medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

# Who should we test for COVID-19?

# A triage model built from national symptom surveys

Saar Shoer[*1,2], Tal Karady[*1,2], Ayya Keshet[*1,2], Smadar Shilo[*1,2,3], Hagai Rossman[*1,2], Amir Gavrieli[1,2], Tomer Meir[1,2], Amit Lavon[1,2], Dmitry Kolobkov[1,2], Iris Kalka[1,2], Anastasia Godneva[1,2], Ori Cohen, Adam Kariv[4], Ori Hoch[4], Mushon Zer-Aviv[4], Noam Castel[4], Anat Ekka Zohar[5], Angela Irony[5], Benjamin Geiger[2], Dorit Hizi[4], Varda Shalev[5,8], Ran Balicer[7], Eran Segal[1,2,†]

[*] Equal contribution
[†] Corresponding author
[1] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel.
[2] Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel.
[3] Pediatric Diabetes Unit, Ruth Rappaport Children's Hospital, Rambam Healthcare Campus, Haifa, Israel.
[4] The public knowledge workshop, Tel Aviv, Israel.
[5] Epidemiology and Database Research Unit, Maccabi Healthcare Services, Tel Aviv, Israel.
[7] Clalit Research Institute, Clalit Health Services, Tel Aviv, Israel.
[8] School of Public Health, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.



NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Abstract

The gold standard for COVID-19 diagnosis is detection of viral RNA in a reverse transcription PCR test. Due to global limitations in testing capacity, effective prioritization of individuals for testing is essential. Here, we devised a model that estimates the probability of an individual to test positive for COVID-19 based on answers to 9 simple questions regarding age, gender, presence of prior medical conditions, general feeling, and the symptoms fever, cough, shortness of breath, sore throat and loss of taste or smell, all of which have been associated with COVID-19 infection. Our model was devised from a subsample of a national symptom survey that was answered over 2 million times in Israel over the past 2 months and a targeted survey distributed to all residents of several cities in Israel. Overall, 43,752 adults were included, from which 498 self-reported as being COVID-19 positive. The model provides statistically significant predictions on held-out individuals and achieves a positive predictive value (PPV) of 46.3% at a 10% sensitivity. As our tool can be used online and without the need of exposure to suspected patients, it may have worldwide utility in combating COVID-19 by better directing the limited testing resources through prioritization of individuals for testing, thereby increasing the rate at which positive individuals can be identified and isolated.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Introduction

The rapid and global spread of COVID-19 led the World Health Organization (WHO) to declare it a pandemic on March 11, 2020. One major factor that contributes to the spread of the virus is the apparently large number of undiagnosed infected individuals. This knowledge gap facilitates the silent propagation of the virus, delays the response of public health officials and results in an explosion in the number of cases [1,2].

One reason for this knowledge gap is insufficient testing. While the current gold standard for COVID-19 diagnosis is detection of viral RNA in a reverse transcription PCR test, the number of tests is limited by financial and logistic constraints. In a time when almost all countries are faced with the same health challenge, resources are scarce. This creates the need for a prioritization mechanism to allocate tests and resources more efficiently towards individuals who are more likely to test positive, leading to earlier identification of COVID-19 patients and reduced spread of the virus. Despite this need, most countries still employ a simplistic testing strategy based on display of symptoms associated with the disease and either close epidemiological contact with a confirmed COVID-19 case or belonging to a high risk group[3]. In practice, these strategies lead to a relatively small fraction of positive tests among those tested and thus to inefficient use of the precious testing resources.

