between two laboratories means that they belong to the same region.
(a) Swabs can only be transported between labs in the same region.

(b) Laboratories in the same region or within 200Km can send swabs.

Figure 5: Map of testing laboratories in Italy. An edge connects two laboratories if they can send swabs between each other overnight.

Finally, we consider the output obtained from the Hierarchical Model, but allowing for reallocation of swabs between laboratories in different regions. This assumption effectively means that regional daily demand can be shared across regions. Because swabs would need to move within a short time (e.g., overnight) we put limitations on the inter-laboratory distance that allows swab transfer. We consider three thresholds of 100Km, 200Km and 400Km and we denote the corresponding data as Model - 100Km, Model - 200Km and Model - 400Km. (We make an exception for the island region of Sardinia, for which moving swabs outside of the region would be unfeasible even if there are other laboratories within 400Km.) Figure 5b shows how laboratories can transfer swabs when using the 200Km threshold. Compare Figures 5a and 5b to note how inter-regional collaboration gives a central planner more opportunities to move swabs between regions in case there should be a day with a demand peak in a particular area. We refer the reader to [35] for the full description of the instance generation process and to [34] for an interactive dashboard presenting the results of the analysis.

Figure 6 shows the number of swabs tested, day-by-day, when using our proposed models Regional, 100Km, 200Km, and 400Km. Note how Model - 400Km fully uses the system’s capacity during the last period of the time horizon, where the number of tested swabs becomes a flat line. During some days, a model with fewer lab transfer capabilities gives more tested swabs than a model in which more transfers are allowed. For example, on April 6, using Model - 400Km results in fewer tests than Model - 200Km and Model - 100Km. This is not surprising if we consider that the models maximise the number of swabs tested over the whole planning horizon, even if this results in fewer tests on any particular day.

Table 1 reports detailed results of the analysis. Each row corresponds to one day, and the columns list the real data and the results obtained using the model. Columns #tests report the number of swabs tested during that day. Columns %gain are the percentage gains obtained using the models, compared to the number of swabs actually tested in Italy in the described period. The last row gives aggregate results for the whole country. Note how using optimisation techniques results in a testing capacity increase between 14% and 39%, under the hypotheses explained above. In particular, assuming that tests in Italy were largely limited by the unavailability of reagent due to a poor allocation of supplies and low inter-regional collaboration, optimising these two aspects can allow to perform up to 250 000 more tests (the increase between Real Data and Model - 400Km) during 13 days.

Figure 7 shows example results from the Regional and 400Km models, for four representative Italian regions.
Figure 6: Number of swab tests performed on the national territory, per day. The dashed line reports official numbers from the Civil Defence agency. The other lines represent the results from applying the Hierarchical Model with intra-regional (orange column) and inter-regional (green, pink and blue lines) swab transfers. The inter-regional transfer radii considered are 100, 200 and 400Km.

Table 1: Number of swabs tested on the national territory, day by day, from official real data and from the results obtained using our model. Columns #tests represent the number of swabs tested each day. Columns %gain are the percentage gains obtained using the model, compared to the real data.
Figure 7: Number of swab tests performed in four Italian regions. The dashed line reports official numbers from the Civil Defence agency. The solid lines represents results obtained with our Hierarchical Model allowing intra-regional (orange line) and inter-regional (blue line) swab transfers.

from the north (Piemonte and Emilia-Romagna), the centre (Lazio) and the south (Sicilia) of the country. Each chart reports the number of swabs analysed per day. Note how, for three of the four regions, the results of the Regional model agree with the real data, validating our modelling assumptions and hinting that regions managed resources locally efficiently. In all four regions, during the last period of the planning horizon, the models use the full laboratory capacity and their curves appear flat, i.e., the limiting factor is not reagent availability anymore, but structural limitations such as testing machines or lab personnel. Note how, in Sicily during the middle period, the 400Km model provides a solution with fewer tested swabs than the Regional model and the Real Data, for two days. However, there was a catch-up on the tests during the last part of the planning horizon, resulting in more tests carried out in total.

6.2 Synthetic data

By conducting a scenario analysis on synthetic data, we want to learn how some aspects of the healthcare system of the area hit by an epidemic can affect its ability to perform large-scale testing of the population. The factors we analyse are:

1. The effects of having a healthcare system with a large level of independent decision-making at the sub-national level (as is the case, e.g., of Italian regions, Spanish autonomous communities, German lander or the United States) and the benefits of increased cross-regional collaboration, common planning, and resource sharing.

