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

Alberto Santini[1]

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

April 20, 2020

1 Introduction

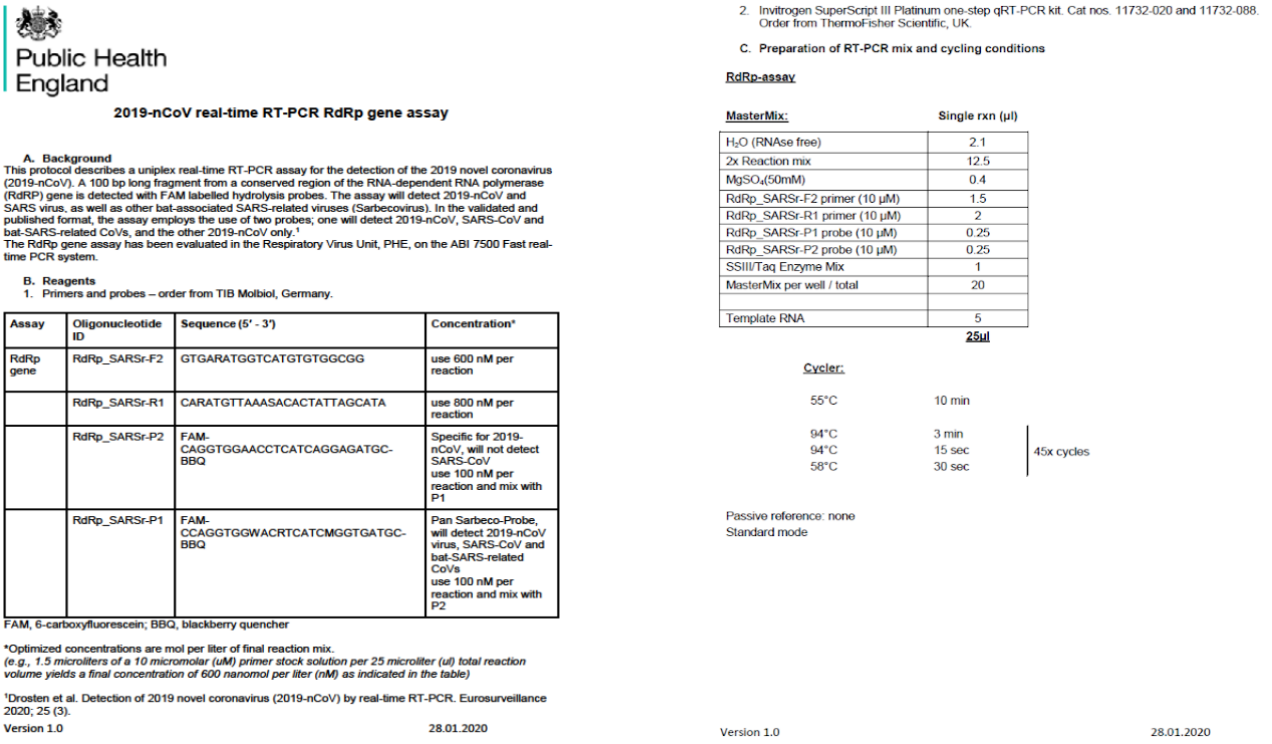
Viruses are pathogens that replicate after penetrating the living cells of other organisms. Outside cells, a virus is nothing more than genetic material (DNA or RNA) surrounded by protective layers of proteins and lipids. Once inside a host cell, the virus uses the cell's structures to replicate its genetic material and assembly 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's 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 don't show symptoms for an average of five days, during which they can spread the virus to others [27]; 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 [21]; up to 44% of patients show no symptoms at all during the infection period, making their diagnosis considerably difficult [19].

A fundamental part of the response put in place to fight epidemics and pandemics is massive testing of the population. 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 [25]. 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 [8], 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 [7]. 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 [3] and Freeman, Walker, and Vrana [11] 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.

During the COVID-19 pandemic, laboratory capacities and reagent availability have become the bottleneck for increasing the number of tests in much of Europe and North America [9, 1]. Therefore, health authorities need



Public Health England

2019-nCoV real-time RT-PCR RdRp gene assay

A. Background

This protocol describes a uniplex real-time RT-PCR assay for the detection of the 2019 novel coronavirus (2019-nCoV). A 100 bp long fragment from a conserved region of the RNA-dependent RNA polymerase (RdRP) gene is detected with FAM labelled hydrolysis probes. The assay will detect 2019-nCoV and SARS virus, as well as other bat-associated SARS-related viruses (Sarbecovirus). In the validated and published format, the assay employs the use of two probes; one will detect 2019-nCoV, SARS-CoV and bat-SARS-related CoVs, and the other 2019-nCoV only.[1]
The RdRp gene assay has been evaluated in the Respiratory Virus Unit, PHE, on the ABI 7500 Fast real-time PCR system.

