

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

Spatio-temporal forecasting of COVID-19 cases in the Netherlands for source and contact tracing

Max C. Keuken^{1,*†}, Jizzo R. Bosdriesz^{2,3,4,5,†}, Anders Boyd², Elisabeth M. den Boogert⁶, Ivo K. Joore⁷, Nicole H.T.M. Dukers – Muijers^{8,9}, Gini van Rijckevorsel^{2,10}, Hannelore M. Götz^{10,11,12}, Irene E. Goverse¹³, Mariska W.F. Petrignani¹⁴, Stijn F.H. Raven¹⁵, Susan van den Hof¹⁰, Kirsten V.C. Wevers-de Boer¹⁶, Maarten F. Schim van der Loeff^{2,3,4,5}, and Amy Matser^{2,3,4,5}

Submission History

Submitted:	24/08/2024
Accepted:	20/02/2025
Published:	07/05/2025

¹ Corona Data team, Public Health Service (GGD) of Amsterdam, Amsterdam, the Netherlands

² Department of Infectious Diseases, Public Health Service (GGD) of Amsterdam, Amsterdam, the Netherlands

³ Department of Internal Medicine, Amsterdam UMC location University of Amsterdam, Infectious Diseases, Amsterdam, the Netherlands

⁴ Amsterdam Institute for Infection and Immunity, Infectious Diseases, Amsterdam, the Netherlands

⁵ Amsterdam Public Health, Amsterdam UMC, Academic Medical Center, Amsterdam, the Netherlands

⁶ Department of Infectious Disease Control, Public Health Service (GGD) Hart voor Brabant, 's-Hertogenbosch, the Netherlands

⁷ Department of Infectious Disease Control and Sexual Health, Public Health Service (GGD) Flevoland, Lelystad, the Netherlands

⁸ Department of Sexual Health, Infectious Diseases and Environmental Health, Living Lab Public Health Mosa, South Limburg Public Health Service (GGD), Heerlen, The Netherlands

⁹ Department of Health Promotion, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands

¹⁰ Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands

¹¹ Department of Infectious Disease Control, Public Health Service (GGD) Rotterdam-Rijnmond, Rotterdam, the Netherlands

¹² Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, the Netherlands

¹³ Department of Infectious Diseases, Public Health Service (GGD) Groningen, Groningen, the Netherlands

¹⁴ Department of Infectious Disease Control, Public Health Service (GGD) Haaglanden, The Hague, the Netherlands

¹⁵ Department of Infectious Diseases, Public Health Service (GGD) region Utrecht, Zeist, the Netherlands

¹⁶ Department of Infectious Disease Control, Public Health Services (GGD) Gelderland Midden, Arnhem, the Netherlands

[†] Equal contribution

Abstract

Source and contact tracing (SCT) is a core public health measure that is used to contain the spread of infectious diseases. It aims to identify a source of infection, and to advise those who have been exposed to this source. Due to the rapid increases in incidence of COVID-19 in the Netherlands, the capacity to conduct a full SCT quickly became insufficient. Therefore, the public health services (PHS) might benefit from a restricted strategy targeted to geographical regions where (predicted) case-to-case transmission is high. In this study, we set out to develop a prediction model for the number of COVID-19 cases per postal code within the Netherlands using geographic and demographic features. The study population consists of individuals residing in one of the participating nine Dutch PHS regions who tested positive for SARS-CoV-2 between 1 June 2020 and 27 February 2021. Using a machine learning random forest regression model, we predicted the top 100 postal codes with the highest number of cases with an accuracy of 49% for the current week, 42% for next week, and 44% for two weeks from present. In addition, the age groups of 20-39 and 40-64 years had a higher prediction accuracy than groups outside these age ranges. The developed model provides a starting point for targeted preventive SCT efforts that incorporate geospatial and demographic characteristics of a neighbourhood. It should nonetheless be noted that during the early stages of the outbreak, the number of available datapoints needed to inform such models are likely insufficient. Given the accuracy and data requirements of the developed model, it is unlikely that this class of models can play a pivotal role in informing policy during the early phases of a future epidemic.

Keywords

COVID-19; SARS-CoV-2; SCT; forecasting; machine learning

*Corresponding Author:

Email Address: mckeuken@gmail.com (Max C. Keuken)



Introduction

In early 2020, the coronavirus disease 2019 (COVID-19) pandemic, caused by the airborne severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), reached the Netherlands. In the absence of vaccines, the main public measure that was used to control the spread of the virus was testing combined with source and contact tracing (SCT) [1]. The main aims of SCT are to identify where and from whom the person who tested positive might have gotten infected, and whom they have had contact within their infectious period, so that preventive measures such as quarantine can be implemented.

In the Netherlands, SCT for notifiable diseases and outbreaks is conducted by the Public Health Services (PHS; in Dutch: Gemeentelijke Gezondheidsdienst or GGD), for example after diagnosis of direct contact diseases like hepatitis C virus but also airborne diseases such as tuberculosis [2–4]. SCT for one SARS-CoV-2 infection took between 8 and 12 hours per index case and thus was a time-consuming process [5]. Call centres for SCT expanded on a regional and national level, during high-incidence periods the demand for SCT was much higher than the available capacity. Depending on the incidence and SCT, the call centre capacity in a region and thoroughness of SCT were modified during the various pandemic waves of infections [5]. During times of high incidence, SCT was prioritized to persons at risk of severe health outcomes (e.g., ≥ 80 years old) or who were deemed critical workers. The remaining persons who tested positive were informed by email and took 30 minutes per case. With decreasing incidence, SCT was extended again to other groups. However, the prioritization was coarse and it was unclear what the actual effects were on persons at risk of severe health outcomes, or in preventing onward transmission [6, 7].

We hypothesized that SCT prioritization would be better served by an evidence-based data-driven prediction model, because with such a model we would be able to identify early local rises in incidence. Predicting transmission risk on an individual level can be challenging, as many factors, both measured and unmeasured, influence transmission dynamics. Nonetheless, previous work has highlighted the substantial contribution of geospatial and demographic factors in the spread of SARS-CoV-2 (e.g., [8–11]). If we could predict new cases on a local geographical level, we could use it for fine-grained prioritization of SCT to new cases residing in specific areas. Such a tailor-made approach could improve societal support for the governmental policy and potentially increase the adherence to the restrictions in place [12]. Such a geospatial and age specific approach would be complimentary to other targeted approaches in the Netherlands such as placing mobile SARS-CoV-2 testing facilities in high-COVID-19-incidence neighbourhoods or mass testing, leading to increases in test uptake [13, 14] or to offer COVID-19 vaccination in neighbourhoods where vaccine uptake had been low with the aim of increasing vaccine uptake [15].

In this study, we aimed to develop a prediction model for the number of cases infected with SARS-CoV-2 at the level of postal codes using several geographic and (socio-) demographic features (e.g., spatial autocorrelation between postal codes, Google search history, population density, and

household size). Furthermore, we aimed to develop an age-group-specific prediction model for the number of individuals infected with SARS-CoV-2. This second model was deemed necessary as SCT prioritization could focus on age groups that are known to have a higher risk of severe health outcomes or on age groups that are thought to have a higher risk of spreading the disease. The main contributions of this study include highlighting the feasibility of employing machine learning techniques for predicting COVID-19 infections in the early stages of an outbreak, albeit with a significant degree of uncertainty. Secondly, the findings of the age-group-specific model stress the importance to account for demographic factors when predicting future COVID-19 cases.

Methods

Description of target and feature data and pre-processing

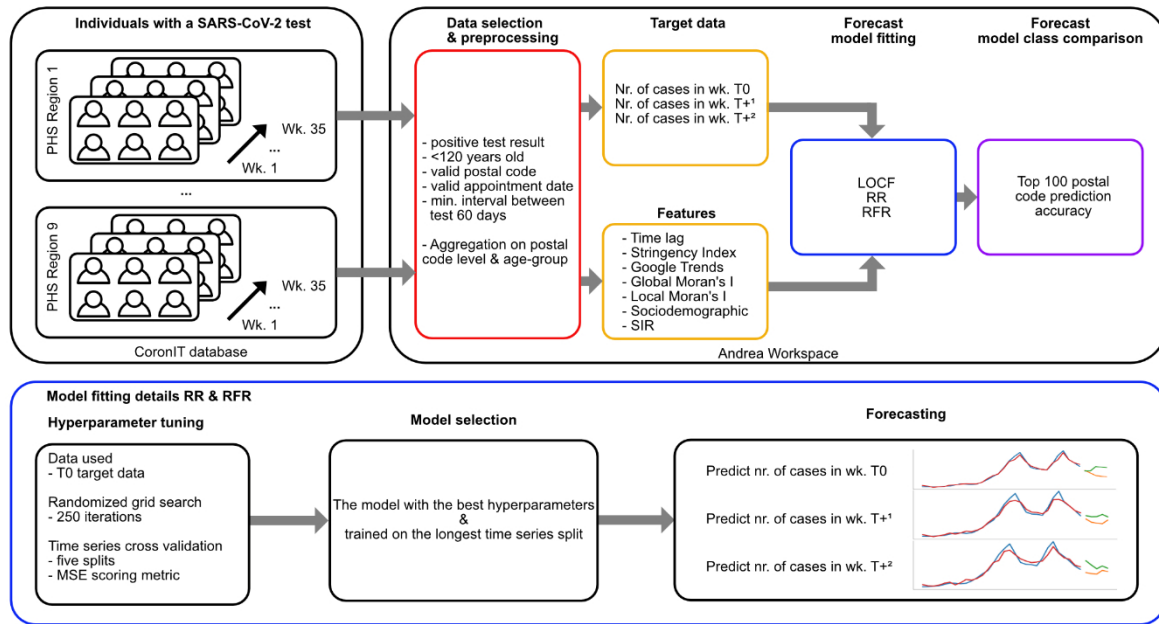
Pre-processing data

Data from individuals who tested positive between 1 June 2020 and 27 February 2021 in one of the SARS-CoV-2 testing facilities and who resided in one of the participating nine PHS regions in the Netherlands were used in this study. The participating PHS regions were Amsterdam, Flevoland, Gelderland-Midden, Groningen, Haaglanden, Hart voor Brabant, Rotterdam-Rijnmond, regio Utrecht, and Zuid-Limburg.

From 1 June 2020, any person in the Netherlands could request a free SARS-CoV-2 nucleic acid amplification test (NAAT) by providing minimal personal information and information on symptoms experienced. For routine purposes, the citizen service number (in Dutch: *burger service nummer* or BSN) was used to obtain information on gender and age from the population registry. Data were stored in a database system, CoronIT. The data used in the current study was retroactively extracted from the CoronIT system by each of the PHS regions separately on various dates between 7 July 2021 and 15 November 2022. These records did not include any self-tests. Data were pre-processed using a custom R script. The relevant steps for the current study entailed calculating age in years at the time of scheduling the test appointment and creating a unique identifier column that did not contain any directly identifiable information. This step was done using an md5 hash function as implemented in the R package “digest” (v0.6.29) with last name, postal code and age in years as input. Then, any directly identifiable information (e.g., name, date of birth and address) was removed from the records before further processing and uploaded to a common anDrea workspace [16]. All further processing was done in Python (v3.8) [17].

Records were excluded from the dataset if they did not have any one of the following: valid postal code, valid appointment date, age of the case under 120 years, or a positive test result. Any positive test result within 60 days after a previously positive test result was excluded as it was unlikely to reflect a new infection [18]. See Figure 1. for a visual representation of the overall dataflow and processing steps.

Figure 1: Data and modelling flow chart



The top row highlights the main data processing and model fitting steps. The individual data were collected per PHS region and extracted from the CoronIT database to further analyze the data in the secure environment Andrea. Based on a number of inclusion criteria, the raw data was cleaned, aggregated per postal code area, and separated into a target and feature dataset. The feature dataset was enriched with additional features such as the Stringency Index, Google Trends, and sociodemographic features of the postal code region. Three models were applied to the dataset and, based on the top 100 postal code prediction accuracy, the models were statistically compared. The bottom row highlights the model fitting steps in more detail. For the RR and RFR models, a number of hyperparameters were tuned using a randomized grid search with 250 iterations. Each iteration was cross-validated using a time series cross-validation method with five splits and the MSE as a scoring metric. Note that one of the RFR tuned hyperparameters was the number of features to consider for the best split. The best fitting RR and RFR model which was trained on the longest time series split was used to predict the number of cases. LOCF: last observation carried forward; MSE: Mean Squared error; RFR: random forest regression; RR: ridge regression; PHS: Public Health Service.

Ethical statement

The Data Protection Officer of the city of Amsterdam concluded that the study complies with the European General Data Protection Regulation (art. 9.2.j & art. 89). Under this law individual informed consent did not have to be obtained, because it was deemed not feasible due to the retrospective nature of the study and population size. The study also complies with the Dutch Law on Medical Treatment Agreement (WGBO art. 7:458). The Medical Ethics Committee of the Amsterdam University Medical Centers reviewed the study and deemed it outside the scope of the Medical Research Involving Human Subject Act and formally waived the need for ethical approval (W20_432#20.479).

Target data

The number of positive COVID-19 cases in a given week (T^0) was aggregated per 4-digit postal code area and over six age groups (i.e., <12, 12–19, 20–39, 40–64, 65–79, and ≥ 80 years old, at the time of the test appointment) and used as the main target variable. In addition to week T^0 , two additional target time series were created to reflect the number of cases in the following week (T^{+1}) and in the week thereafter (T^{+2}) by shifting the T^0 target time series backwards for the corresponding time period. In 2020, there were 4074

distinct 4-digit postal code areas in the Netherlands with a mean of 4272 [standard deviation (SD) 4297] inhabitants. The geospatial distribution of the 1,479 4-digit postal code areas represented in this study is shown in Supplementary Figure 1.

Features

Several features were included to predict the number of positive COVID-19 cases per postal code, age group and week. To capture the inherent sequential nature of the time series, a number of features containing lags (in weeks) of the target variable were added [19]. This meant shifting the T^0 target time series forward with the corresponding number of time periods. The number of lags per postal code and age group were determined using the partial autocorrelation function as implemented in statsmodels (v0.13.2) [20] with the Levinson-Durbin recursion with bias correction method and an α of 0.05 [21]. Based on the resulting histogram of the number of lag features per postal code area and age group, two temporal lag features of 1 and 2 were added. As a result, the observed number of cases of the previous week and two weeks ago were used to inform the prediction on the current week.

To incorporate the different restrictions over time imposed by the Dutch government, the composite metric government response Stringency Index (SI), as provided by the Oxford Coronavirus Government Response Tracker, was used [22].