2. The capillarity and production capacity of bio-chemical and pharmaceutical industries able to prepare the reagent needed to carry out rRT-PCR tests, especially in light of the increasing trend to off-shore such manufacturing capabilities to developing countries (see, e.g., [21, 42]).

3. Reagent-independent laboratory capacity, i.e., the test capacity deriving by the number of available machines and personnel. This corresponds to the full laboratory capacity in the ideal case in which any amount of reagent were immediately available.
4. The logistic capabilities of the country, i.e., how labs can procure reagent from manufacturing facilities and transfer swabs away from saturated labs.

6.2.1 Instance generation

We list in the following the characteristics of the instances we generate. We consider $|L| = 100$ laboratories, which we distribute in the square $S = [0, 100]^2$ in the euclidean plane as explained below. First, we define parameter $\lambda_{LR}$, which is the expected size of regions, i.e., the expected number of labs in each region. Next, we select $n = \lfloor |L|/\lambda_{LR} \rfloor$ random points in $[10, 90]^2$ which we use as initial centroids for the regions (note how we excluded the external boundary of $S$). We place the labs in $S$ by first choosing a random centroid and then a random point in $S$ which lies at distance of at most 20 from the centroid, and repeating this procedure $|L|$ times. In this way, we aim at placing clusters in loose regional clusters. To complete the assignment of labs to regions, we assign each lab to its closest centroid.

(a) Spatial set-up of the instance. Dots represent laboratories, with their size proportional to the average daily demand in their region. Stars represent factories and lines are regional boundaries.

(b) Daily reagent production. Each smaller bar represents one factory and the larger bars represent the total daily production.

Figure 8: Example of a synthetic instance with bumpy reagent production at factories.

Another parameter, $\lambda_{FrR}$, determines how distributed is the country’s reagent production capacity; this parameters represents the expected number of reagent factories per region. We place $|R| = n \cdot \lambda_{FrR}$ factories at random in $S$. Figure 8a shows an example distribution of labs and factories. Dots represent laboratories, stars denote factories, and lines are the Voronoi boundaries (see, e.g., [31, Ch. 5]) associated with regional centroids (i.e., the lines delimit the regions).

To determine the demand of each region, we assume that 40% of the regions will have a criticality, such as a focus of the illness, and will thus require more tests. (For example, during the COVID-19 pandemics, the Chinese province of Wuhan, northern Italian regions, the autonomous communities of La Rioja, Madrid and Castilla-La Mancha in Spain, or the states of Washington and New York in the U.S.A. were foci in their respective countries.) The other regions might, instead, have some spare capacity which they can use, if resources are properly coordinated, to help saturated regions. Let $\bar{m} = 100$ be an ideal number of tests to perform per lab per day (assuming ideal conditions), and let $|L_i|$ be the number of labs in region $i$. The number of tests requested in region $i$ on day $t$ is

$$m_{it} = \begin{cases} 
\bar{m} \cdot |L_i| \cdot 2 \cdot \epsilon & \text{for } 40\% \text{ randomly chosen regions} \\
\bar{m} \cdot |L_i| \cdot 0.75 \cdot \epsilon & \text{for the remaining } 60\% \text{ of regions},
\end{cases}$$

14
We assume here that the “impervious” regions have tighter logistical constraints; for example, islands, mountainous areas, or regions with poorer infrastructure. We set the initial amount of reagent stored at labs and factories to a minimal amount, as follows:

\[ \rho_{t_0} = \bar{m} \cdot \epsilon \quad \rho_{t_0} = \bar{m} \cdot \frac{\lambda_{LXR}}{\lambda_{FXR}} \cdot \epsilon, \]

where \( \epsilon \) is a noise term distributed according to a truncated normal distribution between 0.95 and 1.05. The definition of \( m_t \) is justified by the assumption that the number of swabs to test depends on regional population, which correlates positively with the number of laboratories in the region. The dot size in Figure 8a is directly proportional to the lab’s region average demand in the example instance; note how two bordering regions are “critical”.