B. Reagents

1. Primers and probes – order from TIB Molbiol, Germany.

Assay	Oligonucleotide ID	Sequence (5' - 3')	Concentration*
RdRp gene	RdRp_SARSr-F2	GTGARATGGTCATGTGTGGCGG	use 600 nM per reaction
	RdRp_SARSr-R1	CARATGTTAAASACACTATTAGCATA	use 800 nM per reaction
	RdRp_SARSr-P2	FAM-CAGGTGGAACCTCATCAGGAGATGC-BBQ	Specific for 2019-nCoV, will not detect SARS-CoV use 100 nM per reaction and mix with P1
	RdRp_SARSr-P1	FAM-CCAGGTGGWACRTCATCMGGTGATGC-BBQ	Pan Sarbeco-Probe, will detect 2019-nCoV virus, SARS-CoV and bat-SARS-related CoVs use 100 nM per reaction and mix with P2

FAM, 6-carboxyfluorescein; BBQ, blackberry quencher

*Optimized concentrations are mol per liter of final reaction mix.
(e.g., 1.5 microliters of a 10 micromolar (uM) primer stock solution per 25 microliter (ul) total reaction volume yields a final concentration of 600 nanomol per liter (nM) as indicated in the table)

[1]Drosten et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance 2020; 25 (3).

Version 1.0 28.01.2020

2. Invitrogen SuperScript III Platinum one-step qRT-PCR kit. Cat nos. 11732-020 and 11732-088. Order from ThermoFisher Scientific, UK.

C. Preparation of RT-PCR mix and cycling conditions

RdRp-assay

MasterMix:	Single rxn (µl)
H_2O (RNAse free)	2.1
2x Reaction mix	12.5
$MgSO_4$(50mM)	0.4
RdRp_SARSr-F2 primer (10 µM)	1.5
RdRp_SARSr-R1 primer (10 µM)	2
RdRp_SARSr-P1 probe (10 µM)	0.25
RdRp_SARSr-P2 probe (10 µM)	0.25
SSIII/Taq Enzyme Mix	1
MasterMix per well / total	20
Template RNA	5
	25ul

Cycler:

55°C	10 min	
94°C	3 min	45x cycles
94°C	15 sec	
58°C	30 sec	

Passive reference: none
Standard mode

Version 1.0 28.01.2020

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

to optimise the allocation of resources to their test laboratories, starting from the distribution of reagents 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.

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, wlog, that for each swab a lab needs one unit of reagent. Each lab $l \in L$ starts with a reserve ρ_{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. In this work we are going to assume that each factory $r \in R$ produces f_{rt} units of reagent on day $t \in T$. Factories also store an initial amount ρ_{r0} of reagent at the beginning of the planning horizon. Note that a factory can also model other types of facilities; for example, a warehouse receiving new reagents 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 doesn't 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 tests 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 they assign 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 4.1, we will extend the model to take into account this general case, making the number of swabs assigned to each lab a decision variable.*

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

Remark. *We assume that quantities q^{r} and q^{s} 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 wouldn't 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 4.1, locations are partitioned in areas and we impose area-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's 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 4.2 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 Model formulation

We propose an Integer Programming (IP) model using the following sets of variables. The number of units of reagent moved from factory $r \in R$ to lab $l \in L$ on day $t \in T$ is denoted by variable $x_{rlt} \in \mathbb{N}$. Variables $y_{l_1 l_2 t} \in \mathbb{N}$ represents the number of swabs going from laboratory $l_1 \in L$ to $l_2 \in L$ on day $t \in T$. The quantities of reagent stored, respectively, at factory $r \in R$ and lab $l \in L$ on day $t \in T$ are denoted by variables $\rho_{rt} \in \mathbb{N}$ and $\rho_{lt} \in \mathbb{N}$. Variables $z_{lt} \in \mathbb{N}$ represent the number of swabs stored at lab $l \in L$ during day $t \in T$. Laboratories store swabs if they cannot carry out the tests on day t because of a lack of reagents necessary for the rRT-PCR procedure. We assume wlog that, at the beginning of the time horizon, labs don't have any stored swab ($z_{l0} = 0$ for all $l \in L$).

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

$$\min \quad \sum_{l \in L} z_{l|T|}. \tag{1}$$

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

$$\sum_{r \in R} \sum_{l \in L} x_{rlt} \leq q^{\mathrm{r}} \qquad \forall t \in T \tag{2}$$

$$\sum_{l_1 \in L} \sum_{l_2 \in L} y_{l_1 l_2 t} \leq q^{\mathrm{s}} \qquad \forall t \in T. \tag{3}$$

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's convenient to introduce an auxiliary variable $w_{lt} \in \mathbb{N}$ representing the number of tests carried out at lab $l \in L$ on day $t \in T$. Quantities m_{lt} (number of tests requested each day), $\sum_{l' \in L, l' \neq l} y_{l'lt}$ (swabs moving from l to other labs), $\sum_{l' \in L, l' \neq l} y_{ll't}$ (swabs moving from other labs to l), $z_{l,t-1}$ (swabs backlog from the previous day) and z_{lt} (swabs stored at the end of the day) uniquely determine the value of w_{lt}, according to the linking relation

$$z_{l,t-1} + m_{lt} + \sum_{\substack{l' \in L \\ l' \neq l}} y_{l'lt} = z_{lt} + w_{lt} + \sum_{\substack{l' \in L \\ l' \neq l}} y_{ll't} \quad \forall l \in L, \ \forall t \in T. \tag{4}$$

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.