The SI is a simple additive index between 0 and 100 that aggregates the following ordinal encoded factors: school closing, workplace closing, cancelling of public events, restrictions on gathering size, stay-at-home requirements, restrictions on domestic and international travel. Higher SI values correspond to stricter restrictions imposed by the government. The daily SI values were averaged over the corresponding week numbers and added to the models as a feature.

Google Trends (GT; Google, Inc) is an online portal in which the proportion of searches for specific search terms relative to the most popular searches through Google can be extracted for a given location and time. A noted limitation of GT is that other co-occurring events might skew the relative importance over time. This search behaviour of a population for certain disease terms has been shown to be informative for future COVID-19 outbreaks (e.g., [23, 24]). Through the use of the python package PyTrends (v4.8.0) [25], weekly GT data for the Netherlands were acquired between 1 May 2020 and 30 March 2021. The search terms included 'COVID-19' and all the suggested keywords by GT. The resulting keywords were translated into Dutch as well. In addition to these search terms, a search term related to olfactory symptoms was included, due to their predictive value for COVID-19 outbreaks [26, 27]. The final list of search terms was: 'COVID-19', 'COVID-19 Testen', 'COVID-19 Testing', 'COVID-19 Vaccine', 'Coronavirus disease 2019', 'symptoms of COVID-19', 'symptomen van COVID-19', 'loss of smell', and 'reuk'.

For the spread of an infectious disease, such as COVID-19, it is assumed that the spread is not geographically independent [28, 29]. To quantify the spatial autocorrelation, the global and local Moran's I [30] were calculated per week using the PySAL esda python package (v2.4.1) [31]. To estimate which irregular postal code shape were sharing a border the adjacency was calculated and this was defined by a shared vertex between the postal code shape polygons using the queen contiguity weights [32]. The local and global Moran's I were calculated per week for all age groups separately and combined, and used as a feature to predict the current and future number of individuals with SARS-CoV-2.

The number of people living in a given neighbourhood, number of households, average size of the households, and the population density, all per postal code as of 1 January 2020, were obtained from the Central Bureau for Statistics (CBS) and were added as features [33] because these features can play a role in the spread of an infectious diseases [34, 35].

Finally, to include a feature that focusses on whether the actual number of cases in a particular geographic area deviates from the expected number given the population, the Standardized Incidence Ratio (SIR) was calculated for all postal codes and used as a feature to predict the current and future number of individuals with SARS-CoV-2 [36]. Since the total number of people per age group were not available per postal code, the SIR was only calculated for the entire population per week. The feature importance for the winning T^0 model was computed as the mean accumulation of the impurity decrease for each tree and were provided as a fitted attributed of the model within Scikit-Learn (v1.0.2) [37]. See Table 1 for an overview of the features and target data.

Forecast models

A baseline model "Last observation carried forward" (LOCF) was created where the last observation of the training data was carried forward and used as a naïve prediction for the test data (training and test data described below). In addition, two classes of multi-output regression models [i.e., ridge regression (RR) and random forest regression (RFR) [38, 39]] were fitted to the data per postal code and collapsed over or per age group. This was done for each of the three targets (T^0 , T^{+1} , T^{+2}) which were either log-transformed or left non-transformed. The log-transformation pre-processing step was included in the model comparison to test whether this would improve the model performance. The log-transformation was done by adding a constant of 1 to the data [40]. In total, twelve models were fitted within each of the three model types [three targets (T^0 , T^{+1} , T^{+2}) $\times$ two age models (global, age group specific) $\times$ two scales (log-transformed or non-transformed)].

To improve the convergence of the models, a standard scaler as implemented in Scikit-Learn was fitted and transformed on the training data (30 observations per postal code and age group) and applied to the test data (5 observations per postal code and age group) for both RR and RFR model classes. A scaler was not necessary for the LOCF baseline model as no actual model was fitted to the data. For all models, the mean squared error (MSE) was used as a scoring metric to quantify the model fit where smaller values indicate a better fit.

For all three model types, we calculated the percentage of correctly predicted postal code areas with the highest number of individuals with a positive SARS-CoV-2 test in a given week, and ranked them while only presenting the top 100. This was done for the final week of test data (i.e., the furthest removed from the training time series and considered the most difficult period to predict). This was deemed the main metric of interest as it could be used for prioritization of SCT.

Ridge regression

Since there were more features than aggregated datapoints per week, parameter regularization was needed [41] and hence ridge regression, as implemented in Scikit-Learn (v1.0.2) [37], was chosen over standard linear regression. For brevity we refer to the documentation of Scikit-Learn for the exact mathematical implementation [42]. The regularization parameter α was determined using randomized grid search with the following hyperparameter: α parameter space ranging between 1 and 10,000 with steps of 0.1, whereas the randomized search was executed 250 times with a time series cross validation approach (i.e., five splits, where each split had six test samples), using MSE as the scoring metric was employed where smaller values indicate a better model fit. The data used for the grid search were at the T^0 target time window and the resulting α was used for the T^{+1} and T^{+2} targets. The randomized grid search results indicated that the optimal α , given the data and search space, was 9913.60.

Finally, the RR model that was trained on the longest time series split was used to validate the model by predicting the subsequent five weeks of test data. This was done for the T^0 , T^{+1} and T^{+2} targets.

Table 1: An overview of the observations and features included in the analysis

Type of data	Summary	Description
Training and test data	The number of positive COVID-19 cases in each week, for a given postal code.	The total time series length was 38 weeks. The total number of postal code areas was 1,479 (34,6% of the total postal codes in the Netherlands)
Time lag feature	The shifted target time series one or two weeks forward to account for the inherent sequential nature of time series.	Two temporal lags were included so that the observed number of cases of the previous week and two weeks ago were used to inform the prediction of the current week.
NPI stringency index feature.	A composite metric that reflects the stringency of the government response in force in each week.	The SI is an additive index between 0 and 100 that aggregates the following NPIs: school and workplace closing, cancelling of public events, restrictions on gathering size, stay-at-home requirements and travel restrictions.
Google trends Feature	A metric that reflects the proportion of searches for specific terms relative to the most popular search in each week.	The Dutch and English search terms that were used were 'COVID-19', 'COVID-19 Testen', 'COVID-19 Testing', 'COVID-19 Vaccine', 'Coronavirus disease 2019', 'symptoms of COVID-19', 'symptomen van COVID-19', 'loss of smell', and 'reuk'.
Geospatial features	A metric that reflects the geographical autocorrelation between neighbouring postal codes.	The geospatial autocorrelation between neighbouring postal codes was calculated using the local and global Moran's I.
Demographic features	Several demographic characteristics of a given postal code area, which reflect population size, density and household composition.	The total population size, population density, number of households, and average household size as known on 1 Jan. 2020 were included per postal code area.
SIR feature	A feature that reflects how much the observed cases in a given postal code area deviated from the expected number of cases in each week.	An SIR value <1 indicates lower cases than expected, whereas a SIR of >1 indicates more cases than expected.

For each of the features, training, and test data a short description of what the data entails. See for more details the Methods section.

COVID-19: coronavirus disease 2019; NPI: Non-Pharmaceutical Intervention; SIR: Standardized Incidence Ratio.

Random forest regression

We also chose to include an RFR model [43], as implemented in Scikit-Learn (v1.0.2) [37]. RFR models allow for nonlinearities, are nonparametric, are able to learn interactions without having to explicitly model them, and might be more flexible than ridge regression models [44]. For brevity we refer to the documentation of Scikit-Learn for the exact mathematical implementation [45]. The number of estimators and the maximum depth parameters were determined using a randomized grid cross validation search with the following hyperparameters: n estimators parameter space ranged between 100 and 1,050 with steps of 50, maximum depth ranged between 10 and 110 with steps of 1, and the maximum number of features ranged between 0.1 and 1 with steps of 0.1. The data used for the grid search were at T^0 and the resulting hyperparameters were used for the T^{+1} and T^{+2} targets. The randomized grid search results indicated that the optimal number of estimators, maximum depth and features, given the data and search space, were 950, 81 and

0.3, respectively. All other training and testing procedures were identical to the RR model.

Statistical analysis

To test whether there was a statistical difference in predicting the top 100 postal codes between the different models, scales, and time windows, an ANOVA was used within a Bayesian framework with default uniform priors using the JASP software package (v.0.16.2.0) [46], which relies on the R package BayesFactor (v.0.9.10-2) [47]. The ANOVA used a uniform prior model probability as proposed by Rouder et al. [48]. The assumption of normality was assessed using a Q-Q plot of the residuals. Bayes Factors (BF), which are a continuous measure of support for a given model, were obtained for each of the twelve tested model. The model with the highest BF was considered the best fitting model given the data. The BFs are interpreted using the labels as proposed by Jeffreys [49] and adapted by Wetzels et al. [50] where cut-offs for the BF_{10} are made to indicate the level of evidence for the

alternative hypothesis (H_1): 1.0 (no evidence), between 1 and 3 (anecdotal evidence), between 3 and 10 (moderate evidence), between 10 and 30 (strong evidence), between 30 and 100 (very strong evidence), and larger than 100 (extreme evidence). Inversely, BF_{10} values smaller than 1.0 indicate the amount of evidence for H_0 .

Results

Data description

Data of 424,412 positive SARS-CoV-2 tests from 422,467 unique individuals across 1,479 Dutch 4-digit postal code areas between 1 June 2020 and 27 February 2021 were used. Of all unique cases, 201,640 (47.73%) were male, 219,490 (51.95%) were female and of 1,337 cases the gender was unknown. The mean age for all unique cases was 40 years old (SD 19). The population included 13,620 children of age <12 years, 54,410 persons of age 12-19, 151,452 persons of age 20-39, 158,493 of age 40-64, 36,003 of age 65-79, and 8457 of ≥ 80 . The mean number of positive tests per unique case was 1.01 with a min-max range of 1 to 3. The mean number of self-reported symptoms per case was 2.19 (SD 1.51) with a maximum number of possible reported symptoms of 8. Figure 2 provides an overview of the number of cases over time, PHS regions, postal codes, and age groups.

Forecasting results

Model performance all age groups collapsed

Visual inspection of the model fits indicated that both the RR and RFR models were able to capture the training data to a high degree (Figure 3 for the non-transformed data and Supplementary Figure 2 for the log-transformed data). The mean predicted time series for the RR and RFR seemed reasonable, as they followed the main trends of the observed time series. Note that for the baseline LOCF model there was no actual training as the last observation was carried forward.

Using a Bayesian ANOVA, we tested whether there was a statistical difference in predicting the top 100 postal codes with the highest number of infected individuals between the different models (LOCF, RR, RFR), scales (non-transformed, log-transformed), and time windows (T^{+0} , T^{+1} , T^{+2}).

On average the LOCF model resulted in the lowest MSE, i.e., had more precise predictions (Table 2 and Supplementary Table 1), and at T^0 , 41% of the postal codes predicted to be in the top 100 ranking with the highest number of infections were correctly predicted. However, the RFR model had the highest percentage, with 49% correct predictions; the RR model had the lowest percentage, 37%. At T^{+2} , these percentages were 33% (LOCF), 44% (RFR), and 30% (RR). In scenarios where a model (like the LOCF) exhibits a lower MSE but also lower accuracy compared to other tested models, it suggests that this particular model tends to make a significant number of small errors. Conversely, when a model (like the RFR) has slightly higher MSE but greater accuracy, it indicates that this model occasionally makes larger errors. Given the constraints of limited resources during crisis situations and the necessity to prioritize, rank accuracy was deemed to be the more crucial metric for this particular use case.

Model selection

The primary output from the ANOVA is presented in Table 3. The data was 1.13 times more likely under the best model, which includes the factors model type, scale and time windows and an interaction term between model type and scale (BF_{10} 1.0), than under the second-best model, which included the interaction term between time windows and scale (BF_{10} 0.88).

As the amount of evidence to prefer the best model over the second-to fourth-best models was anecdotal, an analysis of effects was conducted (Table 4). There was strong evidence to include the factors model type and time window (resp. BF_{incl} : >100 and BF_{incl} : 71.98). There was moderate evidence to include the factor scale (BF_{incl} : 3.86) and moderate evidence to include the interaction term between model type and scale (BF_{incl} : 8.18).

Based on the *post hoc* tests (Table 5) for the main effect of model types, there was strong evidence that the RFR model on average resulted in a higher prediction accuracy than the other models tested (resp. LOCF $BF_{10,u}$: 21.74 and RR $BF_{10,u}$: 15.60). There was anecdotal evidence for no difference in model performance between the non-transformed and log-transformed ($BF_{10,u}$: 0.49). As the non-transformed scale is directly interpretable it was decided to report the log-transformed results in the supplementary section. The *post hoc* analysis for the different time windows only indicated anecdotal evidence that the time windows had a different prediction accuracy to one another.

The best model was therefore the RFR model which had similar prediction accuracy for the number of COVID-19 cases in the current week (T^0), the following week (T^{+1}) and in two weeks following (T^{+2}), irrespective whether log-transformed or non-transformed data were used as input.

Feature importance

The top 5 important features that contributed to the prediction of the T^0 RFR model were respectively the GT search terms "COVID-19" and "reuk", the two temporal lags and the local Moran I. See Supplementary Table 4 for the feature importance ranking for the RFR model for all age groups together.