Lab capacities are then defined as a fraction of the ideal number of tests assigned to each lab: \( Q_i = \bar{m} \lambda_{LC} \), where \( \lambda_{LC} \) is a parameter. Values of the parameter larger than 1 correspond to a situation in which reagent availability is the main bottleneck for testing, while smaller values indicate that laboratories have structural deficiencies which limit the number of swabs they can test, even when reagent is abundant. Each lab can procure reagent from its \( \lambda_{LF} \) closest factories, and can send swabs to other laboratories in the same region. A parameter \( \lambda_{LL} \) gives the maximum distance at which two labs in different regions can exchange swabs.

Regional inbound capacities for reagent (\( q_{ir}^{ri} \)) and for swabs (\( q_{is}^{sw} \)) are determined as:

\[
q_{ir}^{ri} = \begin{cases} 
\bar{m} \cdot |L_i| & \text{for } 80\% \text{ of the regions} \\
\bar{m} \cdot |L_i| \cdot 0.75 & \text{for } 20\% \text{ “impervious” regions}
\end{cases}
\]

\[
q_{is}^{sw} = \begin{cases} 
\bar{m} \cdot |L_i| \cdot 0.25 & \text{for } 80\% \text{ of the regions} \\
\bar{m} \cdot |L_i| \cdot 0.10 & \text{for } 20\% \text{ “impervious” regions}
\end{cases}
\]

We assume here that the “impervious” regions have tighter logistical constraints; for example, islands, mountainous areas, or regions with poorer infrastructure. We set the initial amount of reagent stored at labs and factories to a minimal amount, as follows:

\[ \rho_{t_0} = \bar{m} \cdot \epsilon \quad \rho_{t_0} = \bar{m} \cdot \frac{\lambda_{LXR}}{\lambda_{FXR}} \cdot \epsilon, \]

where \( \epsilon \) is distributed according to a truncated normal distribution between 0 and 0.25. Such a choice corresponds to labs and factories having a small fraction of their daily, respectively needed and produced, reagent in the initial stock — a situation that can commonly occur during an epidemic, when safety stocks are quickly depleted.

Next, we describe how we determine factories’ daily production. We propose two production patterns (selected via parameter \( \lambda_{FP} \)): steady and bumpy. Under the steady pattern, each factory \( r \) produces \( f_{r,t} = \bar{m} \cdot \frac{\lambda_{FP}}{\lambda_{FXR}} \cdot \epsilon \) units of reagent each day, where \( \epsilon \) follows a truncated normal distribution between 0.95 and 1.05 and \( \lambda_{FP} \) is a parameter. Note how \( \lambda_{FP} < 1 \) drives the total factory production capacity to be smaller than the total demand, indicating a situation in which there is less reagent than the nation needs. Values of \( \lambda_{FP} \geq 1 \) drive the total reagent production towards meeting the demand; in this case, tests not performed because of lack of reagent are more likely due to poor resource allocation over the national territory, rather than an absolute scarcity of reagent. Under the bumpy pattern, for each factory we first choose two random days of the week \( r_t^1, r_t^2 \in \{0, \ldots, 6\} \), which we call “release days”. The factory will produce reagent at the same

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Num values</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>T</td>
<td>)</td>
<td>Time horizon length</td>
</tr>
<tr>
<td>(\lambda_{LXR})</td>
<td>Labs per region</td>
<td>3</td>
<td>5, 10, 20</td>
</tr>
<tr>
<td>(\lambda_{FXR})</td>
<td>Factories per region</td>
<td>4</td>
<td>0.1, 0.25, 0.5, 1</td>
</tr>
<tr>
<td>(\lambda_{LC})</td>
<td>Lab capacity multiplier</td>
<td>7</td>
<td>0.5, 0.7, 0.9, 1, 1.1, 1.3, 1.5</td>
</tr>
<tr>
<td>(\lambda_{LF})</td>
<td>Number of factories supplying each lab</td>
<td>3</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>(\lambda_{LL})</td>
<td>Radius for inter-regional swab transfers</td>
<td>6</td>
<td>0, 5, 10, 15, 20, 25</td>
</tr>
<tr>
<td>(\lambda_{FP})</td>
<td>Factory production multiplier</td>
<td>5</td>
<td>0.8, 0.9, 1, 1.1, 1.2</td>
</tr>
<tr>
<td>(\lambda_{PP})</td>
<td>Factory production pattern</td>
<td>2</td>
<td>steady, bumpy</td>
</tr>
<tr>
<td>(\lambda_{TS})</td>
<td>Transshipments</td>
<td>2</td>
<td>yes, no</td>
</tr>
</tbody>
</table>

Table 2: Synthetic instance generation parameters.
We report the results of our scenario analysis, highlighting how the parameters affect the testing capacity
waiting to be tested are, respectively, the dependent variables in the two analyses.