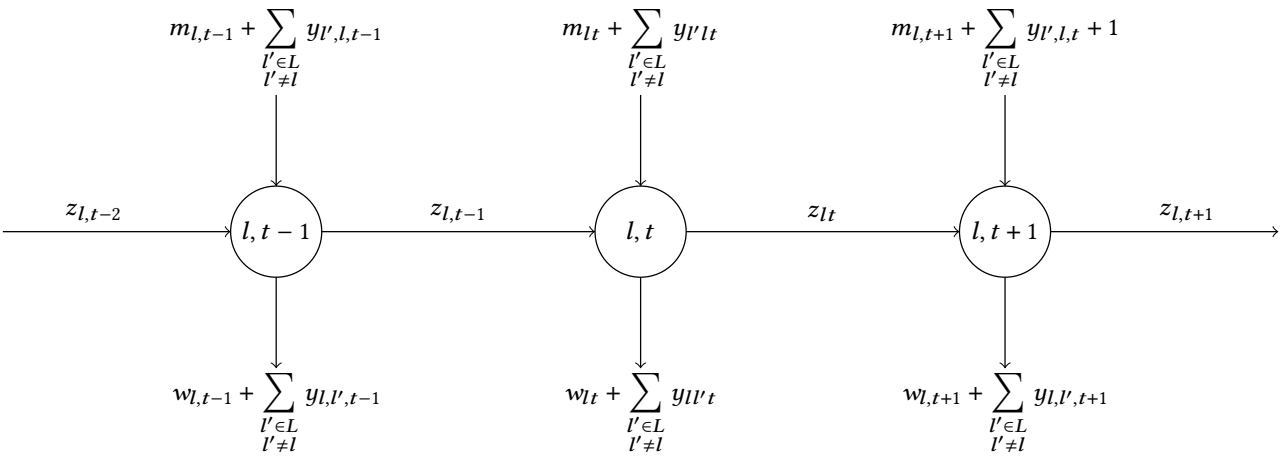


Figure 2: Linking relation between parameter m and variables y, z and w for a lab $l \in L$ at time intervals from $t-1$ to $t+1$.

With linking variables w_{lt}, we can constrain the number of tests performed with the following inequalities:

$$w_{lt} \leq \rho_{l,t-1} + \sum_{r \in R} x_{rlt} \qquad \forall l \in L,\ \forall t \in T \tag{5}$$

$$w_{lt} \leq Q_l \qquad \forall l \in L,\ \forall t \in T. \tag{6}$$

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

$$\rho_{r,t-1} + f_{rt} = \rho_{rt} + \sum_{l \in L} x_{rlt} \qquad \forall r \in R,\ \forall t \in T \tag{7}$$

$$\rho_{l,t-1} + \sum_{r \in R} x_{rlt} = \rho_{lt} + w_{lt} \qquad \forall l \in L,\ \forall t \in T. \tag{8}$$

Equation (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. Equation (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). Note that, because all variables are non-negative, eq. (8) renders eq. (5) redundant.

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

$$x_{rlt} \in \mathbb{N} \qquad \forall r \in R,\ \forall l \in L,\ \forall t \in T \tag{9}$$

$$y_{l_1 l_2 t} \in \mathbb{N} \qquad \forall l_1 \in L,\ \forall l_2 \in L \setminus \{l_1\},\ \forall t \in T \tag{10}$$

$$\rho_{rt} \in \mathbb{N} \qquad \forall r \in R,\ \forall t \in T \tag{11}$$

$$\rho_{lt}, z_{lt}, w_{lt} \in \mathbb{N} \qquad \forall l \in L,\ \forall t \in T \tag{12}$$

3.1 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's already working at full capacity, i.e., it's 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 ρ via the following *indicator*

constraints (see, e.g., [2]):

$$\gamma_{lt} = 0 \;\rightarrow\; w_{lt} \leq Q_l - 1 \qquad \forall l \in L,\ \forall t \in T \tag{13}$$

$$\gamma_{lt} = 0 \;\rightarrow\; \rho_{lt} \geq 1 \qquad \forall l \in L,\ \forall t \in T. \tag{14}$$

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

$$\gamma_{lt} = 0 \;\rightarrow\; \sum_{\substack{l' \in L \\ l' \neq l}} y_{ll't} = 0 \quad \forall l \in L,\ t \in T. \tag{15}$$

Equation (15) guarantees that laboratory l cannot ship out swabs unless it's 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 $\gamma^+_{lt}, \gamma^-lt \in \{0, 1\}$, 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:

$$\gamma^+_{lt} = 1 \;\rightarrow\; \sum_{\substack{l' \in L \\ l' \neq l}} y_{ll't} \geq 1, \qquad \gamma^+_{lt} = 0 \;\rightarrow\; \sum_{\substack{l' \in L \\ l' \neq l}} y_{ll't} = 0 \qquad \forall l \in L,\ \forall t \in T \tag{16}$$

$$\gamma^-_{lt} = 1 \;\rightarrow\; \sum_{\substack{l' \in L \\ l' \neq l}} y_{l'lt} \geq 1, \qquad \gamma^-_{lt} = 0 \;\rightarrow\; \sum_{\substack{l' \in L \\ l' \neq l}} y_{l'lt} = 0 \qquad \forall l \in L,\ \forall t \in T \tag{17}$$

$$\gamma^+_{lt} + \gamma^-_{lt} \leq 1 \qquad \forall l \in L,\ \forall t \in T. \tag{18}$$

Equation (18) enforces that each lab, on each day, cannot both send and receive swabs. Note that this restriction also avoids that a laboratory works as a relay, sending and re-shipping swabs between otherwise distant labs.