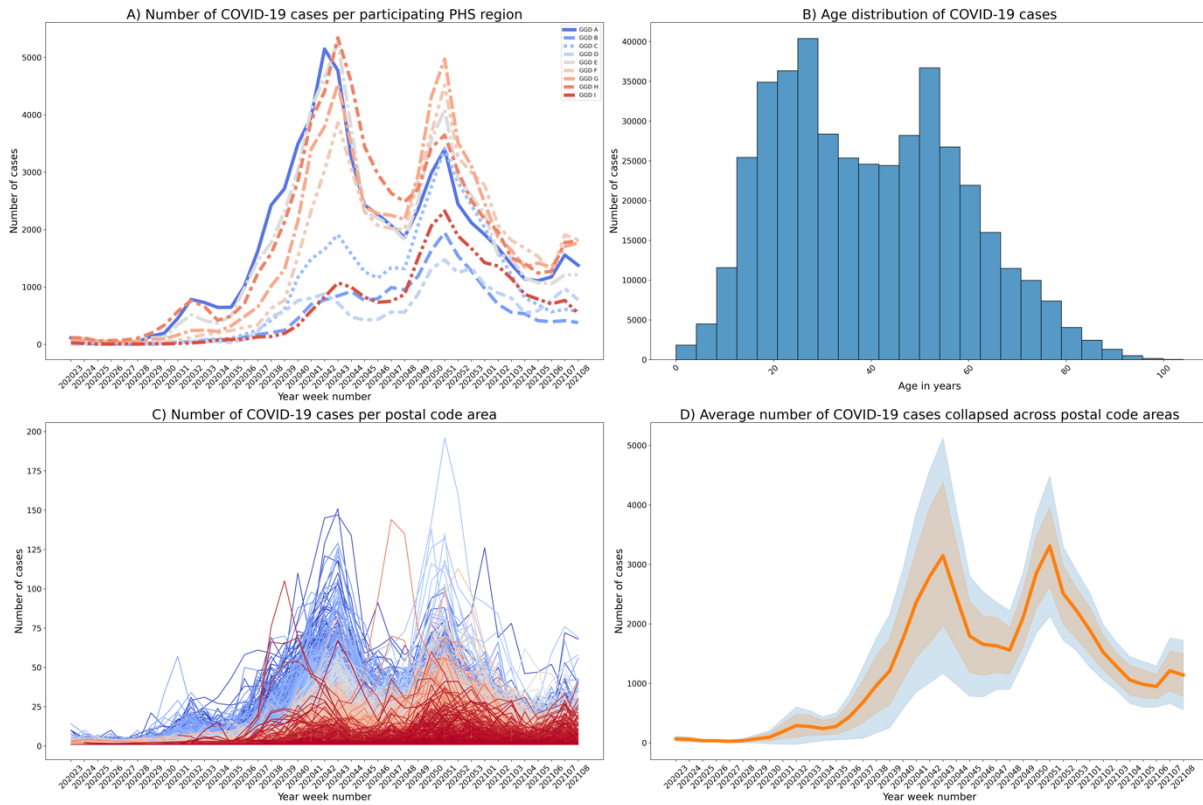
Model performance per age group

Subsequently, we stratified the prediction model with data on age groups. When training the model for the different age groups separately, visual inspection of the model fits indicated that for most age groups the RR and RFR models were able to capture the main trends (Figure 4 for training data and model fit for non-transformed data and Supplementary Figure 3 for log-transformed data).

Using a Bayesian ANOVA, we tested whether there was a statistical difference in predicting the top 100 postal codes between the different models (LOCF, RR, RFR), scales (non-transformed, log transformed), time windows (T^{+0} , T^{+1} , T^{+2}), and six age groups.

Similar to the age collapsed model, the LOCF model resulted in the lowest MSE, i.e., had more precise predictions. However, the RFR model had the highest percentage with correct prediction of the top 100 ranking. There were also

Figure 2: The number of COVID-19 cases over time by PHS region, age, and postal code area



A The number of COVID-19 cases per week per participating PHS region. The time period corresponds to the second COVID-19 infection outbreak within the Netherlands [51]. Note 1. The separate PHS regions are only shown for illustrative purposes and are not explicitly used to inform the model. Note 2. There are geographical temporal differences in the PHS regions where some have a strong first peak followed by a weaker second peak and vice versa [52]. B The age distribution of COVID-19 cases. There is a clear bi-modal distribution with a peak around the late 20s, early 30s and a second peak around the late 50s. C The number of COVID-19 cases per postal code area included in the current study. D The average number of COVID-19 cases over postal code areas which will be used to summarize the model performance. In dark orange the average number of COVID-19 cases per week number collapsed over postal code areas and age groups. The 95% confidence interval is shown in light orange and the first standard deviation is shown in blue.

clear effects of age group as 47% of the top 100 T^{+2} ranking for the 20–39-year-olds were correctly predicted using the RFR model, whereas the prediction accuracy dropped to 17% for the ≥ 80 group (Table 2 and Supplementary Table 2).

Model selection per age group

The primary output from the ANOVA is presented in Table 3 and Supplementary Table 3. The data were 3.64 times more likely under the model where the factors model type, time windows and age groups were included as well as a two-way interaction between all the individual factors and a three-way interaction term between model type, time windows and age group than under the second-best model where scale was included as a main factor ($BF_{10} = 1.0$ and $BF_{10} = 0.28$, respectively).

As the amount of evidence to prefer the best model over the second- or third-best model was moderate, an analysis of effects was conducted (Table 4). There was extreme evidence to include the factors model type, time window and age group (all $BF_{incl} > 100$). There was moderate evidence to exclude

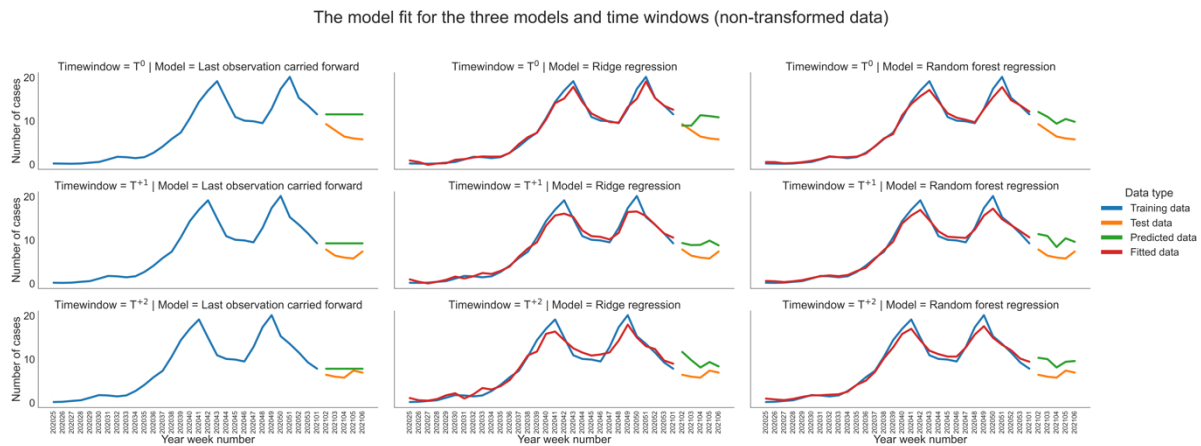
the factor scale ($BF_{incl}: 0.10$). For all the two-way interactions and single three-way interaction between type, time window and age group predictors there was extreme evidence to include them in the model (all $BF_{incl} > 100$).

Based on the *post hoc* tests for the main effect of model types, there was strong and anecdotal evidence that the RFR model outperformed the LOCF model ($BF_{10,u} = 11.14$) and RR model ($BF_{10,u} = 2.46$) (Table 5). The *post hoc* analysis for the different time windows did show anecdotal and moderate evidence that none of the time windows outperformed one another. Similarly, there was moderate evidence that the models fitted on the non-transformed data did not outperform models fitted to the log-transformed data ($BF_{10,u} = 0.21$).

The *post hoc* analysis showed that the prediction accuracy was highest in the age group 20–39 years (all $BF_{10,u} > 100$), followed by the age groups 40–64 years, and then 65–79, <12, and 12–19. The lowest prediction accuracy was found in the ≥ 80 -year-olds.

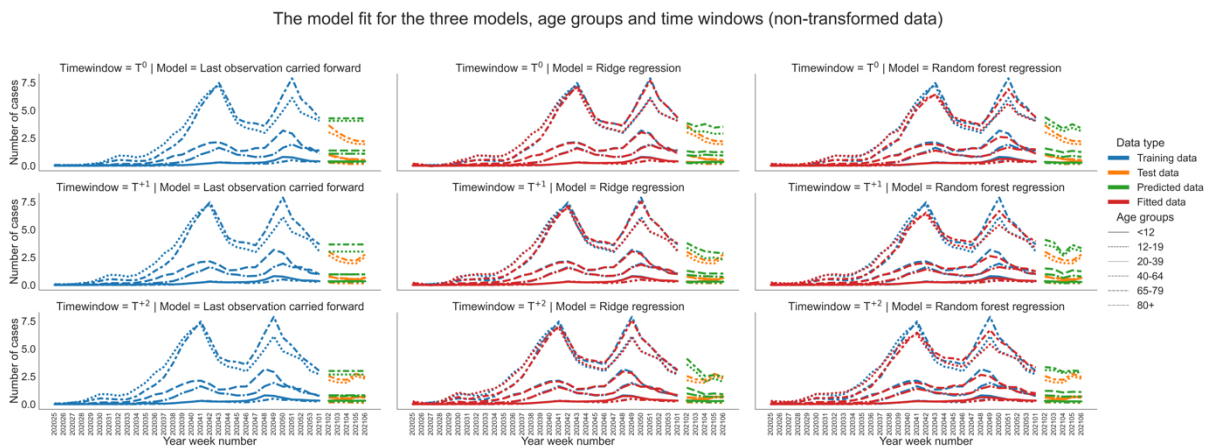
The best age-specific model was therefore a RFR model which had similar prediction accuracy for the number of COVID-19 cases in the current week (T^0), the following week

Figure 3: The mean training, test and predicted number of COVID-19 cases per model and time window for the non-transformed data



The first row shows the model fits for time window T^0 , the second row shows the model fits for time window T^{+1} (i.e., +1 week), and the third row shows the model fits for time window T^{+2} (i.e., +2 weeks). The separate columns correspond to the three different models used (resp. last observation carrier forward, ridge regression and random forest regression). The blue line represents the training data, the orange line shows the test data, the red line visualizes the model fit and finally the green line shows the model predictions. Since no actual model was fitted for the last observation carried forward model, the fitted data line is absent.

Figure 4: The mean training, test and predicted number of COVID-19 cases per model, time window and age group for the non-transformed data



The first row shows the model fits for time window T^0 , the second row shows the model fits for time window T^{+1} (i.e., +1 week), and the third row shows the model fits for time window T^{+2} (i.e., +2 weeks). The separate columns correspond to the three different models used (resp. last observation carrier forward, ridge regression and random forest regression). The blue line represents the training data, the orange line shows the test data, the red line visualizes the model fit and finally the green line shows the model predictions. Since no actual model was fitted for the last observation model, the fitted data line is absent.

(T^{+1}) and in the week following T^{+1} (T^{+2}), irrespective of log-transformed or non-transformed data. In this model, the prediction accuracy was highest for the 20-39 and 40-64 age groups. On average the prediction accuracy of the RFR model over the three time-windows for the age groups <12, 12-19, 20-39, 40-64, 65-79 and ≥ 80 were respectively 23%, 25%, 53%, 39%, 28% and 20%.

Feature importance

The top 5 important features that contributed to the prediction of the T^0 RFR model per age model were respectively the GT search terms “Reuk” and COVID-19, the first and second temporal lag for the age group 40-64 and

the local Moran I for the age group 20-39. See Supplementary Table 4 for the feature importance ranking for the RFR model per age groups.

Discussion

We set out to investigate whether a regression-based model could predict the intensity and characteristics of future COVID-19 outbreaks on the postal code level using several geospatial, governmental, demographic, and behavioural features. The results from the RFR model indicate that it was possible to predict 44% of the top 100 in terms of numbers of new cases on a group-level two-weeks in advance. Such a forecasting horizon could provide time for policy makers to

Table 2: The last observation carried forward, ridge regression, and random forest regression model performance for all age groups collapsed and per age group (non-transformed)

Model	Age (years)	Time window = T ⁰		Time window = T ⁺¹		Time window = T ⁺²	
		MSE mean (std)	Top 100 correctly predicted	MSE mean (std)	Top 100 correctly predicted	MSE mean (std)	predicted predicted
All age groups together							
LOCF		84.86 (53.40)	41%	56.11 (41.29)	33%	41.52 (35.07)	33%
RR		91.99 (39.07)	37%	97.44 (41.75)	30%	168.46 (134.17)	30%
RFR		117.87 (92.65)	49%	112.99 (61.61)	42%	110.97 (43.69)	44%
Per age group							
LOCF	<12	6.75 (4.68)	15%	4.95 (3.97)	13%	4.15 (3.72)	24%
	12–19		21%		20%		16%
	20–39		38%		46%		35%
	40–64		34%		32%		19%
	65–79		27%		22%		21%
	>=80		19%		18%		13%
RR	<12	6.91 (3.43)	17%	7.27 (3.31)	23%	9.92 (6.14)	26%
	12–19		21%		20%		27%
	20–39		39%		42%		30%
	40–64		37%		36%		26%
	65–79		22%		24%		26%
	>=80		19%		18%		15%
RFR	<12	7.88 (5.45)	18%	7.56 (3.68)	20%	7.47 (3.07)	30%
	12–19		29%		21%		26%
	20–39		58%		54%		47%
	40–64		41%		40%		37%
	65–79		27%		29%		29%
	>=80		24%		18%		17%

The average Mean Squared Error (MSE), percentage correctly predicted top 100 over all postal code areas and rank mismatch per model type and time window. The MSE indicates the average difference in observed and predicted values. The top 100 correctly predicted postal code areas indicate what percentage of predicted top 100 postal code areas with the highest number of COVID-19 cases were correctly predicted irrespective of absolute rank.

MSE: Mean Squared Error; LOCF: Last Observation Carried Forward; RR: Ridge Regression; RFR: Random Forest; std: standard deviation.

design and implement SCT prioritization on a geographical level. In addition, the results per age group indicate that the prediction accuracy for the categories 20–39 and 40–64 years was the highest.

There are however a number of limitations that caution overinterpretation of the results. A major limitation of our study and for using routine SARS-CoV-2 testing data to inform such models in general is that the test uptake differs between subgroups of the population [55]. It is likely that certain subgroups are underrepresented in the testing data as males, persons with diabetes, a lower education level, and older age tended to have a lower testing probability in the Netherlands [56]. In addition, the introduction of self-tests in May 2021, the age-specific adoption of self-tests, and the change in policy regarding testing might have also introduced certain data selection biases. Such underrepresentation might introduce unintended biases in the model, which could result in an underestimation of future caseloads in a given age group or postal code [57, 58]. Another important limitation is the focus on the postal code of residency and not explicitly accounting for employment location. In the Netherlands, the

majority of the working population commutes to a jobsite in a different postal code [59]. This daily commute likely adds an additional layer of complexity in predicting case load by residency postal code. Due to privacy concerns, employment information was not available for inclusion in the current analysis. To address this limitation, we included the stringency index as one of the features, since one of the first non-pharmaceutical interventions in the Netherlands was to close offices (with essential jobs exempted). During the study period, this restriction was partially lifted for a short period in the summer of 2021, allowing employees to return to their offices 50% of the time.