While in the following we comment on the insights obtained.

The coloured bars correspond to the quantities released by the factories on their release days, while grey back-
ground bars count the total amount of reagent released on that day, by all factories. The figure shows that,
even if factories have, in principle, a steady production rate (only fluctuating between −5% and +5% due to
the noise term \( \xi \)), making new reagent available twice per week produces an uneven total daily distribution.

Last, we introduce transshipment parameter \( \lambda_{TS} \), which takes values yes or no. This parameter does not affect
instance generation; rather, it enables or disables the transshipment of reagent between laboratories in the
model. If transshipments are not allowed, it forces all variables \( x_{l \ell'} \) to take value zero.

Table 2 summarises all instance generation parameters we used. Because in our scenario analysis we vary
each parameter independently, we ran the Hierarchical Model over the full grid of 60 480 instances. We ran
the model twice for each instance, once enabling and once disabling transshipments.

### 6.2.2 Scenario analysis

We report the results of our scenario analysis, highlighting how the parameters affect the testing capacity
and commenting on which real-life decisions a planner can take to increase the number of swabs tested. A
better testing capacity depends on both structural factors (such as a strong reagent industrial base) which
require long-term decisions, and organisational factors (such as more inter-regional collaboration) which are
influenced by operational decisions.

Figure 9 shows how the percentage of swabs tested within the end of the time horizon changes, when varying
each parameter. Each chart corresponds to one parameter; we use line charts for the seven numeric param-
eters, and box plots for \( \lambda_{PP} \) and \( \lambda_{TS} \), which are categorical. The values reported in each chart refer to the
average over all instances generated or solved with the given parameter values.

Two linear regression analyses provide additional findings. They use the instance parameters as independent
variables. We centre and standardise numeric parameters, and transform \( \lambda_{PP} \) and \( \lambda_{TS} \) into numeric parameters,
respectively, with value 0 if the production was steady and 1 if bumpy, and 0 when transshipments are not
allowed and 1 when they are. The percentage of swabs tested and the average time a swab spent in a lab
waiting to be tested are, respectively, the dependent variables in the two analyses. Table 3 reports the results,
while in the following we comment on the insights obtained.

<table>
<thead>
<tr>
<th>ind. variable</th>
<th>% swabs tested coef</th>
<th>p-value</th>
<th>swab wait time coef</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>42.98</td>
<td>&lt; 10⁻³</td>
<td>0.36</td>
<td>&lt; 10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>[T]</td>
<td>0.64</td>
<td>&lt; 10⁻³</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{LR} )</td>
<td>-0.94</td>
<td>&lt; 10⁻³</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{FX} )</td>
<td>3.65</td>
<td>&lt; 10⁻³</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{LC} )</td>
<td>5.26</td>
<td>&lt; 10⁻³</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{LF} )</td>
<td>3.00</td>
<td>&lt; 10⁻³</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{LL} )</td>
<td>0.50</td>
<td>&lt; 10⁻³</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{FP} )</td>
<td>0.63</td>
<td>&lt; 10⁻³</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{PP} )</td>
<td>-12.86</td>
<td>&lt; 10⁻³</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>( \lambda_{TS} )</td>
<td>1.21</td>
<td>&lt; 10⁻³</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 3: Results of a linear regression analysis using the parameters as independent variables and the primary
(number of swabs tested, expressed as a percentage of the total demand) and secondary (average waiting time
of a swab before it is tested) objectives as dependent variables. Columns coef report the linear regression
coefficients, after centring and standardising the inputs. All independent variables resulted significant, with
p-values always smaller than 0.001.
Figure 9: Variation in the percentage of swabs tested (y axis), when varying each parameter (x axis). In each chart, the line shows the average over all instances generated with the given parameter value and the shaded area around the line is the 95% confidence interval. For parameters $\lambda_{PP}$ and $\lambda_{TS}$, as they take categorical values, we show a box plot instead of a line plot. The central line indicates the median, the box spans the two central quartiles, while whiskers extend to the rest of the distribution (excluding outliers).
Steadiness of reagent supply. Parameter $\lambda_{pp}$ is the one with the largest impact in our analysis: a steady availability of reagent allows the laboratories to operate without interruptions, testing more swabs and with shorter delays. The average percentage of swabs tested when $\lambda_{pp} =$ steady is double than when $\lambda_{pp} =$ bumpy. Recall that a factory in our model can represent a real production site, but also a warehouse or a shipment from abroad. A physical factory will be able to maintain a steadier production rate, while relying on shipments will resemble more the “bumpy” scenario. The analysis, thus, hints to the fact that a solid industrial sector and a well-function supply chain can have the greatest impact on testing capacity; not necessarily because they allow to procure more reagent in total, but because they provide a steadier flow of it. A country that cannot count on such production capacity can mitigate the adverse effects of irregular supply with larger orders (see the positive impact of parameter $\lambda_{pp}$), more inventory capacity and a quick dispatch of reagent to labs.