4 Model extensions

In this section we propose extensions to the model presented above, to account for more realistic scenarios.

4.1 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 reagents) 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 within each region, within different capacities.

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 $\delta_{rl} \in \{0, 1\}$ determines whether a factory $r \in R$ can ship reagent overnight to laboratory $l \in L$. Analogously, parameter $\mu_{l_1 l_2} \in \{0, 1\}$ 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^{r}_i and q^{s}_i 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 number of reagents 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. We present an **Extended Model**, taking into account the above considerations, in Appendix A.

4.2 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 [12] and Kalyanmoy [14]. In the following, we consider the hierarchical objective approach, in which the planner first optimises for their main objective (number of tests performed) and, 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 (21)–(42). We can obtain the solution to the hierarchical problem by solving the following **Hierarchical Model**:

$$\min \quad \sum_{l \in L} \sum_{t \in T} z_{lt} \tag{19}$$

$$\text{s.t.} \sum_{l \in L} z_{l|T|} = Z^* \tag{20}$$

$$(22)\text{–}(42),$$

where eq. (19) minimises the sum of all amounts of swabs stored at labs waiting to be tested and eq. (20) imposes that the considered solutions match the optimal value with respect to objective function (21). Objective function (19) is equivalent to minimising the average swab waiting time because (20) fixes the total number of swabs tested. Note that, when minimising the primary objective, two untested swab 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.

5 Related work

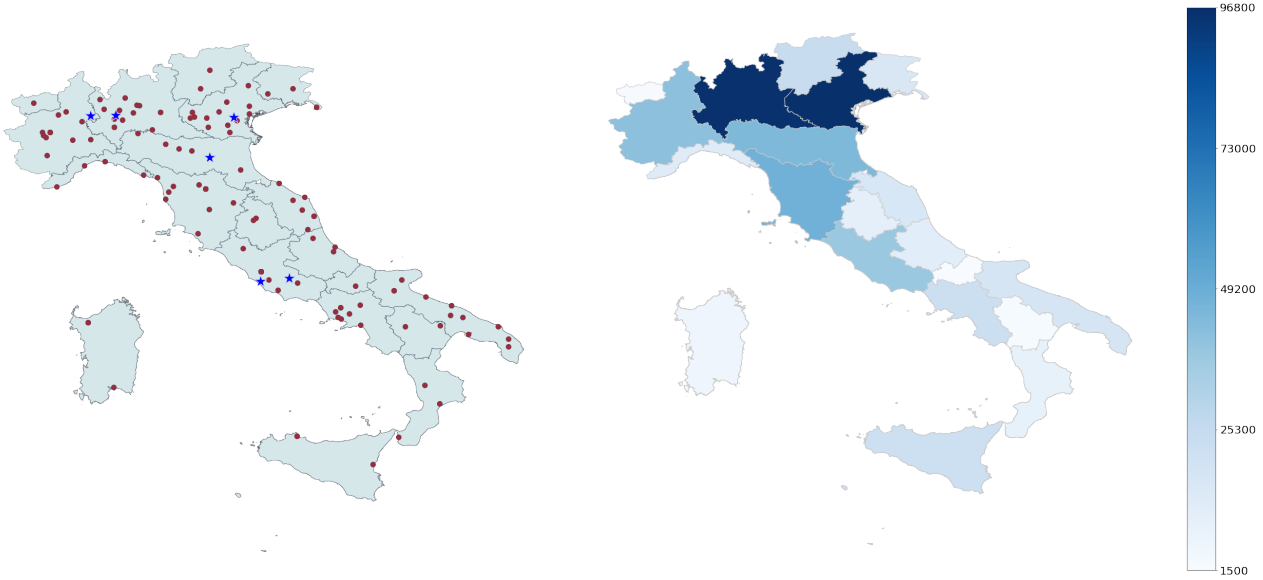
[[WIP: I think we can find relationships with Material Requirement Planning and other similar problems. We have to highlight how our problem is different, though, and so one needs to develop a new model. I think key differences are regional partitioning (I haven't seen MRP with this feature), the possibility of moving orders between production units (or, in our jargon, moving swabs between labs), and the hierarchical objective function. Also, different from classical MRP, we don't take costs into account (inventory, manpower, etc.)... we are in an emergency, after all! Also refer other COVID-related optimisation work. Reference work by the Bertsimas group [5, 4]. Other references [16, 15, 24].]]

6 Computational analysis

We want to assess the impact of optimising the allocation of reagents and swabs to laboratories on total testing capacity. 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 an extensive sensitive analysis to highlight 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 [23]. To solve the model, we use commercial solver Gurobi [13] 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.