Unfortunately, the prediction accuracy for the oldest age groups, with a higher risk of developing adverse complications [60, 61], was the lowest. The predictive power of the age-based models were influenced by the sample size, and so, unfortunately, smaller age groups were more difficult to predict. These age-specific results illustrate that the developed model is ill-suited to prioritize SCT for the elderly. This makes the model incompatible with most COVID-19 containment policies in the Netherlands and other countries within the

Table 3: The last observation carried forward, ridge regression, and random forest regression model comparison predicting the top postal codes with the highest number of COVID-19 cases for all age groups together and per age group

Model	P(M)	P(M data)	BF _M	BF ₁₀	error %
All age groups together					
MT+S+TW+MT*S	0.053	0.264	6.469	1.000	**
MT+S+TW+MT*S+MT*TW	0.053	0.234	5.493	0.884	10.697
MT+S+TW+MT*S+S*TW	0.053	0.134	2.776	0.505	3.782
MT+S+TW+MT*S+MT*TW+S*TW	0.053	0.102	2.047	0.386	2.558
MT+S+TW	0.053	0.060	1.145	0.226	2.078
MT+TW	0.053	0.058	1.117	0.221	1.737
MT+S+TW+MT*S+MT*TW+S*TW+MT*S*TW	0.053	0.056	1.060	0.210	5.480
MT+S+TW+MT*TW	0.053	0.027	0.491	0.100	3.165
MT+S+TW+S*TW	0.053	0.026	0.480	0.098	2.976
MT+TW+MT*TW	0.053	0.024	0.434	0.089	1.793
MT+S+TW+MT*TW+S*TW	0.053	0.011	0.201	0.042	2.081
MT	0.053	0.002	0.044	0.009	1.615
MT+S	0.053	0.001	0.025	0.005	1.850
MT+S+MT*S	0.053	9.90E-01	0.018	0.004	1.877
TW	0.053	1.15E-01	0.002	4.36E-01	1.615
Nullmodel	0.053	9.19E-02	0.002	3.48E-01	1.615
S+TW	0.053	5.87E-02	0.001	2.22E-01	1.975
S	0.053	4.51E-02	8.11E-01	1.71E-01	1.615
S+TW+S*TW	0.053	2.30E-02	4.13E-01	8.68E-02	1.983
Per age group (first 10 models)					
MT+TW+AG+MT*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.554	205.986	1.000	**
MT+S+TW+AG+MT*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.152	29.806	0.275	11.160
MT+S+TW+AG+MT*TW+MT*AG+S*AG+TW*AG+MT*TW*AG	0.006	0.113	21.200	0.205	3.150
MT+S+TW+AG+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*TW*AG	0.006	0.037	6.453	0.068	20.527
MT+S+TW+AG+MT*TW+S*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.035	6.013	0.063	3.965
MT+S+TW+AG+MT*S+MT*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.023	3.832	0.041	4.393
MT+S+TW+AG+MT*S+MT*TW+MT*AG+S*AG+TW*AG+MT*TW*AG	0.006	0.018	3.114	0.033	2.981
MT+S+TW+AG+MT*S+MT*TW+MT*AG+S*AG+TW*AG+MT*S*AG+MT*TW*AG	0.006	0.018	2.967	0.032	2.884
MT+S+TW+AG+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*TW*AG+S*TW*AG	0.006	0.012	1.981	0.021	3.353
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.009	1.511	0.016	40.181

MT: Model Type; TW: Time Window; S: non-transformed or log Scale; AG: Age group; P(M): Prior model probability; P(M|data): posterior model probability; BF_M: the change from prior odds to posterior odds; BF₁₀: the Bayes factor relative to the best model; error %: indicates the precision of the numerical approximation; * indicates an interaction term. ** As the estimated Bayes factors are relative to the best performing model, there is no relative error associated with the best performing model.

EU as they were aimed at protecting vulnerable groups, such as the elderly [62]. However, from a SCT perspective, these results might be promising given that young adults and middle aged age groups have been shown to contribute

disproportionally to COVID-19 transmission relative to their population size [63] and tend to be the main drivers of superspreading events [64]. Relatedly, due to the inherent biases of natural experiments, it is challenging to estimate

Table 4: Analysis of effect predicting the top 100 postal codes with the highest number of COVID-19 cases for all age groups together and per age group. The prior and posterior inclusion probabilities, and the inclusion BF for each predictor

Effects	P(incl)	P(excl)	P(incl data)	P(excl data)	BF _{incl}
All age groups together					
MT	0.737	0.263	1.000	3.34E-01	1,069.135
S	0.737	0.263	0.915	0.085	3.864
TW	0.737	0.263	0.995	0.005	71.975
MT*S	0.316	0.684	0.790	0.210	8.175
MT*TW	0.316	0.684	0.453	0.547	1.792
S*TW	0.316	0.684	0.328	0.672	1.059
MT*S*TW	0.053	0.947	0.056	0.944	1.060
Per age group					
MT	0.886	0.114	1.000	5.22E-12	2.46E+16
S	0.886	0.114	0.446	0.554	0.103
TW	0.886	0.114	1.000	5.44E-12	2.36E+16
MT*S	0.503	0.497	0.097	0.903	0.106
MT*TW	0.503	0.497	1.000	1.19E-02	82,990.993
S*TW	0.503	0.497	0.122	0.878	0.138
MT*S*TW	0.120	0.880	0.012	0.988	0.092
AG	0.886	0.114	1.000	5.22E-12	2.46E+16
MT*AG	0.503	0.497	1.000	9.08E-06	1.09E+11
S*AG	0.503	0.497	0.222	0.778	0.282
TW*AG	0.503	0.497	1.000	5.66E-12	1.75E+17
MT*S*AG	0.120	0.880	0.031	0.969	0.237
MT*TW*AG	0.120	0.880	1.000	4.36E-02	168,494.483
S*TW*AG	0.120	0.880	0.020	0.980	0.150
MT*S*TW*AG	0.006	0.994	7.12E-01	0.999	0.118

MT: Model Type; TW: Time Window; S: non-transformed or log Scale; AG: age group; P(incl): prior inclusion probability and is computed by summing up the prior model probabilities of all models which contain a given predictor [53]; P(excl): prior exclusion probability and is computed by summing up the posterior model probabilities of all models which contain a given predictor (53); P(Incl|data): posterior inclusion probability; P(excl|data): posterior exclusion probability; BF_{incl}: the inclusion Bayes factor.

the effects of employing SCT prioritization on the observed caseloads. As a result, it is uncertain how the RFR model would perform in future outbreaks in absence of the previously employed SCT prioritization.

At the time of writing, four years after the first confirmed COVID-19 case within the Netherlands, it remains challenging to share (medical) data for scientific purposes between PHS regions. Within the current study it was not feasible to readily expand the number of participating regions nor was it trivial to get ethical approval for a longer time window to analyse due to the large number of organisations involved. It was therefore not possible to study whether the results observed also apply to later epidemiological waves of the pandemic. Given these challenging conditions, it is a notable strength of this study that data from nine different PHS regions were able to be combined. In future, it is advisable to develop or adopt a (inter)national collaborative data infrastructure (such as The European Surveillance System (TESSy) [65]) which has clear data governance and GDPR compliant regulations to ensure that data usage is facilitated. Indeed, pandemic preparedness efforts within the Netherlands should be aligned with the European Health Emergency Preparedness and Response Authority (HERA) and the IT system ATHINA to rapidly detect and respond to (cross-border) health emergencies and facilitate knowledge and data exchange [64]. As noted, the

use of mathematical models to predict the development of the COVID-19 pandemic is widespread, but many of these models suffer from numerous biases and limitations [66, 67]. Indeed, one can argue that forecasting models, such as the ones used here, generally have a limited role in the containment phase of an epidemic, as data points are still scarce and the underlying dynamics are largely unknown [68, 69]. Indeed, we were not able to include any cases from the initial phase of the pandemic as large-scale testing in the Netherlands only became available in June 2020 [70]. Conventional models such as linear regression, albeit less precise than advanced machine learning models when large amounts of data are available [71–73], might be of better value in the early stages of an outbreak as these are less data hungry [74]. Other areas where models did have had a positive impact on informing policy was by calculating the effective reproduction number (R_{eff}) [75, 76].

Promising features and previous modelling work

A substantial number of recent studies have used geospatial and socio-demographic information to predict COVID-19 cases using a number of different machine learning model classes (e.g., deep learning models such as convoluted

Table 5: Post hoc comparisons between model types, scale and time windows for predicting the top 100 postal codes with the highest number of COVID-19 cases for all age groups together and per age group

Comparison			Prior odds	Posterior odds	BF _{10, U}	error %
All age groups together						
Model types	LOCF	RFR	0.587	12.770	21.740	1.93E-03
	LOCF	RR	0.587	0.275	0.468	4.65E-01
	RFR	RR	0.587	9.162	15.598	3.36E-03
Scale	Log	Non-transformed	1.000	0.491	0.491	0.001
Time windows	T ⁺¹	T ⁺²	0.587	0.286	0.486	5.71E-01
	T ⁺¹	T ⁰	0.587	1.537	2.616	0.009
	T ⁺²	T ⁰	0.587	0.680	1.158	0.005
Per age group						
Model types	LOCF	RFR	0.587	6.544	11.141	3,03E-04
	LOCF	RR	0.587	0.212	0.361	0.013
	RFR	RR	0.587	1.447	2.463	0.009
Scale	Log	Non-transformed	1.000	0.205	0.205	0.029
Time windows	T ⁺¹	T ⁺²	0.587	0.167	0.284	0.014
	T ⁺¹	T ⁰	0.587	0.153	0.261	0.014
	T ⁺²	T ⁰	0.587	0.234	0.399	0.013
Age groups (years)	<12	12-19	0.260	0.229	0.879	0.005
	<12	20-39	0.260	1.82E+11	7.01E+11	6.10E-11
	<12	40-64	0.260	22,047.112	84,822.342	6.36E-07
	<12	65-79	0.260	14.006	53.885	5.70E-05
	<12	>=80	0.260	0.188	0.723	0.005
	12-19	20-39	0.260	1.39E+10	5.33E+10	5.72E-09
	12-19	40-64	0.260	663.347	2,552.108	1.04E-05
	12-19	65-79	0.260	0.439	1.687	0.007
	12-19	>=80	0.260	5.578	21.459	1.64E-03
	20-39	40-64	0.260	47.266	181.846	1.38E-02
	20-39	65-79	0.260	1.47E+09	5.64E+09	6.03E-08
	20-39	>=80	0.260	1.33E+13	5.11E+13	5.29E-11
	40-64	65-79	0.260	25.346	97.516	5.69E-04
	40-64	>=80	0.260	6.13E+09	2.36E+10	1.45E-08
	65-79	>=80	0.260	20,328.035	78,208.499	7.05E-07

LOCF: Last Observation Carried Forward; RR: Ridge Regression; RFR: Random Forest; The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons [54]. Individual comparisons are based on the default t-test with a Cauchy ($0, 1 = \frac{1}{\sqrt{2}}$) prior. The "U" in the Bayes factor denotes that it is uncorrected.

neural networks, extreme gradient boosting and random forest regression models [77–93]). The feature importance analysis highlighted three interesting results: the geospatial autocorrelation between postal codes was informative for the model to predict the number of COVID-19 cases; demographic characteristics such as number of households in each postal code were not; that the feature importance differed per age group. Complimentary to the features which were included in the current study, several promising features have emerged such as mobility [94, 95] and climate [96, 97] that could be included in future work. In light of pandemic preparedness, future work should include ablation studies to identify which features are most informative in predicting COVID-19 cases. Another line of promising work is to solely focus on predicting COVID-19 cases in long-term care facilities as the inhabitants have an increased vulnerability to COVID-19 [98–100, e.g.].

The resulting insights will be crucial in designing and preparing effective data collection strategies.

The current study contributes to this body of work by highlighting the differences in model performance for the different age groups and stresses the need to have age-specific models when predicting future COVID-19 cases. Direct comparison in model performance of the current study with previous studies is however challenging as data types, time windows, geospatial scale, modelling approaches, and employed (accuracy) metrics differ substantially. Despite these differences, the reported performance of the various deep learning models suggests potential for enhancing the accuracy of COVID-19 case prediction compared to the RFR, with one crucial caveat. An inherent discrepancy exists between the early phase of an emerging epidemic, when SCT is thought to be most effective, and the limited data available for data-intensive modelling techniques like RFR and, particularly, deep

learning models [74, 101]. As a result, existing epidemic forecast models tend to have large margins of uncertainty, limiting the direct applicability for informing policy in the early stages of an epidemic [69, 102, 103]. This is highlighted in the current study as we were only able to predict approximately 44% of the top 100 outbreak regions two-weeks in advance. A prediction accuracy which we deem to be insufficient to warrant the deployment of the RFR model in the field. 6

Conclusion

In conclusion, it is possible to use machine learning methods to predict new infections during the early phases of an outbreak, but with a large margin of uncertainty. This is due to the insufficient number of datapoints available in this early phase. Therefore, results should be handled with caution when such models are used to inform policy during the early phases of outbreak. Especially given that the current results indicate that vulnerable groups were the hardest to predict. Regardless of which model is used, outbreak management would benefit from having high quality standardized data from the onset. This would require preferably a data platform already in place that facilitates the exchange of information between stakeholders [65, 104].

Acknowledgements

We would like to acknowledge Suzan van Dijken, Ymke Hamelink, Rosa van Hoorn, Rana Jajou, Kevin Konings, Jeroen Kuiper, Bram Meima, Feiko Ritsema, and Juul Tönis for contributing to the process of data acquisition.

Funding

This research is part of the project CONTROL and is financially supported by ZonMw (#10430022010022). There was no additional external funding received for this study.

Competing interests

The authors have the following competing interests: The institution of M. F. Schim van der Loeff received study funding for an investigator-initiated study from GSK; he served on advisory boards of MSD and Novosan. There are no patents, products in development or marketed products associated with this research to declare.

Author contribution

Conceptualization: MCK, JRB, AM, EMdB, IKJ, NHTMD-M, GvR, HMG, IEG, MWFP, CFHR, SvdH, KVCWdB, MFSvdL; Data curation: MCK, JRB, AM; Formal analysis: MCK; Funding acquisition: AM, EMdB, IKJ, NHTMD-M, GvR, HMG, IEG, MWFP, CFHR, SvdH, KVCWdB, MFSvdL; Investigation: MCK, JRB, AAM; Methodology: MCK; Project administration: AM; Supervision: MCK, JRB, AM; Visualization: MCK; Writing – Original Draft

Preparation: MCK, JRB, AM, AB; Writing – Review & Edition: EMdB, IKJ, NHTMD-M, GvR, HMG, IEG, MWFP, CFHR, SvdH, KVCWdB, MFSvdL.