Here, we present a model that provides statistically significant estimates of the probability of an individual to test positive for SARS-CoV-2 infection in a PCR test, based on a national symptom survey that we distributed in Israel. Notably, while most studies describing the clinical characteristics of COVID-19 cases were based on symptoms of hospitalized patients [4], our survey data allowed us to also study symptoms of milder cases and reveal which symptoms hold the highest predictive power for COVID-19 diagnosis. Using our model, the risk for a positive COVID-19 test can be evaluated in less than a minute and without added costs, or risk of exposure to a suspected patient. Our approach can thus be used globally to make more efficient use of available testing capacities, by significantly increasing the fraction of positive tests obtained, and by rapidly identifying individuals that should be isolated until definitive test results are obtained.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

# Methods

## Data

In this study, we utilized data that originates from two versions of a one-minute survey that was developed and deployed by our research group in the early stages of the COVID-19 spread in Israel [5]. The online version of the survey includes questions relating to age, gender, prior medical conditions, smoking habits, self-reported symptoms and geographical location (see section 1a in the supplementary appendix). Questions regarding prior medical conditions and symptoms included in the survey were carefully chosen by medical professionals. Each participant is asked to fill the survey once a day for himself and for family members that are unable to fill it for themselves (e.g. children and the elderly). The survey is anonymous to maintain the privacy of the participants, and has been collected since March 14th, 2020. As the number of COVID-19 diagnosed individuals in Israel rose, in some cities more than others, a shortened version of the survey was deployed using an Interactive Voice Response (IVR) platform. This version of the survey included information on respondents' age group, gender, presence of prior medical conditions, general feeling and a partial list of symptoms, including fever, cough, shortness of breath, sore throat and loss of taste or smell (see Section 1b in the supplementary appendix). Cities were targeted to participate in the IVR version of the survey according to the number of diagnosed patients and an increased concern for COVID-19 outbreaks (Table S1). Starting April 5, 2020, citizens in the targeted cities were contacted and responses were collected anonymously. This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline [6].

## Study Design and Population

Overall, 695,586 and 66,447 responses were collected up until April 26th, 2020, from the online and IVR versions of the survey, respectively. Since children express different clinical manifestations of COVID-19 infection [7,8], we decided to focus our analysis only on adults (age above 20 years old). To avoid translation discrepancies, we included only responses in Hebrew. We also excluded responses that did not meet quality control criteria, such as a reasonable age (<120 years old) and body temperature (35-43°C) or responses suspected as spam (see Section 3 in the supplementary appendix) (Figure 1). A total of 33,737 responses were

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

eventually included from the IVR version of the survey. Since surveyed cities were at relatively high risk (Table S1), the prevalence of COVID-19 diagnosed responders in the IVR data was 1.14%, 6 times higher than the national prevalence of 0.18% [3]. These cities also had very high response rates, between 6% to 16% of the cities' population (Table S1). From the online version of the survey, we randomly sampled a single response for each individual that was recognized by a unique identifier we created (see Section 4 in the supplementary appendix), and when an individual reported a COVID-19 diagnosis, we randomly sampled one response only from those which included a positive diagnosis answer. A total of 131,166 responses were identified in the online version, of which 0.09% reported a positive COVID-19 diagnosis, which is closer to the national prevalence of 0.18% [3]. The characteristics of these unique responders are described in Table S2 of the supplementary appendix.

For the integration of the two survey versions, all 33,737 IVR responses were combined together with all 114 uniquely identified responders in the online survey that self-reported COVID-19 diagnosis and a random sample of 9,901 undiagnosed responders in the online version, to maintain the same diagnosis prevalence as in the IVR version (Figure 1). Overall, 43,752 responses were eventually included in the study, of which 498 self-reported as being COVID-19 diagnosed. The characteristics of these responders are described in Table 1.

**Predicting the Outcome of a COVID-19 Test**

We defined survey self-reporting of a COVID-19 laboratory confirmed diagnosis as our outcome. We constructed two models. The first, which we term the primary model, was constructed from the integrated responses from both the IVR and online versions of the survey and included the reduced set of questions that were surveyed in the IVR version. The second model, which we term the extended features model, was constructed using only responses from the online version, and included additional symptoms and questions that were not part of the IVR survey. We trained both the primary and extended features models using Logistic Regression. In addition, in order to capture nonlinear interactions and interactions amongst features, in both cases we also constructed models using a Gradient Boosting Decision Trees algorithm [9]. To test the validity of our models, all model constructions were done using the framework of cross-validation, in which model performance is evaluated on a subset of the data that was not used in the model's construction. For more

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

information on handling of missing values and the models construction process, see Section 6 in the supplementary appendix.