6.1 Italy, April 1–13, 2020

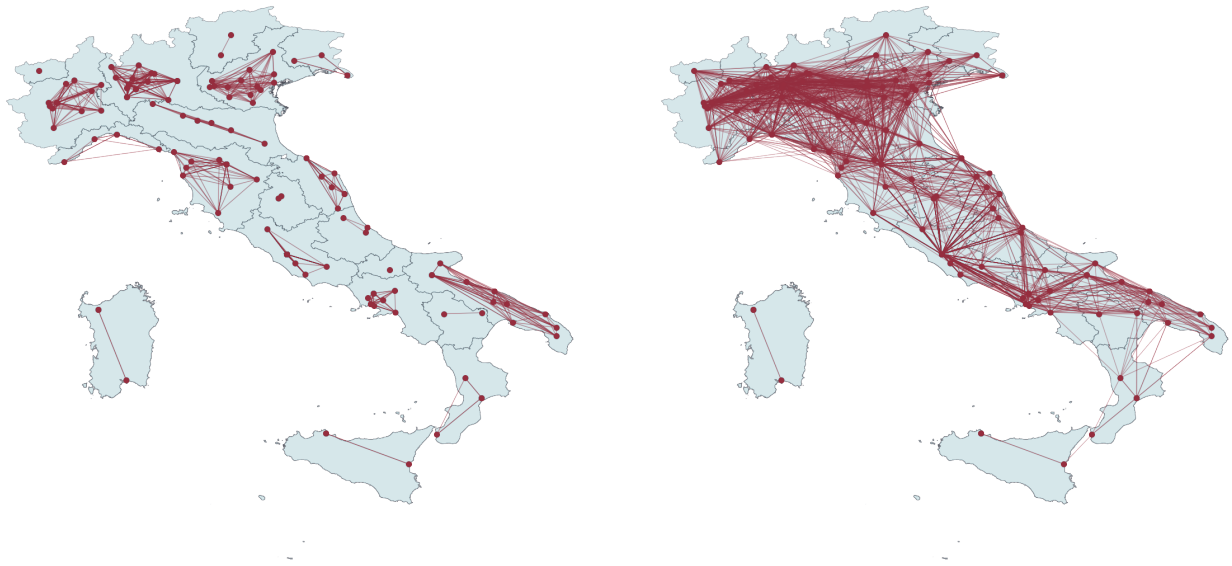
We create a model based on data sources relative to the COVID-19 pandemic in Italy [20, 17, 18, 26, 6]. We use data for the period 01–13 April, 2020 for the entire national territory. Figure 3a shows the distribution of laboratories and reagent factories in Italy [17]: dots represents labs which are officially authorised to execute COVID-19 swab tests by the Italian Health Ministry, while stars represent factories whose testing kits (including reagents, extraction kits, probes, negative controls) have been certified for use in COVID-19 swab testing.



(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 3: Regional-level data for Italy.



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

Figure 3b 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.

We analyse the number of swabs tested, day by day and region by region, using five different sources. The first is the official data released by Italy's "*Dipartimento della Protezione Civile*", which we refer to as **Real Data**.

Next, we consider the output obtained by the extended model presented in Appendix A, 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 [10]. We assume that labs can procure reagents from their geographically closest factory, and that procurement is optimised in a centralised way, with the planner deciding the quantities of reagent sent to each laboratory, to maximise the number of tested swabs. Because news sources and press releases by regional health authorities of Basilicata, Campania, Emilia-Romagna, Marche, Lombardia, Piemonte, Puglia, Srdegna, Sicilia, Toscana, Umbria and Veneto (i.e., 12 out of the 20 Italian regions) hint that reagents were a major bottleneck in expanding test capabilities, we expect this model to be able to increase the number of swabs tested, even if it doesn't allow inter-regional swab reassignment. Figure 4a shows how laboratories cluster within each region: an edge between two laboratories means that they belong to the same region.

Finally, we consider the output obtained from the extended 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 4b shows how laboratories can transfer swabs when using the 200Km threshold. Compare, e.g., Figures 4a and 4b to note how now a central planner has 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 [23] for the full description of the instance generation process and to Santini [22] for an interactive dashboard presenting the results of the analysis.

Figure 5 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 reagents 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 from *Real Data* to *Model - 400Km*) during 13 days.

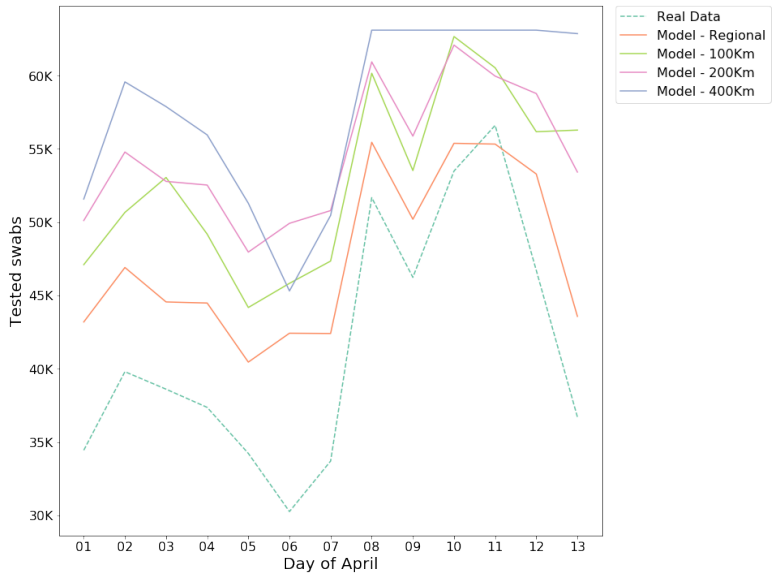


Figure 5: Number of swab tests performed on the national territory, per day. The dashed line reports official numbers from the Civil Defence agency. The other lines represent the results from applying the *Hierarchical Model* with intra-regional (orange column) and inter-regional (green, pink and blue lines) swab transfers. line inter-regional transfer radii considered are 100, 200 and 400Km.