Open science & data availability statement

All code used to pseudonymize, process and analyze the data are available on OSF (https://osf.io/khcb8/?view_only=fafa72443a704209814dd517f1cab7e9). The aggregated PHS data on postal code and week number used to train the models are available on the above OSF page. To ensure that the shared data does not contain sensitive medical information that can be linked to an individual, any cell that contained less than ten observations was set to 0. This restriction was imposed by the privacy office of the hosting organization. A request for the non-filtered data can be send to mbooi@ggd.amsterdam.nl.

References

1. Eames KTD, Keeling MJ. Contact tracing and disease control. *Proc R Soc Lond B*. 2003 Dec 22;270(1533):2565–71. <https://doi.org/10.1098/rspb.2003.2554>
2. Götz HM, van Doornum G, Niesters HG, den Hollander JG, Thio HB, de Zwart O. A cluster of acute hepatitis C virus infection among men who have sex with men – results from contact tracing and public health implications. *AIDS*. 2005 Jun 10;19(9):969–74. <https://doi.org/10.1097/01.aids.0000171412.61360.f8>
3. Koster B, Borgen K, Meijer H, van der Plas S, Kuyvenhoven V. Large scale contact tracing after a case of open tuberculosis in a supermarket, the Netherlands, January - February 2005. *Weekly releases*. 2005;10:2645. <https://doi.org/10.2807/esw.10.08.02648-en>
4. BE, Götz HM, van Gorp ECM, Verbon A, Rokx C, Boucher CAB, et al. Partner Notification for Reduction of HIV-1 Transmission and Related Costs among Men Who Have Sex with Men: A Mathematical Modeling Study. Khudyakov YE, editor. *PLoS ONE*. 2015 Nov 10;10(11):e0142576. <https://doi.org/10.1371/journal.pone.0142576>
5. Rouvoet A. Roadmap Testen, Traceren, Vaccineren Januari-maart 2021. GGD GHOR; 2021. Available from: <https://ggdghor.nl/wp-content/uploads/2021/02/Aanbiedingsbrief-en-RoadmapQ1TestenTracerenVaccineren.pdf>.
6. Lash RR, Moonan PK, Byers BL, Bonacci RA, Bonner KE, Donahue M, et al. COVID-19 Case Investigation and Contact Tracing in the US, 2020. *JAMA Netw Open*. 2021 Jun 3;4(6):e2115850. <https://doi.org/10.1001/jamanetworkopen.2021.15850>
7. Fetzer T, Graeber T. Measuring the scientific effectiveness of contact tracing: Evidence from a

- natural experiment. *Proc Natl Acad Sci USA*. 2021 Aug 17;118(33):e2100814118. <https://doi.org/10.1073/pnas.2100814118>
8. Pourghasemi HR, Pouyan S, Heidari B, Farajzadeh Z, Fallah Shamsi SR, Babaei S, et al. Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *International Journal of Infectious Diseases*. 2020 Sep;98:90–108. <https://doi.org/10.1016/j.ijid.2020.06.058>
9. Guliyev H. Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics*. 2020 Aug;38:100443. <https://doi.org/10.1016/j.spasta.2020.100443>
10. Ting RSK, Aw Yong YY, Tan MM, Yap CK. Cultural Responses to Covid-19 Pandemic: Religions, Illness Perception, and Perceived Stress. *Front Psychol*. 2021 Jul 23;12:634863. <https://doi.org/10.3389/fpsyg.2021.634863>
11. Thomas LJ, Huang P, Yin F, Luo XI, Almquist ZW, Hipp JR, et al. Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. *Proc Natl Acad Sci USA*. 2020 Sep 29;117(39):24180–7. <https://doi.org/10.1073/pnas.2011656117>
12. Christensen T. The Social Policy Response to COVID-19 – The Failure to Help Vulnerable Children and Elderly People. *Public Organiz Rev*. 2021 Dec;21(4):707–22. <https://doi.org/10.1007/s11115-021-00560-2>
13. Vink M, Iglói Z, Fanoy EB, van Beek J, Boelsums T, de Graaf M, et al. Community-based SARS-CoV-2 testing in low-income neighbourhoods in Rotterdam: Results from a pilot study. *J Glob Health*. 2022 Oct 1;12:05042. <https://doi.org/10.7189/jogh.12.05042>
14. Heijmink L, Tönis J, Gilhuis N, Gerkema M, Hart L, Raven S, et al. Epidemiological evaluation of mass testing in a small municipality in the Netherlands during the SARS-CoV-2 epidemic. *Epidemiol Infect*. 2022;150:e193. <https://doi.org/10.1017/S0950268822001777>
15. Rijksinstituut voor volksgezondheid en milieu (RIVM). RIVM.nl. 2022. Invloed van prikbusen op vaccinatiegraad. Available from: <https://www.rivm.nl/gedragsonderzoek/invloed-van-prikbusen-op-vaccinatiegraad>
16. Koster T. anDrea - A Digital Research Environment. 2021.
17. van Rossum G, de Boer J. Interactively testing remote servers using the python programming language. *CWI Quarterly*. 1991;4(4):283–304.
18. Dhillon RA, Qamar MA, Gilani JA, Irfan O, Waqar U, Sajid MI, et al. The mystery of COVID-19 reinfections: A global systematic review and meta-analysis. *Annals of Medicine and Surgery*. 2021 Dec;72:103130. <https://doi.org/10.1016/j.amsu.2021.103130>
19. Wei WWS. *Time Series Analysis*. Oxford; 2013. (The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis; vol. 2). <https://doi.org/10.1093/oxfordhb/9780199934898.013.0022>
20. Seabold S, Perktold J. *Statsmodels: Econometric and Statistical Modeling with Python*. In Austin, Texas; 2010. p. 92–6. <https://doi.org/10.25080/Majora-92bf1922-011>
21. Shaman P. Generalized Levinson–Durbin sequences, binomial coefficients and autoregressive estimation. *Journal of Multivariate Analysis*. 2010 May;101(5):1263–73. <https://doi.org/10.1016/j.jmva.2010.01.004>
22. Hale T, Angrist N, Hale AJ, Kira B, Majumdar S, Petherick A, et al. Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves. Seale H, editor. *PLoS ONE*. 2021 Jul 9;16(7):e0253116. <https://doi.org/10.1371/journal.pone.0253116>
23. Mavragani A, Gkillas K. COVID-19 predictability in the United States using Google Trends time series. *Sci Rep*. 2020 Dec;10(1):20693. <https://doi.org/10.1038/s41598-020-77275-9>
24. Prasanth S, Singh U, Kumar A, Tikkiwal VA, Chong PHJ. Forecasting spread of COVID-19 using google trends: A hybrid GWO-deep learning approach. *Chaos, Solitons & Fractals*. 2021 Jan;142:110336. <https://doi.org/10.1016/j.chaos.2020.110336>
25. Hogue J, DeWilde B. *PyTrends*. 2022. Available from: <https://github.com/GeneralMills/pytrends>
26. Bosdriesz JR, Ritsema F, Leenstra T, Petrignani MWF, Bruisten SM, Coyer L, et al. Self-reported symptoms as predictors of SARS-CoV-2 infection in the general population living in the Amsterdam region, the Netherlands. Yon DK, editor. *PLoS ONE*. 2022 Jan 28;17(1):e0262287. <https://doi.org/10.1371/journal.pone.0262287>
27. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. *Int Forum Allergy Rhinol*. 2020 Jul;10(7):839–47. <https://doi.org/10.1002/alr.22580>
28. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data*. Wiley; 2004. (Wiley Series in Probability and Statistics). <https://doi.org/10.1002/0471662682>
29. Giuliani D, Dickson MM, Espa G, Santi F. Modelling and predicting the spatio-temporal spread of COVID-19 in Italy. *BMC Infect Dis*. 2020 Dec;20(1):700. <https://doi.org/10.1186/s12879-020-05415-7>
30. Moran PAP. A test for the serial independence of residuals. *Biometrika*. 1950;37(1–2):178–81. <https://doi.org/10.1093/biomet/37.1-2.178>
31. Rey SJ, Anselin L. *PySAL: A Python Library of Spatial Analytical Methods*. The Review of Regional Studies. 2007;37(1):23. <https://doi.org/10.52324/001c.8285>

32. Suryowati K, Bektı RD, Faradila A. A Comparison of Weights Matrices on Computation of Dengue Spatial Autocorrelation. IOP Conf Ser: Mater Sci Eng. 2018 Apr;335:012052. <https://doi.org/10.1088/1757-899X/335/1/012052>
33. Regio en Ruimte C. CBS.nl. 2022 [cited 2023 Jan 4]. GGD-regio's naar postcode 2020. Available from: <https://www.cbs.nl/nl-nl/maatwerk/2020/47/ggd-regio-s-naar-postcode-2020>.
34. House T, Keeling MJ. Household structure and infectious disease transmission. Epidemiol Infect. 2009 May;137(5):654–61. <https://doi.org/10.1017/S0950268808001416>
35. Tarwater PM, Martin CF. Effects of population density on the spread of disease. Complexity. 2001 Jul;6(6):29–36. <https://doi.org/10.1002/cplx.10003>
36. Breslow NE, Day NE. Statistical methods in cancer research: Volume 2. The design and analysis of cohort studies. Journal of Epidemiology & Community Health. 1989 Mar 1;43(1):92–3. <https://doi.org/10.1002/cplx.10003>
37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825–30. <https://doi.org/10.5555/1953048.2078195>
38. Petropoulos F, Apiletti D, Assimakopoulos V, Babai MZ, Barrow DK, Ben Taieb S, et al. Forecasting: theory and practice. International Journal of Forecasting. 2022 Jul;38(3):705–871. <https://doi.org/10.1016/j.ijforecast.2021.11.001>
39. Borchani H, Varando G, Bielza C, Larrañaga P. A survey on multi-output regression: Multi-output regression survey. WIREs Data Mining Knowl Discov. 2015 Sep;5(5):216–33. <https://doi.org/10.1002/widm.1157>
40. Ekwaru JP, Veugelers PJ. The Overlooked Importance of Constants Added in Log Transformation of Independent Variables with Zero Values: A Proposed Approach for Determining an Optimal Constant. Statistics in Biopharmaceutical Research. 2018 Jan 2;10(1):26–9. <https://doi.org/10.1080/19466315.2017.1369900>
41. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 2000 Feb;42(1):80–6. <https://doi.org/10.1080/00401706.2000.10485983>
42. Scikit-Learn. Scikit-learn.org. 2024 [cited 2024 Jan 15]. 1.1.2.1 Regression. Available from: https://scikit-learn.org/stable/modules/linear_model.html#ridge-regression.
43. Breiman L. Random Forests. Machine Learning. 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>
44. Grömping U. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. The American Statistician. 2009 Nov;63(4):308–19. <https://doi.org/10.1198/tast.2009.08199>
45. Scikit-Learn. Scikit-learn.org. 2024 [cited 2024 Jan 15]. 1.10.6. Tree algorithms: ID3, C4.5, C5.0 and CART. Available from: <https://scikit-learn.org/stable/modules/tree.html#tree-algorithms-id3-c4-5-c5-0-and-cart>.
46. JASP T. JASP. 2020.
47. Morey RD, Rouder JN. BayesFactor. 2015.
48. Rouder JN, Morey RD, Speckman PL, Province JM. Default Bayes factors for ANOVA designs. Journal of Mathematical Psychology. 2012 Oct;56(5):356–74. <https://doi.org/10.1016/j.jmp.2012.08.001>
49. Jeffreys SH. The theory of probability. 3rd ed. Oxford University Press;
50. Wetzels R, Matzke D, Lee MD, Rouder JN, Iverson GJ, Wagenmakers EJ. Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. Perspect Psychol Sci. 2011 May;6(3):291–8. <https://doi.org/10.1177/1745691611406923>
51. CBS. cbs.nl. 2021. In tweede golf overleden bijna 11 duizend meer mensen dan verwacht. Available from: <https://www.cbs.nl/nl-nl/nieuws/2021/27/in-tweede-golf-overleden-bijna-11-duizend-meer-mensen-dan-verwacht>.
52. RIVM. RIVM.nl. 2022. Resultaat Pienter Corona Onderzoek Ronde 6. Available from: <https://www.rivm.nl/pienter-corona-onderzoek/resultaten/ronde-6>.
53. van den Bergh D, van Doorn J, Marsman M, Draws T, van Kesteren EJ, Derks K, et al. A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. L'Année Psychologique. 2020;120:73–96. <https://doi.org/10.3917/anpsy1.201.0073>
54. Westfall PH, Johnson WO, Utts JM. A Bayesian Perspective on the Bonferroni Adjustment. Biometrika. 1997;84(2):419–27. Available from: <http://www.jstor.org/stable/2337467>
55. Rijksinstituut voor volksgezondheid en milieu (RIVM). RIVM.nl. 2022. Testen op corona: een overzicht van testgedrag. Available from: <https://www.rivm.nl/gedragsonderzoek/testen-op-corona>.
56. McDonald SA, Soetens LC, Schipper CMA, Friesema I, van den Wijngaard CC, Teirlinck A, et al. Testing behaviour and positivity for SARS-CoV-2 infection: insights from web-based participatory surveillance in the Netherlands. BMJ Open. 2021 Dec;11(12):e056077. <https://doi.org/10.1136/bmjopen-2021-056077>