Capillarity of the reagent industry. Having more small reagent production sites works better than having few, larger ones (parameter $\lambda_{FR}$). When $\lambda_{FR}$ moves from 0.1 to 1.0, the average percentage of swabs tested climbs from 40.83% to 51.46%. An efficient supply chain which is able to move material over longer distances (parameter $\lambda_{LF}$) also has a positive impact on testing capacity. When each lab can procure from its closest factory ($\lambda_{LF} = 1$), the analysis shows that 41.36% of swabs are tested. This number increases to 47.86% and then to 49.61%, when $\lambda_{LF}$ increases to, respectively, 2 and 3.

Figure 10: Variation in the percentage of swabs tested ($y$ axis), when varying parameter $\lambda_{LL}$ ($x$ axis). The chart on the left refers to instances with $\lambda_{LC} < 1$, while that on the right, to instances with $\lambda_{LC} \geq 1$.

Lab capacity and inter-regional capacity sharing. Other two factors which impact the amount of swabs tested are lab capacity and the degree to which regions can share this capacity by transferring swabs between each other. Parameter $\lambda_{LC}$ clearly influences testing capabilities: for $\lambda_{LC} < 1$ largely less than half of the swabs are tested within the end of the time horizon. When $\lambda_{LC} \geq 1$, i.e., when the total lab capacity is at least as large as the average demand, gains are marginal, as exhibited by the curve flattening in Figure 9. It is at this point that we would expect capacity sharing via increased inter-regional collaboration to have the greatest effect. The extent at which capacity is shared depends on parameter $\lambda_{LL}$. It might appear, looking at Figure 9, that the impact of $\lambda_{LL}$ is small. However, if we divide the analysis of its impact between instances with $\lambda_{LC} < 1$ and instances with $\lambda_{LC} \geq 1$, the situation is different, as reported in Figure 10. Note how the curve referring to instances with $\lambda_{LC} \geq 1$ is much steeper, corresponding to a larger percentage increase in swabs tested. We can conclude that countries with insufficient lab infrastructures should primarily focus on procuring test machinery and training personnel to carry out rRT-PCR tests. Countries which already have a developed test lab network, on the other hand, benefit more from inter-regional collaboration and centralised planning than they do from increasing lab capacity.
Other factors. Allowing transshipments produces an increase in swabs tested, from 43.30% to 45.85% on average (an increase of +5.03%), placing parameter $\lambda_{TS}$’s impact behind several others. We suspect that the reason is that there are few cases in which moving reagent in one direction gives different results compared to moving swabs in the opposite direction, as mentioned in Section 5.2. Transshipments’ effect is larger if we focus on instances with bumpy reagent production. In this case they increase the number of swabs tested by +11.14%. This result suggests that transshipments can smooth irregular production patterns: a country which cannot procure reagent steadily should increase internal collaboration by moving both swabs and reagent. The percentage of swabs tested decreases when there are fewer, larger regions (parameter $\lambda_{LR}$). This is because, with fewer regions, the impact of one region being critical, i.e., with larger demands $m_{it}$, is proportionally larger. The effect of region size is rather modest, as indicated by a small regression coefficient associated with $\lambda_{LR}$. Regarding the planning horizon length $|T|$, its impact is unclear. On one hand, a longer time horizon allows more opportunities for optimisation. On the other hand, there is a reduction in tested swabs when moving from $|T| = 5$ to $|T| = 7$, which is hard to explain. In practice, a planner would want to balance the advantages usually conferred by longer-term planning with the quality of the demand predictions that he can obtain from epidemiological models. Finally, while the signs of the regression coefficients in the two regression analyses tend to coincide (meaning that parameters which have a positive impact on the number of swabs tested also have a positive impact on shortening wait times), the signs disagree for parameters $\lambda_{LR}$ and $|T|$, making it difficult to derive actionable recommendations for decision makers based on these two parameters.