	Real Data	**Model - Regional**		**Model - 100Km**		**Model - 200Km**		**Model - 400Km**	
Day of April	**#tests**	**#tests**	**%gain**	**#tests**	**%gain**	**#tests**	**%gain**	**#tests**	**%gain**
1	34455	43202	25.39	47110	36.73	47110	36.73	47110	36.73
2	39809	46916	17.85	50674	27.29	50674	27.29	50674	27.29
3	38617	44567	15.41	53055	37.39	53055	37.39	53055	37.39
4	37375	44492	19.04	49190	31.61	49190	31.61	49190	31.61
5	34237	40463	18.19	44187	29.06	44187	29.06	44187	29.06
6	30271	42433	40.18	45835	51.42	45835	51.42	45835	51.42
7	33713	42414	25.81	47353	40.46	47353	40.46	47353	40.46
8	51680	55454	7.30	60163	16.41	60163	16.41	60163	16.41
9	46244	50207	8.57	53528	15.75	53528	15.75	53528	15.75
10	53495	55385	3.53	62666	17.14	62666	17.14	62666	17.14
11	56609	55332	-2.26	60536	6.94	60536	6.94	60536	6.94
12	46720	53292	14.07	56175	20.24	56175	20.24	56175	20.24
13	36717	43582	18.70	56287	53.30	56287	53.30	56287	53.30
Total	539942	617739	14.41	686759	27.19	709996	31.49	750464	38.99

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.

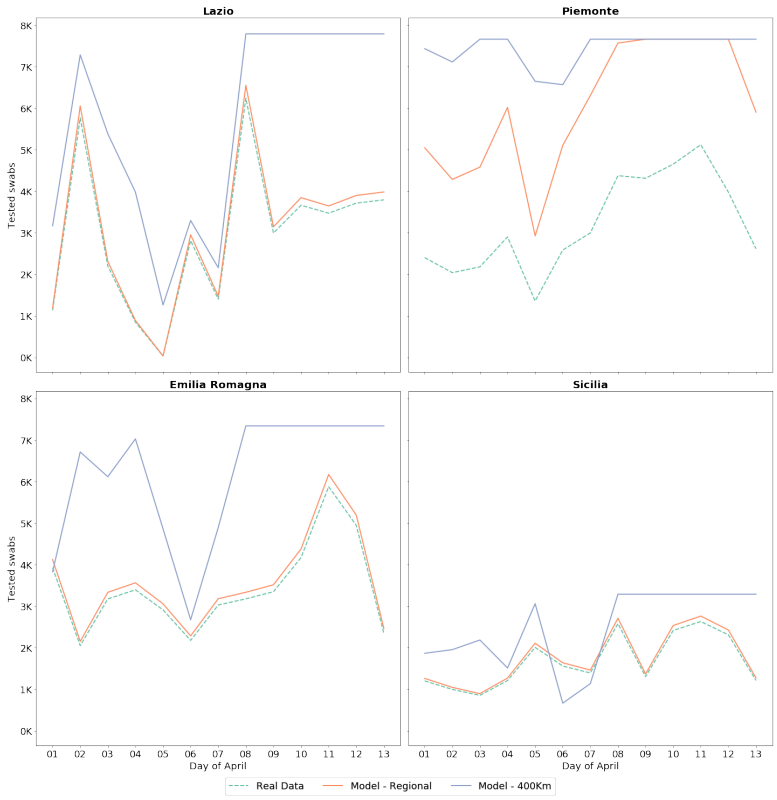


Figure 6: Number of swab tests performed in four Italian regions. The dashed line reports official numbers from the Civil Defence agency. The solid lines represents results obtained with our *Hierarchical Model* allowing intra-regional (orange line) and inter-regional (blue line) swab transfers.

Figure 6 shows example results from the *Regional* and *400Km* models, for four representative Italian regions 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. In these 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 isn't reagent availability anymore, but structural limitations such as machines. 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

[[WIP: Add analysis on synthetic data.]]

7 Conclusions

[[WIP: Add conclusions.]]

Acknowledgements

The authors would like to thank Michele Iacono of Roche Diagnostics S.p.A. and Xavier Jiménez Fàbrega of the Catalan Health Department for comments and fruitful discussions on this work. Alberto Santini was partially supported by grant "RTI2018-095197-B-I00" from the Spanish Ministry of Economy and Competitiveness.

References

[1] Robert Baird. "Why Widespread Coronavirus Testing Isn't Coming Anytime Soon". In: *The New Yorker* (Mar. 24, 2020). URL: https://www.newyorker.com/news/news-desk/why-widespread-coronavirus-testing-isnt-coming-anytime-soon.

[2] Pierre Bonami, Andrea Lodi, Andrea Tramontani, and Sven Wiese. "On mathematical programming with indicator constraints". In: *Mathematical programming* 151 (2015), pp. 191–223. DOI: 10.1007/s10107-015-0891-4.

[3] Stephen Bustin. "Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems". In: *Journal of molecular endocrinology* 29 (1 2002), pp. 23–39. DOI: 10.1677/jme.0.0290023.