57. Bastos SB, Morato MM, Cajueiro DO, Normey-Rico JE. The COVID-19 (SARS-CoV-2) uncertainty tripod in Brazil: Assessments on model-based predictions with large under-reporting. *Alexandria Engineering Journal*. 2021 Oct;60(5):4363–80. <https://doi.org/10.1016/j.aej.2021.03.004>
58. Alballa N, Al-Turaiki I. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review. *Informatics in Medicine Unlocked*. 2021;24:100564. <https://doi.org/10.1016/j.imu.2021.100564>
59. CBS. [www.cbs.nl. Meer dan de helft van de werknemers is forens](https://www.cbs.nl/nl-nieuws/2013/23/meer-dan-de-helft-van-de-werknemers-is-forens). Available from: <https://www.cbs.nl/nl-nieuws/2013/23/meer-dan-de-helft-van-de-werknemers-is-forens>.
60. Tiruneh SA, Tesema ZT, Azanaw MM, Angaw DA. The effect of age on the incidence of COVID-19 complications: a systematic review and meta-analysis. *Syst Rev*. 2021 Dec;10(1):80. <https://doi.org/10.1186/s13643-021-01636-2>
61. Zhang H, Wu Y, He Y, Liu X, Liu M, Tang Y, et al. Age-Related Risk Factors and Complications of Patients With COVID-19: A Population-Based Retrospective Study. *Front Med*. 2022 Jan 11;8:757459. <https://doi.org/10.3389/fmed.2021.757459/full>
62. Copenhagen: WHO Regional Office for Europe and Stockholm: European Centre for Disease Prevention and Control. COVID-19 Contact tracing: country experiences and way forward. 2022. Available from: <https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-contact-tracing-report-ECDC-WHO-EURO.pdf>
63. Monod M, Blenkinsop A, Xi X, Hebert D, Bershan S, Tietze S, et al. Age groups that sustain resurging COVID-19 epidemics in the United States. *Science*. 2021 Mar 26;371(6536):eabe8372. <https://doi.org/10.1126/science.abe8372>
64. Lau MSY, Grenfell B, Thomas M, Bryan M, Nelson K, Lopman B. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proc Natl Acad Sci USA*. 2020 Sep 8;117(36):22430–5. <https://doi.org/10.1073/pnas.2011802117>
65. ECDC. ECDC. The European Surveillance System (TESSy). Available from: <https://www.ecdc.europa.eu/en/publications-data/european-surveillance-system-tessey>.
66. Heneghan CJ, Jefferson T. Why COVID-19 modelling of progression and prevention fails to translate to the real-world. *Advances in Biological Regulation*. 2022 Dec;86:100914. <https://doi.org/10.1016/j.jbior.2022.100914>
67. Lucas RE. Econometric policy evaluation: a critique. In: *The Phillips Curve and Labor Markets*. North-Holland Pub. Co.; 1976. p. 19–46. (Carnegie-Rochester Conference Series on Public Policy; vol. 1).
68. Adiga A, Chen J, Marathe M, Mortveit H, Venkatramanan S, Vullikanti A. Data-Driven Modeling for Different Stages of Pandemic Response. *J Indian Inst Sci*. 2020 Oct;100(4):901–15. <https://doi.org/10.1007/s41745-020-00206-0>
69. Rosenfeld R, Tibshirani RJ. Epidemic tracking and forecasting: Lessons learned from a tumultuous year. *Proc Natl Acad Sci USA*. 2021 Dec 21;118(51):e2111456118. <https://doi.org/10.1073/pnas.2111456118>
70. Ministerie van Justitie en Veiligheid. www.rijksoverheid.nl. 2023. Ontwikkelingen coronavirus in 2020. Available from: <https://www.rijksoverheid.nl/onderwerpen/coronavirus-tijdslijn/2020>
71. Fernandez-Delgado M, Cernadas E, Barro S, Amorim D. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? <https://doi.org/10.5555/2627435.2697065>
72. Osisanwo FY, Akinsola JET, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J. Supervised Machine Learning Algorithms: Classification and Comparison. *IJCTT*. 2017 Jun 25;48(3):128–38. <https://doi.org/10.14445/22312803/IJCTT-V48P126>
73. Ye J, Hua M, Zhu F. Machine Learning Algorithms are Superior to Conventional Regression Models in Predicting Risk Stratification of COVID-19 Patients. *RMHP*. 2021 Jul;Volume 14:3159–66. <https://doi.org/10.2147/RMHP.S318265>
74. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014 Dec;14(1):137. <https://doi.org/10.1186/1471-2288-14-137>
75. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B*. 2007 Feb 22;274(1609):599–604. <https://doi.org/10.1098/rspb.2006.3754>
76. Rijksoverheid. coronadashboard [rijksoverheid.nl](https://coronadashboard.rijksoverheid.nl/landelijk/reproductiegetal#). 2020 [cited 2023 May 23]. Coronadashboard - Ontwikkeling van het virus - Reproductiegetal. Available from: <https://coronadashboard.rijksoverheid.nl/landelijk/reproductiegetal#>
77. Ahasan R, Alam MdS, Chakraborty T, Hossain MdM. Applications of GIS and geospatial analyses in COVID-19 research: A systematic review. *F1000Res*. 2022 Jan 28;9:1379. <https://doi.org/10.12688/f1000research.27544.2>
78. Fatima M, O'Keefe KJ, Wei W, Arshad S, Gruebner O. Geospatial Analysis of COVID-19: A

- Scoping Review. IJERPH. 2021 Feb 27;18(5):2336. <https://doi.org/10.3390/ijerph18052336>
79. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. A review. Science of The Total Environment. 2020 Oct;739:140033. <https://doi.org/10.1016/j.scitotenv.2020.140033>
80. Zhao Y, Hu M, Jin Y, Chen F, Wang X, Wang B, et al. Predicting the transmission trend of respiratory viruses in new regions via geospatial similarity learning. International Journal of Applied Earth Observation and Geoinformation. 2023 Dec;125:103559. <https://doi.org/10.1016/j.jag.2023.103559>
81. Yahya BM, Yahya FS, Thannoun RG. COVID-19 prediction analysis using artificial intelligence procedures and GIS spatial analyst: a case study for Iraq. Appl Geomat. 2021 Sep;13(3):481–91. <https://doi.org/10.1007/s12518-021-00365-4>
82. Ward T, Johnsen A, Ng S, Chollet F. Forecasting SARS-CoV-2 transmission and clinical risk at small spatial scales by the application of machine learning architectures to syndromic surveillance data. Nat Mach Intell. 2022 Oct 21;4(10):814–27. <https://doi.org/10.1038/s42256-022-00538-9>
83. Vahedi B, Karimzadeh M, Zoraghein H. Spatiotemporal prediction of COVID-19 cases using inter- and intra-county proxies of human interactions. Nat Commun. 2021 Nov 8;12(1):6440. <https://doi.org/10.1038/s41467-021-26742-6>
84. Galasso J, Cao DM, Hochberg R. A random forest model for forecasting regional COVID-19 cases utilizing reproduction number estimates and demographic data. Chaos, Solitons & Fractals. 2022 Mar;156:111779. <https://doi.org/10.1016/j.chaos.2021.111779>
85. Lucas B, Vahedi B, Karimzadeh M. A spatiotemporal machine learning approach to forecasting COVID-19 incidence at the county level in the USA. Int J Data Sci Anal. 2023 Apr;15(3):247–66. <https://doi.org/10.1007/s41060-021-00295-9>
86. Dairi A, Harrou F, Zeroual A, Hittawe MM, Sun Y. Comparative study of machine learning methods for COVID-19 transmission forecasting. Journal of Biomedical Informatics. 2021 Jun;118:103791. <https://doi.org/10.1016/j.jbi.2021.103791>
87. Mir SA, Bhat MS, Rather GM, Mattoo D. Role of big geospatial data in the COVID-19 crisis. In: Data Science for COVID-19. Elsevier; 2022. p. 589–609. <https://doi.org/10.1016/B978-0-323-90769-9.00031-1>
88. Ayris D, Imtiaz M, Horbury K, Williams B, Blackney M, Hui See CS, et al. Novel deep learning approach to model and predict the spread of COVID-19. Intelligent Systems with Applications. 2022 May;14:200068. <https://doi.org/10.1016/j.iswa.2022.200068>
89. Pavlyutin M, Samoyavcheva M, Kochkarov R, Pleshakova E, Korchagin S, Gataullin T, et al. COVID-19 Spread Forecasting, Mathematical Methods vs. Machine Learning, Moscow Case. Mathematics. 2022 Jan 9;10(2):195. <https://doi.org/10.3390/math10020195>
90. Park YM, Kearney GD, Wall B, Jones K, Howard RJ, Hylock RH. COVID-19 Deaths in the United States: Shifts in Hot Spots over the Three Phases of the Pandemic and the Spatiotemporally Varying Impact of Pandemic Vulnerability. IJERPH. 2021 Aug 26;18(17):8987. <https://doi.org/10.3390/ijerph18178987>
91. Zhang S, Wang M, Yang Z, Zhang B. A Novel Predictor for Micro-Scale COVID-19 Risk Modeling: An Empirical Study from a Spatiotemporal Perspective. IJERPH. 2021 Dec 16;18(24):13294. <https://doi.org/10.3390/ijerph182413294>
92. Razavi-Termeh SV, Sadeghi-Niaraki A, Farhangi F, Choi SM. COVID-19 Risk Mapping with Considering Socio-Economic Criteria Using Machine Learning Algorithms. IJERPH. 2021 Sep 14;18(18):9657. <https://doi.org/10.3390/ijerph18189657>
93. Lounis M, Khan FM. Predicting COVID-19 cases, deaths and recoveries using machine learning methods. Eng Appl Sci Lett. 2021 Dec 31;4(4):43–9. <https://doi.org/10.30538/psrp-easl2021.0079>
94. Niraula P, Mateu J, Chaudhuri S. A Bayesian machine learning approach for spatio-temporal prediction of COVID-19 cases. Stoch Environ Res Risk Assess. 2022 Aug;36(8):2265–83. <https://doi.org/10.1007/s00477-021-02168-w>
95. Herrera M, Godoy-Faúndez A. Exploring the Roles of Local Mobility Patterns, Socioeconomic Conditions, and Lockdown Policies in Shaping the Patterns of COVID-19 Spread. Future Internet. 2021 Apr 28;13(5):112. <https://doi.org/10.3390/fi13050112>
96. Liu X, Huang J, Li C, Zhao Y, Wang D, Huang Z, et al. The role of seasonality in the spread of COVID-19 pandemic. Environmental Research. 2021 Apr;195:110874. <https://doi.org/10.1016/j.envres.2021.110874>
97. Pramanik M, Udmale P, Bisht P, Chowdhury K, Szabo S, Pal I. Climatic factors influence the spread of COVID-19 in Russia. International Journal of Environmental Health Research. 2022 Apr 3;32(4):723–37. <https://doi.org/10.1080/09603123.2020.1793921>
98. European Centre for Disease Prevention and Control. Surveillance of COVID-19 in long-term care facilities in the EU/EEA, 2020–2023. LU: Publications Office; 2024. <https://doi.org/10.2900/371227>
99. Gmehlin CG, Munoz-Price LS. Coronavirus disease 2019 (COVID-19) in long-term care facilities: A review of epidemiology, clinical presentations, and containment interventions. Infect Control Hosp Epidemiol. 2022 Apr;43(4):504–9. <https://doi.org/10.1017/ice.2020.1292>

100. Pang X, Lee BE, Gao T, Rosychuk RJ, Immaraj L, Qiu JY, et al. Early warning COVID-19 outbreak in long-term care facilities using wastewater surveillance: correlation, prediction, and interaction with clinical and serological statuses. *The Lancet Microbe*. 2024 Oct;5(10):100894. [https://doi.org/10.1016/S2666-5247\(24\)00126-5](https://doi.org/10.1016/S2666-5247(24)00126-5)
101. Thomas Craig KJ, Rizvi R, Willis VC, Kassler WJ, Jackson GP. Effectiveness of Contact Tracing for Viral Disease Mitigation and Suppression: Evidence-Based Review. *JMIR Public Health Surveill*. 2021 Oct 6;7(10):e32468. <https://doi.org/10.2196/32468>
102. Ioannidis JPA, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. *International Journal of Forecasting*. 2022 Apr;38(2):423–38. <https://doi.org/10.1016/j.ijforecast.2020.08.004>
103. Saltelli A, Bammer G, Bruno I, Charters E, Di Fiore M, Didier E, et al. Five ways to ensure that models serve society: a manifesto. *Nature*. 2020 Jun 25;582(7813):482–4. <https://doi.org/10.1038/d41586-020-01812-9>
104. Dykstra P, Bluysen PM, Bleeker-Rovers C, Derde L, van Doorslaer E, Kuipers S, et al. Met de kennis van straks: de wetenschap goed voorbereid op pandemieën. KNAW; 2022. Available from: <https://www.knaw.nl/publicaties/met-de-kennis-van-straks-de-wetenschap-goed-voorbereid-op-pandemieen>.



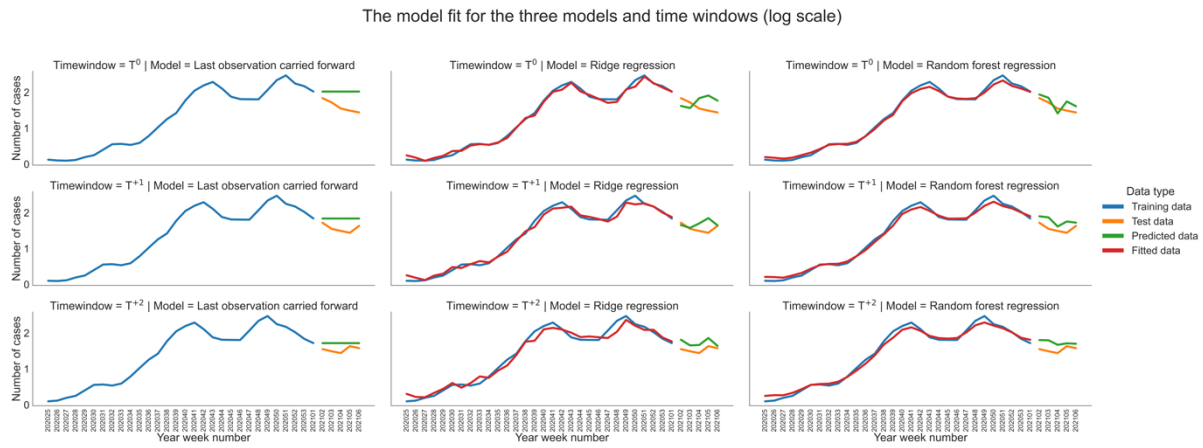
Supporting information

Supplementary Figure 1: The spatial coverage of the postal code areas included in the current study



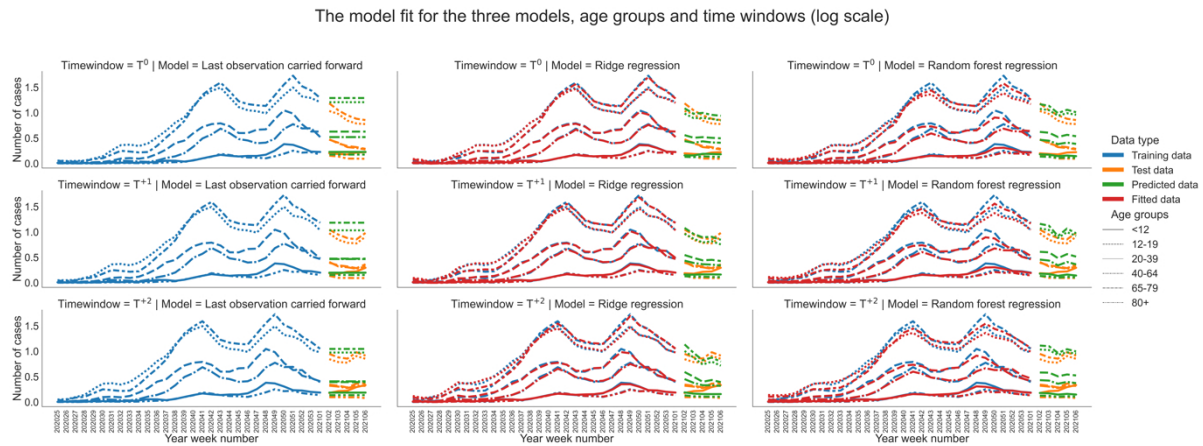
In black the postal code areas included in the current analysis, whereas in white the postal code areas are shown which were not part of the analysis. The postal code areas are provided by the © Centraal Bureau voor de Statistiek and © ESRI Nederland.