<table>
<thead>
<tr>
<th>% swabs moved</th>
<th>avg % swabs tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0, 5]</td>
<td>38.38</td>
</tr>
<tr>
<td>(5, 10]</td>
<td>36.24</td>
</tr>
<tr>
<td>(10, 15]</td>
<td>43.34</td>
</tr>
<tr>
<td>(15, 20]</td>
<td>56.59</td>
</tr>
<tr>
<td>(20, 25]</td>
<td>67.34</td>
</tr>
<tr>
<td>(25, 30]</td>
<td>64.72</td>
</tr>
<tr>
<td>(30, 35]</td>
<td>84.97</td>
</tr>
</tbody>
</table>

Table 4: Relation between the percentage of swabs moved between laboratories (column “% swabs moved”) over all swabs collected, and the success of the testing campaign as indicated by the percentage of swabs tested over all swabs collected (column “avg % swabs tested”). The second column contains averages over all values of the first column belonging to each interval.

To summarise the impact of resource sharing, Table 4 shows how the quantity of swabs tested changes depending on the quantity of swabs moved. Indicator “% swabs moved” is the percentage of swabs sent from one lab to another, over the total number of swabs collected. The table shows the correlation between swabs sharing and the effectiveness of the test campaign, independently from how swab sharing was achieved. In our simulations swab sharing depends on instance generation parameters, but in real life it can depend on actions taken by the planner.

7 Conclusions

We have proposed a model for the problem of assigning reagent and swabs to laboratories in a country which wants to maximise the number of rRT-PCR tests during a viral epidemic. After describing an Integer Programming model for a basic version of the problem, we have strengthened and extended the formulation to include real-life aspects of the organisation of healthcare systems. The complete model takes into account that many countries manage healthcare at a regional level and that labs can function as reagent transshipment points, and allows to optimise two hierarchical objectives: primarily, maximising the number of tests performed and, secondarily, minimising the average time a swab waits before it is tested.

We performed two computational studies: the first based on real-life data from Italy and the second on synthetic instances. Under the assumptions we used, the main one being that reagent availability is often a bottleneck for testing more swabs (see, e.g., [13, 1, 43]), the analysis has revealed two main findings:
Increased inter-regional coordination has the potential of increasing the testing capabilities of the country as a whole, as shown by the analysis of Italian data. In particular, moving swabs from critical regions to regions with spare capacity produced an increase of up to 40% in the number of swabs tested in the Italian scenario. The study performed on synthetic data confirmed this finding, but found a more modest effect. The recommendation, in fact, is that countries which do not have adequate lab infrastructure (i.e., which would have insufficient test capacity even if fully stocked with reagent) first focus on expanding their lab network, e.g., by purchasing new machines and training more lab personnel. Countries with an advanced laboratory infrastructure, instead, will benefit more from inter-regional collaboration.

Steady availability of reagent has a large impact on the number of swabs tested and their waiting time. A solid chemical industrial base, with many (although possibly small) factories producing reagent daily gives an advantage compared with having to rely on a discontinuous supply, e.g., when purchasing from international markets. If the supply is not steady, then we recommend to allow labs to work as small warehouses by letting them forward reagent to other labs, i.e., enabling transshipments.

On the short term, increasing inter-regional collaboration requires political will and an efficient supply chain which can be used to move swabs and reagent. On the other hand, increasing reagent production capacity involves longer-term investments and balancing the economic interests of pharmaceutical companies with national interests.

Future research on optimising the delivery of mass-scale test programmes can develop integrating other test types, such as immunological tests, i.e., tests which detect the presence of antibodies which a person has developed in response to the viral infection. These tests are too slow for early detection of positive patients; e.g., T and B cells response to COVID-19 were detected around one week after the onset of symptoms [39], while swab tests can identify the virus even before symptoms appear. However, because antibodies persist in the blood even after a patient does not host the virus any longer, health authorities use immunological tests for mass-screening of populations (including asymptomatic patients) with the aim of assessing the prevalence of the virus, the level of population immunity and other epidemiological dynamics. It would be interesting to study not only how to increase capacity for immunological tests, but also to study the problem of selecting the patients to test with higher priority, in order to create an accurate epidemiological picture of the population while using a restricted supply of antibody tests such as ELISA (enzyme-linked immunosorbent assay) [41] or neutralization assays [45].