[4] Operations Research Center. *Optimization can solve the ventilator shortage*. 2020. URL: https://www.covidanalytics.io/ventilator_allocation.

[5] Operations Research Center. *Ventilator Pooling: Formulation and Data Sources*. Tech. rep. Massachussets Institute of Technology, 2020. URL: https://www.covidanalytics.io/ventilator_documentation_pdf.

[6] Covid19Italia. *Open Data Coronavirus*. 2020. URL: https://www.covid19italia.info/opendata/.

[7] Center for Disease Control and Prevention. *CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel*. Tech. rep. CDC-006-00019, Revision: 03. U.S. Department of Health and Human Services, Mar. 30, 2020. URL: https://www.fda.gov/media/134922/download.

[8] Center for Disease Control and Prevention. *Processing of Sputum Specimens for Nucleic Acid Extraction*. Tech. rep. U.S. Department of Health and Human Services, Feb. 7, 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/downloads/processing-sputum-specimens.pdf.

[9] European Centre for Disease Prevention and Control. *Coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK*. Tech. rep. Seventh update. Mar. 25, 2020. URL: https://www.ecdc.europa.eu/sites/default/files/documents/RRA-seventh-update-Outbreak-of-coronavirus-disease-COVID-19.pdf.

[10] George France, Francesco Taroni, and Andrea Donatini. "The Italian health-care system". In: *Health Economics* 14 (S1 2005), S187–S202.

[11] Willard Freeman, Stephen Walker, and Kent Vrana. "Quantitative RT-PCR: pitfalls and potential". In: *Biotechniques* 26.1 (1999), pp. 112–125. DOI: 10.2144/99261rv01.

[12] Nyoman Gunantara. "A review of multi-objective optimization: Methods and its applications". In: *Cogent Engineering* 5.1 (2018). DOI: 10.1080/23311916.2018.1502242.

[13] Gurobi Optimization LLC. *Gurobi Optimizer Reference Manual*. 2020. URL: https://www.gurobi.com.

[14] Deb Kalyanmoy. "Multi-Objective Optimization". In: *Seach Methodologies*. Ed. by Edmund Burke and Graham Kendall. 2014, pp. 403–449. DOI: 10.1007/978-1-4614-6940-7_15.

[15] Lorenzo Lampariello and Simone Sagratella. *Effectively managing diagnostic tests to monitor the COVID-19 outbreak in Italy*. Tech. rep. Optimization Online, 2020. URL: http://www.optimization-online.org/DB_FILE/2020/03/7680.pdf.

[16] Sanjay Mehrotra, Hamed Rahimian, Masoud barah, Fengqiao Luo, and Schantz Karolina. *A Model of Supply-Chain Decisions for Resource Sharing withan Application to Ventilator Allocation to Combat COVID-19*. Tech. rep. Optimization Online, 2020. URL: http://www.optimization-online.org/DB_FILE/2020/04/7719.pdf.

[17] Ministero della Sanità — Direzione Generale della Prevenzione Sanitaria. *Pandemia di COVID-19 — Aggiornamento delle indicazioni sui test diagnostici e sui criteri da adottare nella determinazione delle priorità. Aggiornamento delle indicazioni relative alla diagnosi di laboratorio*. Apr. 2020. URL: www.trovanorme.salute.gov.it/norme/renderNormsanPdf?anno=2020&codLeg=73799&parte=1%20&serie=null.

[18] OpenToscana. *Open Data Covid19*. 2020. URL: http://dati.toscana.it/dataset/open-data-covid19.

[19] Xingfei Pan, Dexiong Chen, Yong Xia, Xinwei Wu, Tangsheng Li, Xueting Ou, Liyang Zhou, and Jing Liu. "Asymptomatic cases in a family cluster with SARS-CoV-2 infection". In: *The Lancet Infectious Diseases* 20.4 (2020), pp. 410–411. DOI: 10.1016/S1473-3099(20)30114-6.

[20] Presidenza del Consiglio dei Ministri — Dipartimento della Protezione Civile. *COVID-19 Italia: Monitoraggio della Situazione*. 2020. URL: https://github.com/pcm-dpc/COVID-19.

[21] Steveb Sanche, Yen Ting Lin, Chonggang Xu, Ethan Romero-Severson, Nick Hengartner, and Rulan Ke. "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2." In: *Emerging infectious diseases* 26.7 (2020). DOI: 10.3201/eid2607.200282.

[22] Alberto Santini. *Increase COVID-19 Swab Testing Capacity via Optimisation*. 2020. URL: https://santini.in/covid/.

[23] Alberto Santini. *Optimisation tools for the COVID-19 pandemic*. 2020. URL: https://github.com/alberto-santini/covid-optimisation.

[24] Ruggiero Seccia. *The Nurse Rostering Problem in COVID-19 emergency scenario*. Tech. rep. 2020. URL: http://www.optimization-online.org/DB_FILE/2020/03/7712.pdf.

[25] Selma Souf. "Recent advances in diagnostic testing for viral infections". In: *Bioscience Horizons* 9 (2016). DOI: 10.1093/biohorizons/hzw010.

[26] Task force COVID-19 del Dipartimento Malattie Infettive e Servizio di Informatica, Istituto Superiore di Sanità. *Epidemia COVID-19, Aggiornamento Nazionale*. Apr. 2020. URL: https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_16-aprile-2020.pdf.