**Primary Model**

The primary model was constructed using responses to both the IVR and the online version of the survey. Features included in this model were determined by the IVR version, since it included a subset of the online version questions. These consisted of age group, gender, presence of prior medical conditions, general feeling, and the following symptoms: fever, cough, shortness of breath, sore throat and loss of taste or smell.

**Extended Features Model**

The extended features model was constructed using only responses from the online version of the survey, as it had 14 additional features that were not available in the IVR version. This extended list added dry cough and moist cough (instead of general cough in the primary model), fatigue, muscle pain, rhinorrhea, diarrhea, nausea or vomiting, chills, confusion and reporting on presence of specific prior medical conditions separately (as opposed to the presence of *any* of the prior medical conditions in the primary model).

**Baseline Models**

To assess the contribution of reported symptoms and prior medical conditions to both the primary model and the extended features model (in both the Logistic Regression and the Gradient Boosting Decision Trees versions), we constructed baseline models using only age group and gender information to predict our outcome.

**Analysis of feature contributions**

To gain insight into the features that contribute most to the predicted probability of being diagnosed with COVID-19 of our models, we analyzed feature contribution in the Gradient Boosting Decision Trees models using SHAP (SHapley Additive exPlanation) [10]. SHAP aims to interpret the output of a machine learning model by estimating the Shapley value of each feature, which represents the average change in the output of the model, by conditioning on that feature while introducing other features one at a time, over all possible features ordering.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Analyzing feature contributions in each of the models allowed us to compare the inner workings of each model and to identify which features dominated in each prediction.

We further analyzed SHAP interaction values, which uses the 'Shapely interaction index' to capture local interaction effects between features [10]. Interaction values are calculated for each pair of the model's features, and for each individual prediction of the model, allowing us to uncover interaction patterns between pairs of features. We placed particular emphasis on the contribution of participants' age and the interaction of age with all other features.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Results

Our primary model for prediction of a positive COVID-19 test result, which was constructed using Logistic Regression, achieved an area under the Receiver-Operating-Characteristic (auROC) of 0.737 (CI: 0.712-0.759), and an area under the Precision-Recall (auPR) of 0.144 (CI: 0.119-0.177). This model substantially outperforms the baseline model, which uses only age group and gender and achieved an auROC of 0.556 (CI: 0.527-0.581), and an auPR of 0.013 (CI: 0.012-0.015) (Figure 2A-B). Aside from discrimination performance measures, we also tested whether the model was calibrated. In a perfectly calibrated model, the distribution of the predicted probabilities is equal to the distribution of outcomes observed in the training data. We found that our primary model is well calibrated across the relevant prediction range (Figure 2C). The model has a positive predictive value (PPV) of 46.3%, at 10% sensitivity, while the baseline model only has 0.014%, at 10% sensitivity.

As an additional validation for the risk scores obtained, we compared the model predictions on the online survey data that was not used in the model construction process (n=121,151), with the actual number of confirmed COVID-19 patients in Israel over time. Notably, we found that the average predicted probability of individuals to test positive for COVID-19 according to our model, is highly correlated with the number of new confirmed COVID-19 cases 4 days later (pearson $r=0.90$, $p<10^{-8}$) (Figure 3).

To better understand which features contribute to the probability of being diagnosed with COVID-19, and to examine feature interactions, we analyzed the primary model constructed using the Gradient Boosting Decision Trees algorithm (see Methods). This model showed similar performance to the Logistic Regression model (Figure 2D-F, Table S3), with a positive predictive value (PPV) of 52.6%, at 10% sensitivity, while the baseline model only has 0.025%, at 10% sensitivity. Predictions on the online survey data not used in the model's construction process were highly correlated with the predictions of the primary Logistic Regression model (pearson $r=0.91$, $p<10^{-8}$).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Analysis of feature contributions was performed using Shapley values (see Methods). Loss of taste or smell and cough had the largest overall contribution to the model (Figure 4A), when analyzing the mean absolute SHAP value of features on the entire data. Since the primary model contained a limited number of features, we compared its feature contributions to those obtained from the extended features model, also constructed using a Gradient Boosting Decision Trees algorithm. Notably, loss of taste or smell was the most contributing feature in both the primary model and the extended features model, which contained 14 additional features (Figure 4), as well as in an odds ratio analysis (Figure S1). Although the extended features model included 23 features - 14 additional features over the primary model, all symptoms included in the primary model were among the 12 features the algorithm found to be most contributing (Figure 4B).