Acknowledgements

The author would like to thank Michele Iacono of Roche Diagnostics S.p.A. and Xavier Jiménez Fàbrega of the Health Department of Catalonia for comments and fruitful discussions on this work. The author would also like to thank two anonymous reviewers for suggesting important improvements to this paper.

Funding: the author was partially supported by grant “RTI2018-095197-B-I00” from the Spanish Ministry of Economy and Competitiveness.

Declarations of interest: none.

References


A Extended Model formulation

The extended model, which incorporates the considerations made in Section 5, uses a new variable $u_{lt} \in \mathbb{Z}^+$, which represents the number of new swabs assigned to lab $l \in L$ on day $t \in T$. The model reads as follows:

\begin{align}
\min & \quad \sum_{l \in L} z_{l|T} \\
\text{s.t.} & \quad \sum_{r \in R} \sum_{l \in L} x_{rlt} + \sum_{l \in L} x_{l't} \leq q^R \\ & \quad \sum_{l_i \in L} y_{l_i,t} \leq q^w_i \\ & \quad x_{rlt} = 0 \\ & \quad x_{l't} = 0 \\ & \quad y_{l_i,t} = 0 \\ & \quad z_{lt-1} + u_{lt} + \sum_{l' \in L} y_{l't} = z_t + w_t + \sum_{l' \in L} y_{l't} \\ & \quad \sum_{l \in L} u_{lt} = m_{it} \\ & \quad w_{lt} \leq \rho_{l,t-1} + \sum_{r \in R} y_{rlt} \\ & \quad w_{lt} \leq Q_t \\ & \quad \rho_{r,t-1} + f_{rt} = \rho_{rt} + \sum_{l \in L} x_{rlt} \\ & \quad \rho_{l,t-1} + \sum_{r \in R} y_{rlt} + \sum_{l' \in L} x_{l',t-1} = \rho_{lt} + w_t + \sum_{l' \in L} y_{l't} \\ & \quad y_{lt} = 0 \quad \rightarrow \quad w_{lt} \leq Q_t - 1 \\ & \quad y_{lt} = 0 \quad \rightarrow \quad \rho_{lt} \geq 1 \\ & \quad y_{lt} = 1 \quad \rightarrow \quad (w_{lt} = Q_t) \lor (\rho_{lt} = 0) \\ & \quad y_{lt} = 0 \quad \rightarrow \quad \sum_{l' \in L} y_{l't} = 0 \\ & \quad y^+_t = 1 \quad \rightarrow \quad \sum_{l' \in L} y_{l't} \geq 1, \quad \gamma^+_t = 0 \quad \rightarrow \quad \sum_{l' \in L} y_{l't} = 0 
\end{align}
\[ Y_{lt}^+ + Y_{lt}^- \leq 1 \]

\[
(x_{rlt} = 0) \lor (x_{r1lt} = 0) \]

\[
(y_{rlt} = 0) \lor (y_{r1lt} = 0) \]

\[
(x_{rlt} = 0) \lor (x_{r1lt} = 0) \]

\[ x_{rlt} \in \mathbb{Z}^+ \]

\[ y_{l1lt}, x_{l1lt} \in \mathbb{Z}^+ \]

\[ \rho_{rt} \in \mathbb{Z}^+ \]

\[ \rho_{lt}, z_{lt}, w_{lt}, u_{lt} \in \mathbb{Z}^+ \]

\[ Y_{lt}, Y_{lt}' \in \{0, 1\} \]

Without loss of generality, one can linearise indicator constraints of the type \( y = 0 \rightarrow ax \leq b \), for a binary variable \( y \) and a non-negative integer variable \( x \), introducing a non-negative slack variable \( s \) and transforming inequality \( ax \leq b \) into \( ax + s \leq b \). Then the “big-M” inequality \( s \leq My \) enforces the indicator constraint.

Similarly, one can linearise constraints of type \( (x = 0) \lor (y = 0) \), where both \( x \) and \( y \) are integer an non-negative. Consider two binary variables \( s_x, s_y \) which will take value 1 if, respectively, \( x \) and \( y \) are strictly positive. Then the following three inequalities linearise the original constraint: \( x \leq Ms_x, y \leq Ms_y, s_x + s_y \leq 1 \).