Supplementary Figure 2: The average training, test and predicted number of COVID-19 cases per model and time window for the log-transformed data



The first row shows the model fits for time window T^0 , the second row shows the model fits for time window T^{+1} (i.e., +1 week), and the third row shows the model fits for time window T^{+2} (i.e., +2 weeks). The separate columns correspond to the three different models used (resp. last observation carried forward, ridge regression and random forest regression). The blue line represents the training data, the orange line shows the test data, the red line visualizes the model fit and finally the green line shows the model predictions. Since no model was fitted for the last observation model, the fitted data line is absent.

Supplementary Figure 3: The average training, test and predicted number of COVID-19 cases per model, time window and age group for the log-transformed data



The first row shows the model fits for time window T^0 , the second row shows the model fits for time window T^{+1} (+ 1 week), and the third row shows the model fits for time window T^{+2} (+2 weeks). The separate columns correspond to the three different models used (resp. last observation carried forward, ridge regression and random forest regression). The blue line represents the training data, the orange line shows the test data, the red line visualizes the model fit and finally the green line shows the model predictions. Since no model was fitted for the last observation model, the fitted data line is absent.

Supplementary Table 1: The last observation carried forward, ridge regression and random forest regression model performance for all age groups collapsed (log-transformed data)

Model	Time Window	MSE mean (std.)	Top 100 correctly predicted
LOCF	T^0	0.72 (0.68)	41%
	T^{+1}	0.62 (0.61)	33%
	T^{+2}	0.58 (0.56)	33%
RR	T^0	0.6 (0.21)	42%
	T^{+1}	0.79 (0.49)	37%
	T^{+2}	1.24 (1.13)	37%
RFR	T^0	0.75 (0.47)	47%
	T^{+1}	0.76 (0.41)	41%
	T^{+2}	0.75 (0.50)	46%

The average Mean Squared Error (MSE), percentage correctly predicted top 100 over all postal code areas and rank mismatch per model type and time window. The MSE indicates the average difference in observed and predicted values. The top 100 correctly predicted postal code areas indicate what percentage of predicted top 100 postal code areas with the highest number of COVID-19 cases were correctly predicted irrespective of absolute rank.

LOCF: Last Observation Carried Forward; RR: Ridge Regression; RFR: Random Forest; std: standard deviation.



Supplementary Table 2: The last observation carried forward, ridge regression and random forest regression model performance per age group (log-transformed data)

Model	Time Window	MSE mean (std.)	Age group (years)	Top 100 Correctly predicted
LOCF	T^0	0.45 (0.39)	<12	15%
			12-19	21%
			20-39	38%
			40-64	34%
			65-79	27%
			≥ 80	19%
	T^{+1}	0.40 (0.37)	<12	13%
			12-19	20%
			20-39	46%
			40-64	32%
			65-79	22%
			≥ 80	18%
	T^{+2}	0.38 (0.35)	<12	24%
			12-19	16%
			20-39	35%
			40-64	19%
			65-79	21%
			≥ 80	13%
RR	T^0	0.28 (0.07)	<12	17%
			12-19	25%
			20-39	43%
			40-64	35%
			65-79	25%
			≥ 80	19%
	T^{+1}	0.31 (0.08)	<12	20%
			12-19	17%
			20-39	45%
			40-64	30%
			65-79	24%
			≥ 80	18%
	T^{+2}	0.40 (0.16)	<12	24%
			12-19	26%
			20-39	32%
			40-64	30%
			65-79	25%
			≥ 80	14%
RFR	T^0	0.32 (0.13)	<12	18%
			12-19	31%
			20-39	57%
			40-64	40%
			65-79	28%
			≥ 80	25%
	T^{+1}	0.32 (0.08)	<12	18%
			12-19	22%
			20-39	54%
			40-64	34%
			65-79	34%
			≥ 80	22%
	T^{+2}	0.32 (0.08)	<12	28%
			12-19	32%
			20-39	46%
			40-64	36%
			65-79	32%
			≥ 80	17%

The average Mean Squared Error (MSE), percentage correctly predicted top 100 over all postal code areas and rank mismatch per model type and time window. The MSE indicates the average difference in observed and predicted values. The top 100 correctly predicted postal code areas indicate what percentage of predicted top 100 postal code areas with the highest number of COVID-19 cases were correctly predicted irrespective of absolute rank.

LOCF.: Last Observation Carried Forward; RR: Ridge Regression; RFR: Random Forest; std: standard deviation.

Supplementary Table 3: The last observation carried forward, ridge regression and random forest regression model comparison predicting the top 100 postal codes per age group (all models)

Models	P(M)	P(M data)	BFM	BF10	error %
MT+TW+AG+MT*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.554	205.986	1.000	**
MT+S+TW+AG+MT*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.152	29.806	0.275	11.160
MT+S+TW+AG+MT*TW+MT*AG+S*AG+TW*AG+MT*TW*AG	0.006	0.113	21.200	0.205	3.150
MT+S+TW+AG+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*TW*AG	0.006	0.037	6.453	0.068	20.527
MT+S+TW+AG+MT*TW+S*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.035	6.013	0.063	3.965
MT+S+TW+AG+MT*S+MT*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.023	3.832	0.041	4.393
MT+S+TW+AG+MT*S+MT*TW+MT*AG+S*AG+TW*AG+MT*TW*AG	0.006	0.018	3.114	0.033	2.981
MT+S+TW+AG+MT*S+MT*TW+MT*AG+S*AG+TW*AG+MT*S*AG+MT*TW*AG	0.006	0.018	2.967	0.032	2.884
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*TW*AG+S*TW*AG	0.006	0.012	1.981	0.021	3.353
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+TW*AG+MT*TW*AG	0.006	0.009	1.511	0.016	40.181
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*AG+MT*TW*AG	0.006	0.006	1.002	0.011	4.252
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*TW*AG	0.006	0.005	0.900	0.010	4.054
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+TW*AG+MT*S*TW+MT*TW*AG	0.006	0.005	0.899	0.010	69.000
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*AG+MT*TW*AG+S*TW*AG	0.006	0.003	0.538	0.006	3.446
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW+MT*S*AG+MT*TW*AG	0.006	0.002	0.368	0.004	3.880
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*TW*AG+S*TW*AG	0.006	0.002	0.341	0.004	3.255
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW+MT*TW*AG	0.006	0.002	0.291	0.003	3.398
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW+MT*S*AG+MT*TW*AG+S*TW*AG	0.006	0.002	0.258	0.003	5.894
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW+MT*TW*AG+S*TW*AG	0.006	7.16E-01	0.119	0.001	3.061
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW+MT*S*AG+MT*TW*AG	0.006	7.12E-01	0.118	0.001	4.859
+S*TW*AG+MT*S*TW*AG					
MT+TW+AG+MT*TW+MT*AG+TW*AG	0.006	2.34E-02	0.004	4.23E-02	2.274
MT+TW+AG+MT*AG+TW*AG	0.006	8.82E-03	0.001	1.59E-02	2.069
MT+S+TW+AG+MT*TW+MT*AG+TW*AG	0.006	5.09E-03	8.45E-01	9.19E-03	2.946
MT+S+TW+AG+MT*AG+TW*AG	0.006	1.93E-03	3.20E-01	3.48E-03	2.651
MT+S+TW+AG+MT*TW+S*TW+MT*AG+TW*AG	0.006	9.18E-04	1.53E-01	1.66E-03	3.725
MT+S+TW+AG+MT*TW+MT*AG+S*AG+TW*AG	0.006	9.07E-04	1.51E-01	1.64E-03	2.792
MT+S+TW+AG+MT*S+MT*TW+MT*AG+TW*AG	0.006	7.54E-04	1.25E-01	1.36E-03	2.622
MT+S+TW+AG+S*TW+MT*AG+TW*AG	0.006	3.40E-04	5.64E-02	6.14E-04	3.124
MT+S+TW+AG+MT*AG+S*AG+TW*AG	0.006	3.32E-04	5.51E-02	6.00E-04	5.095
MT+S+TW+AG+MT*S+MT*AG+TW*AG	0.006	2.88E-04	4.78E-02	5.20E-04	3.785
MT+S+TW+AG+MT*TW+S*TW+MT*AG+S*AG+TW*AG	0.006	1.79E-04	2.97E-02	3.23E-04	3.964
MT+S+TW+AG+MT*S+MT*TW+MT*AG+S*AG+TW*AG	0.006	1.54E-04	2.55E-02	2.77E-04	8.976
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+TW*AG	0.006	1.43E-04	2.37E-02	2.57E-04	2.964
MT+S+TW+AG+S*TW+MT*AG+S*AG+TW*AG	0.006	5.94E-05	9.86E-03	1.07E-04	5.401
MT+S+TW+AG+MT*S+MT*AG+S*AG+TW*AG	0.006	5.39E-05	8.95E-03	9.74E-05	11.457
MT+S+TW+AG+MT*S+S*TW+MT*AG+TW*AG	0.006	5.37E-05	8.92E-03	9.70E-05	3.170
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG	0.006	2.76E-05	4.59E-03	4.99E-05	4.589
MT+S+TW+AG+MT*S+MT*TW+MT*AG+S*AG+TW*AG+MT*S*AG	0.006	2.76E-05	4.58E-03	4.99E-05	7.507
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+TW*AG+MT*S*TW	0.006	2.63E-05	4.37E-03	4.75E-05	2.622
MT+S+TW+AG+MT*TW+S*TW+MT*AG+S*AG+TW*AG+S*TW*AG	0.006	2.61E-05	4.33E-03	4.71E-05	6.599
MT+S+TW+AG+MT*S+S*TW+MT*AG+S*AG+TW*AG	0.006	8.58E-06	1.43E-03	1.55E-05	2.822
MT+S+TW+AG+MT*S+MT*AG+S*AG+TW*AG+MT*S*AG	0.006	8.39E-06	1.39E-03	1.52E-05	6.177
MT+S+TW+AG+S*TW+MT*AG+S*AG+TW*AG+S*TW*AG	0.006	7.58E-06	1.26E-03	1.37E-05	3.228
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW	0.006	5.25E-06	8.72E-04	9.48E-06	3.658
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*AG	0.006	5.10E-06	8.46E-04	9.21E-06	6.232
MT+TW+AG+TW*AG	0.006	3.85E-06	6.39E-04	6.95E-06	2.186
MT+S+TW+AG+MT*S+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+S*TW*AG	0.006	3.81E-06	6.33E-04	6.88E-06	3.143
MT+TW+AG+MT*TW+TW*AG	0.006	2.67E-06	4.43E-04	4.82E-06	3.385
MT+S+TW+AG+MT*S+S*TW+MT*AG+S*AG+TW*AG+MT*S*AG	0.006	1.59E-06	2.63E-04	2.87E-06	7.302
MT+S+TW+AG+MT*S+S*TW+MT*AG+S*AG+TW*AG+S*TW*AG	0.006	1.13E-06	1.88E-04	2.05E-06	3.025
MT+S+TW+AG+MT*TW+TW*AG	0.006	1.02E-06	1.69E-04	1.83E-06	48.071
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW+MT*S*AG	0.006	9.48E-07	1.57E-04	1.71E-06	4.092
MT+S+TW+AG+TW*AG	0.006	8.53E-07	1.42E-04	1.54E-06	6.447
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*AG+S*TW*AG	0.006	7.46E-07	1.24E-04	1.35E-06	4.088
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW+S*TW*AG	0.006	7.20E-07	1.19E-04	1.30E-06	4.871
MT+S+TW+AG+MT*S+S*TW+MT*AG+S*AG+TW*AG+MT*S*AG+S*TW*AG	0.006	1.91E-07	3.17E-05	3.45E-07	3.160
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+TW*AG+MT*S*TW+MT*S*AG+S*TW*AG	0.006	1.36E-07	2.26E-05	2.46E-07	3.836
MT+S+TW+AG+S*TW+TW*AG	0.006	1.32E-07	2.19E-05	2.38E-07	2.838
MT+S+TW+AG+MT*S+TW*AG	0.006	1.22E-07	2.03E-05	2.21E-07	3.514
MT+S+TW+AG+MT*TW+S*TW+TW*AG	0.006	9.60E-08	1.59E-05	1.73E-07	5.848
MT+S+TW+AG+S*AG+TW*AG	0.006	9.55E-08	1.59E-05	1.72E-07	2.878
MT+S+TW+AG+MT*S+MT*TW+TW*AG	0.006	8.42E-08	1.40E-05	1.52E-07	4.585
MT+S+TW+AG+MT*TW+S*AG+TW*AG	0.006	6.75E-08	1.12E-05	1.22E-07	2.548
MT+S+TW+AG+MT*S+S*TW+TW*AG	0.006	1.98E-08	3.29E-06	3.58E-08	3.117
MT+S+TW+AG+S*TW+S*AG+TW*AG	0.006	1.61E-08	2.68E-06	2.91E-08	3.504
MT+S+TW+AG+MT*S+S*AG+TW*AG	0.006	1.57E-08	2.60E-06	2.83E-08	7.322
MT+S+TW+AG+MT*S+MT*TW+S*TW+TW*AG	0.006	1.37E-08	2.27E-06	2.47E-08	3.304
MT+S+TW+AG+MT*TW+S*TW+S*AG+TW*AG	0.006	1.19E-08	1.97E-06	2.15E-08	3.070
MT+S+TW+AG+MT*S+MT*TW+S*AG+TW*AG	0.006	1.06E-08	1.75E-06	1.91E-08	3.638
MT+S+TW+AG+MT*S+S*TW+S*AG+TW*AG	0.006	2.45E-09	4.07E-07	4.42E-09	3.376
MT+S+TW+AG+MT*S+MT*TW+S*TW+TW*AG+MT*S*TW	0.006	2.36E-09	3.92E-07	4.27E-09	3.288
MT+S+TW+AG+MT*S+MT*TW+S*TW+S*AG+TW*AG	0.006	1.90E-09	3.16E-07	3.44E-09	6.149
MT+S+TW+AG+S*TW+S*AG+TW*AG+S*TW*AG	0.006	1.77E-09	2.93E-07	3.19E-09	2.801
MT+S+TW+AG+MT*TW+S*TW+S*AG+TW*AG+S*TW*AG	0.006	1.39E-09	2.31E-07	2.51E-09	3.456
MT+S+TW+AG+MT*S+MT*TW+S*TW+S*AG+TW*AG+MT*S*TW	0.006	2.90E-10	4.81E-08	5.24E-10	3.147
MT+S+TW+AG+MT*S+S*TW+S*AG+TW*AG+S*TW*AG	0.006	2.64E-10	4.38E-08	4.76E-10	2.684