[27] Di Wu, Tiantian Wu, Qun Liu, and Zhicong Yang. "The SARS-CoV-2 outbreak: what we know". In: *International Journal of Infectious Diseases* (2020). DOI: 10.1016/j.ijid.2020.03.004.

A Extended Model formulation

The extended model which incorporates the considerations made in Section 4 uses a new variable $u_{lt} \in \mathbb{N}$, which represents the number of new swabs assigned to lab $l \in L$ on day $t \in T$. The model reads as follows:

$$\min \quad \sum_{l \in L} z_{l|T|} \tag{21}$$

$$\text{s.t.} \quad \sum_{r \in R} \sum_{l \in L_i} x_{rlt} \leq q_i^{\mathrm{r}} \qquad \forall i \in \{1, \ldots, n\},\ \forall t \in T \tag{22}$$

$$\sum_{l_1 \in L} \sum_{l_2 \in L_i} y_{l_1 l_2 t} \leq q_i^{\mathrm{s}} \qquad \forall i \in \{1, \ldots, n\},\ \forall t \in T \tag{23}$$

$$x_{rlt} = 0 \qquad \forall r \in R,\ \forall l \in L \ :\ \delta_{rl} = 0,\ \forall t \in T \tag{24}$$

$$y_{l_1 l_2 t} = 0 \qquad \forall l_1, l_2 \in L \ :\ \mu_{l_1 l_2} = 0,\ \forall t \in T \tag{25}$$

$$z_{l,t-1} + u_{lt} + \sum_{\substack{l' \in L \\ \mu_{l'l}=1}} y_{l'lt} = z_{lt} + w_{lt} + \sum_{\substack{l' \in L \\ \mu_{ll'}=1}} y_{ll't} \qquad \forall l \in L,\ \forall t \in T \tag{26}$$

$$\sum_{l \in L_i} u_{lt} = m_{it} \qquad \forall i \in \{1, \ldots, n\},\ \forall t \in T \tag{27}$$

$$w_{lt} \leq \rho_{l,t-1} + \sum_{\substack{r \in R \\ \delta_{rl}=1}} x_{rlt} \qquad \forall l \in L,\ \forall t \in T \tag{28}$$

$$w_{lt} \leq Q_l \qquad \forall l \in L,\ \forall t \in T \tag{29}$$

$$\rho_{r,t-1} + f_{rt} = \rho_{rt} + \sum_{\substack{l \in L \\ \delta_{rl}=1}} x_{rlt} \qquad \forall r \in R,\ \forall t \in T \tag{30}$$

$$\rho_{l,t-1} + \sum_{\substack{r \in R \\ \delta_{rl}=1}} x_{rlt} = \rho_{lt} + w_{lt} \qquad \forall l \in L,\ \forall t \in T \tag{31}$$

$$\gamma_{lt} = 0 \ \rightarrow \ w_{lt} \leq Q_l - 1 \qquad \forall l \in L,\ \forall t \in T \tag{32}$$

$$\gamma_{lt} = 0 \ \rightarrow \ \rho_{lt} \geq 1 \qquad \forall l \in L,\ \forall t \in T \tag{33}$$

$$\gamma_{lt} = 0 \ \rightarrow \ \sum_{\substack{l' \in L \\ \mu_{ll'}=1}} y_{ll't} = 0 \qquad \forall l \in L,\ t \in T \tag{34}$$

$$\gamma_{lt}^{+} = 1 \ \rightarrow \ \sum_{\substack{l' \in L \\ l' \neq l}} y_{ll't} \geq 1, \quad \gamma_{lt}^{+} = 0 \ \rightarrow \ \sum_{\substack{l' \in L \\ l' \neq l}} y_{ll't} = 0 \qquad \forall l \in L,\ \forall t \in T \tag{35}$$

$$\gamma_{lt}^{-} = 1 \ \rightarrow \ \sum_{\substack{l' \in L \\ l' \neq l}} y_{l'lt} \geq 1, \quad \gamma_{lt}^{-} = 0 \ \rightarrow \ \sum_{\substack{l' \in L \\ l' \neq l}} y_{l'lt} = 0 \qquad \forall l \in L,\ \forall t \in T \tag{36}$$

$$\gamma_{lt}^{+} + \gamma_{lt}^{-} \leq 1 \qquad \forall l \in L,\ \forall t \in T \tag{37}$$

$$x_{rlt} \in \mathbb{N} \qquad \forall r \in R,\ \forall l \in L \ :\ \delta_{rl} = 1,\ \forall t \in T \tag{38}$$

$$y_{l_1 l_2 t} \in \mathbb{N} \qquad \forall l_1 \in L,\ \forall l_2 \in L \ :\ \mu_{l_1 l_2} = 1,\ \forall t \in T \tag{39}$$

$$\rho_{rt} \in \mathbb{N} \qquad \forall r \in R,\ \forall t \in T \tag{40}$$

$$\rho_{lt}, z_{lt}, w_{lt}, u_{lt} \in \mathbb{N} \qquad \forall l \in L,\ \forall t \in T \tag{41}$$

$$\gamma_{lt} \in \{0, 1\} \qquad \forall l \in L,\ \forall t \in T \tag{42}$$