Continued

Supplementary Table 3: Continued

Models	P(M)	P(M data)	BFM	BF10	error %
MT+S+TW+AG+MT*S+MT*TW+S*TW+S*AG+TW*AG+S*TW*AG	0.006	2.21E-10	3.67E-08	4.00E-10	8.032
MT+S+TW+AG+MT*S+MT*TW+S*TW+S*AG+TW*AG+MT*S*TW+S*TW*AG	0.006	3.18E-11	5.28E-09	5.74E-11	2.664
MT+TW+AG+MT*AG	0.006	1.20E-13	1.99E-11	2.17E-13	2.847
MT+AG+MT*AG	0.006	1.08E-13	1.79E-11	1.95E-13	2.038
MT+AG	0.006	6.94E-14	1.15E-11	1.25E-13	1.965
MT+TW+AG	0.006	5.99E-14	9.94E-12	1.08E-13	2.849
MT+TW+AG+MT*TW+MT*AG	0.006	3.06E-14	5.07E-12	5.52E-14	2.141
MT+S+TW+AG+MT*AG	0.006	2.46E-14	4.09E-12	4.45E-14	3.437
MT+S+AG+MT*AG	0.006	2.18E-14	3.62E-12	3.94E-14	2.645
MT+S+AG	0.006	1.46E-14	2.42E-12	2.63E-14	2.867
MT+TW+AG+MT*TW	0.006	1.32E-14	2.19E-12	2.38E-14	2.421
MT+S+TW+AG	0.006	1.24E-14	2.07E-12	2.25E-14	3.141
MT+S+TW+AG+MT*TW+MT*AG	0.006	6.38E-15	1.06E-12	1.15E-14	3.649
MT+S+TW+AG+MT*S+MT*AG	0.006	4.23E-15	7.02E-13	7.63E-15	15.750
MT+S+TW+AG+S*TW+MT*AG	0.006	3.78E-15	6.28E-13	6.83E-15	2.581
MT+S+AG+MT*S+MT*AG	0.006	3.26E-15	5.41E-13	5.88E-15	2.560
MT+S+TW+AG+MT*TW	0.006	2.74E-15	4.55E-13	4.95E-15	3.575
MT+S+TW+AG+MT*AG+S*AG	0.006	2.26E-15	3.75E-13	4.08E-15	2.562
MT+S+AG+MT*AG+S*AG	0.006	2.11E-15	3.50E-13	3.81E-15	3.686
MT+S+AG+MT*S	0.006	2.11E-15	3.50E-13	3.80E-15	2.535
MT+S+TW+AG+S*TW	0.006	2.02E-15	3.36E-13	3.66E-15	7.089
MT+S+TW+AG+MT*S	0.006	1.79E-15	2.98E-13	3.24E-15	2.737
MT+S+AG+S*AG	0.006	1.28E-15	2.12E-13	2.31E-15	2.481
MT+S+TW+AG+S*AG	0.006	1.11E-15	1.85E-13	2.01E-15	2.594
MT+S+TW+AG+MT*TW+S*TW+MT*AG	0.006	9.84E-16	1.63E-13	1.78E-15	2.725
MT+S+TW+AG+MT*S+MT*TW+MT*AG	0.006	9.32E-16	1.55E-13	1.68E-15	4.532
MT+S+TW+AG+MT*TW+MT*AG+S*AG	0.006	6.00E-16	9.97E-14	1.08E-15	2.471
MT+S+TW+AG+MT*S+S*TW+MT*AG	0.006	5.38E-16	8.93E-14	9.72E-16	2.629
MT+S+TW+AG+MT*TW+S*TW	0.006	4.44E-16	7.37E-14	8.02E-16	3.049
MT+S+TW+AG+MT*S+MT*TW	0.006	4.12E-16	6.84E-14	7.44E-16	5.341
MT+S+TW+AG+S*TW+MT*AG+S*AG	0.006	3.66E-16	6.07E-14	6.61E-16	3.087
MT+S+TW+AG+MT*S+MT*AG+S*AG	0.006	3.56E-16	5.91E-14	6.43E-16	3.898
MT+S+AG+MT*S+MT*AG+S*AG	0.006	3.00E-16	4.98E-14	5.42E-16	2.696
MT+S+TW+AG+MT*S+S*TW	0.006	2.85E-16	4.73E-14	5.15E-16	3.735
MT+S+TW+AG+MT*TW+S*AG	0.006	2.51E-16	4.17E-14	4.54E-16	2.713
MT+S+AG+MT*S+S*AG	0.006	1.90E-16	3.16E-14	3.44E-16	2.968
MT+S+TW+AG+MT*S+S*AG	0.006	1.74E-16	2.89E-14	3.14E-16	6.020
MT+S+TW+AG+S*TW+S*AG	0.006	1.72E-16	2.86E-14	3.11E-16	2.896
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG	0.006	1.69E-16	2.80E-14	3.05E-16	5.685
MT+S+TW+AG+MT*TW+S*TW+MT*AG+S*AG	0.006	9.77E-17	1.62E-14	1.76E-16	4.028
MT+S+TW+AG+MT*S+MT*TW+MT*AG+S*AG	0.006	9.20E-17	1.53E-14	1.66E-16	6.108
MT+S+TW+AG+MT*S+MT*TW+S*TW	0.006	6.80E-17	1.13E-14	1.23E-16	5.595
MT+S+TW+AG+MT*S+S*TW+MT*AG+S*AG	0.006	5.41E-17	8.98E-15	9.76E-17	3.924
MT+S+TW+AG+MT*TW+S*TW+S*AG	0.006	3.86E-17	6.41E-15	6.98E-17	2.598
MT+S+TW+AG+MT*S+MT*AG+S*AG+MT*S*AG	0.006	3.77E-17	6.27E-15	6.82E-17	3.418
MT+S+TW+AG+MT*S+MT*TW+S*AG	0.006	3.70E-17	6.15E-15	6.69E-17	3.458
MT+S+AG+MT*S+MT*AG+S*AG+MT*S*AG	0.006	3.24E-17	5.38E-15	5.85E-17	2.618
MT+S+TW+AG+MT*S+S*TW+S*AG	0.006	2.66E-17	4.42E-15	4.80E-17	3.490
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+MT*S*TW	0.006	2.21E-17	3.67E-15	3.99E-17	2.838
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG	0.006	1.45E-17	2.41E-15	2.62E-17	4.014
MT+S+TW+AG+MT*S+MT*TW+MT*AG+S*AG+MT*S*AG	0.006	1.07E-17	1.78E-15	1.94E-17	7.818
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*S*TW	0.006	9.80E-18	1.63E-15	1.77E-17	3.301
MT+S+TW+AG+MT*S+S*TW+MT*AG+S*AG+MT*S*AG	0.006	5.64E-18	9.37E-16	1.02E-17	2.919
MT+S+TW+AG+MT*S+MT*TW+S*TW+S*AG	0.006	5.51E-18	9.14E-16	9.94E-18	2.968
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+MT*S*TW	0.006	2.25E-18	3.74E-16	4.07E-18	3.576
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+MT*S*AG	0.006	1.52E-18	2.52E-16	2.74E-18	3.209
MT+S+TW+AG+MT*S+MT*TW+S*TW+S*AG+MT*S*TW	0.006	9.17E-19	1.52E-16	1.66E-18	3.564
TW+AG+TW*AG	0.006	4.20E-19	6.97E-17	7.58E-19	1.970
MT+S+TW+AG+MT*S+MT*TW+S*TW+MT*AG+S*AG+MT*S*TW+MT*S*AG	0.006	2.54E-19	4.21E-17	4.58E-19	4.118
S+TW+AG+TW*AG	0.006	8.49E-20	1.41E-17	1.53E-19	2.221
S+TW+AG+S*TW+TW*AG	0.006	1.45E-20	2.40E-18	2.61E-20	5.993
S+TW+AG+S*AG+TW*AG	0.006	7.48E-21	1.24E-18	1.35E-20	2.401
AG	0.006	2.18E-21	3.62E-19	3.94E-21	1.809
S+TW+AG+S*TW+S*AG+TW*AG	0.006	1.12E-21	1.86E-19	2.02E-21	2.326
TW+AG	0.006	8.27E-22	1.37E-19	1.49E-21	2.338
S+AG	0.006	4.48E-22	7.43E-20	8.08E-22	2.502
S+TW+AG	0.006	1.65E-22	2.74E-20	2.99E-22	2.219
S+TW+AG+S*TW+S*AG+TW*AG+S*TW*AG	0.006	1.07E-22	1.78E-20	1.93E-22	2.691
S+AG+S*AG	0.006	3.63E-23	6.02E-21	6.55E-23	3.535
S+TW+AG+S*TW	0.006	2.49E-23	4.14E-21	4.50E-23	2.614
S+TW+AG+S*AG	0.006	1.33E-23	2.21E-21	2.40E-23	2.330
S+TW+AG+S*TW+S*AG	0.006	2.13E-24	3.54E-22	3.85E-24	5.487
MT	0.006	3.18E-44	5.28E-42	5.75E-44	1.809
MT+S	0.006	7.44E-45	1.23E-42	1.34E-44	13.594
MT+TW	0.006	4.37E-45	7.25E-43	7.89E-45	2.449
Nullmodel	0.006	3.68E-45	6.10E-43	6.64E-45	1.809
MT+S+MT*S	0.006	9.42E-46	1.56E-43	1.70E-45	2.757
MT+S+TW	0.006	9.07E-46	1.51E-43	1.64E-45	2.639
S	0.006	7.52E-46	1.25E-43	1.36E-45	1.809
TW	0.006	4.84E-46	8.03E-44	8.73E-46	1.809
MT+TW+MT*TW	0.006	3.82E-46	6.34E-44	6.90E-46	2.364
MT+S+TW+S*TW	0.006	1.27E-46	2.10E-44	2.28E-46	2.651
MT+S+TW+MT*S	0.006	1.26E-46	2.10E-44	2.28E-46	3.427
S+TW	0.006	9.71E-47	1.61E-44	1.75E-46	2.622
MT+S+TW+MT*TW	0.006	7.47E-47	1.24E-44	1.35E-46	2.657
MT+S+TW+MT*S+S*TW	0.006	1.82E-47	3.03E-45	3.29E-47	3.282

Continued

Supplementary Table 3: Continued

Models	P(M)	P(M data)	BFM	BF10	error %
S+TW+S*TW	0.006	1.39E-47	2.31E-45	2.52E-47	2.496
MT+S+TW+MT*S+MT*TW	0.006	1.08E-47	1.80E-45	1.96E-47	2.610
MT+S+TW+MT*TW+S*TW	0.006	1.07E-47	1.78E-45	1.94E-47	2.584
MT+S+TW+MT*S+MT*TW+S*TW	0.006	1.54E-48	2.56E-46	2.79E-48	2.711
MT+S+TW+MT*S+MT*TW+S*TW+MT*S*TW	0.006	2.28E-49	3.78E-47	4.11E-49	3.915

MT: Model Type; TW: Time Window; S: natural or log Scale; AG: Age group; P(M): Prior model probability; P(M|data): posterior model probability; BFM: the change from prior odds to posterior odds; BF10: the Bayes factor relative to the best model; error %: indicates the precision of the numerical approximation. ** As the estimated Bayes factors are relative to the best performing model, there is no relative error associated with the best performing model.



Supplementary Table 4: The feature importance for the T⁰ RFR model predicting the top 100 postal codes with the highest number of COVID-19 cases

	Feature	Feature importance
All age groups together	GT COVID-19	0.00014
	GT Reuk	0.00009
	Temporal lag 1	0.00008
	Temporal lag 2	0.00007
	Local Moran I	0.00005
	Global Moran I	0.00004
	Stringency index	0.00002
	SIR	0.00002
	GT COVID-19 Testen	0.00002
	GT COVID-19 vaccine	0.00001
	GT Loss of smell	0.00001
	GT Coronavirus disease 2019	0.00001
	GT Symptoms of COVID-19	0.00001
	GT COVID-19 Testing	0.00001
	GT Symptomen van COVID-	0.00000
	Total number of inhabitants in a postal code	0.00000
	Average size of household in a postal code	0.00000
	Population density of a postal code	0.00000
	Total number of households in a postal code	0.00000
Per age group	GT Reuk	0.00008
	GT COVID-19	0.00006
	Temporal lag 1 (40-64)	0.00003
	Local Moran I (20-39)	0.00003
	Temporal lag 2 (40-64)	0.00002
	Global Moran I (20-39)	0.00002
	Local Moran I (>=80)	0.00002
	Local Moran I (65-79)	0.00002
	Local Moran I (40-64)	0.00002
	Global Moran I (40-64)	0.00002
	Temporal lag 1 (20-39)	0.00002
	Local Moran I (12-19)	0.00002
	Local Moran I (<12)	0.00002
	Temporal lag 2 (20-39)	0.00002
	Global Moran I (12-19)	0.00002
	SIR	0.00002
	Global Moran I (65-79)	0.00001
	Stringency index	0.00001
	Temporal lag 1 (12-19)	0.00001
	Global Moran I (<12)	0.00001
	Temporal lag 2 (12-19)	0.00001
	Global Moran I (>=80)	0.00001
	GT COVID-19 Testen	0.00001
	GT COVID-19 Vaccine	0.00001
	Temporal lag 1 (65-79)	0.00001
	GT Loss of smell	0.00000
	GT Coronavirus disease 2019	0.00000
	Temporal lag 2 (65-79)	0.00000
	GT COVID-19 Testing	0.00000
	GT Symptoms of COVID-19	0.00000
	GT Symptomen van COVID-19	0.00000
	Temporal lag 1 (<12)	0.00000
	Temporal lag 2 (<12)	0.00000
	Temporal lag 1 (>=80)	0.00000
	Temporal lag 2 (>=80)	0.00000
	Total number of inhabitants in a postal code	0.00000
	Average size of household in a postal code	0.00000
	Population density of a postal code	0.00000
	Total number of households in a postal code	0.00000

GT: Google Trends; SIR: Standardized Incidence Ratio.