As age was reported as a dominant factor in COVID-19 infection and its clinical manifestation [11], we examined the interaction of age with every symptom using SHAP interaction values (see Methods). The contribution of age to the probability of being diagnosed with COVID-19 is the highest in the oldest age group (>70 years old) (Figure 5B). Presence of cough and loss of taste or smell exhibits a sharp transition-type (sigmoid-like) interaction with age, such that above the age of 40 years old, presence of each of these symptoms sharply increases the model's predicted probability of COVID-19 infection (Figure 5G-H). In contrast, shortness of breath and sore throat show a more gradual (parabolic-like) interaction with age with presence of these symptoms increasing the model's prediction more gradually as the age of the subject being predicted increases (Figure 5I-J). Negative answers to all these features show no interaction with age. Other examined features, such as fever and general feeling, do not show such interactions with age.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Discussion

In this study we constructed a model that predicts the probability of individuals to test positive for COVID-19. Our model is based on 9 simple questions that every person can easily answer in less than a minute from the comfort of their home. Our model can assist the worldwide fight against COVID-19 by better prioritizing the limited tests available without additional costs or risk of exposure to suspected patients, thereby increasing the rate at which positive individuals can be identified and isolated.

In Israel, as well as in many other countries, due to limited testing resources, suspected patients are only tested if they were exposed to a COVID-19 confirmed patient as well as exhibited acute respiratory symptoms [3]. By taking an unbiased approach to predicting COVID-19 diagnosis from symptoms data, our analysis highlights the importance of additional symptoms. Of note, anosmia and ageusia that were less described in patients in the early stages of the COVID-19 pandemic [4,12] were the most impactful features in both models for predicting COVID-19 diagnosis as well as in an odds-ratio analysis. This is in line with recent literature demonstrating the importance of these symptoms in early detection and identification of the disease [13,14,15]. Our model also successfully recapitulated patterns of the disease that are described in the literature, such as its complex relationship with age [11]. In addition, our model unraveled several patterns that are not described in the literature, such as the different patterns of interactions that particular symptoms have with age, suggesting variation of the clinical manifestation in different age groups. Although our analysis is purely predictive and not causal, these new patterns may be used to devise better testing policies, and pave the way for future studies that can uncover new aspects of the disease that were not studied to date.

Analysis of an extended features model that included 23 features compared to 9 in the primary model, validated our choice of questions in the shortened version of the survey and suggested that fatigue should also be considered. In addition, the extended features model suggested that while dry cough has an essential role in predicting COVID-19 diagnosis, moist cough does not and thus may help distinguish between cases of COVID-19 and other infections. Some of the most contributing symptoms to the prediction of a COVID-19 diagnosis are

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

currently not included in the Israeli testing policy [3], such as loss of taste or smell and sore throat. Our analysis suggests that adding these symptoms to the testing policy may help discriminate which individuals should be tested, and improve testing prioritization.

While informative, our feature contribution analysis has several limitations. First, we did not include children in our datasets and thus, symptoms such as nausea or vomiting and diarrhea that were mostly described in children [7,8], may have a more significant part in models designed for younger age groups. Second, although we included a large set of prior medical conditions that may have a role in COVID-19 susceptibility, some of these conditions are not highly prevalent in our dataset and their contribution may thus be underestimated in our model. Finally, body temperature was the only non-mandatory question in our survey, and may thus have higher predictive power than portrayed within our model.

Several studies attempted to simulate and predict different aspects of COVID-19, such as hospital admissions, diagnosis, prognosis and mortality risk, using mostly age, body temperature, medical tests and symptoms [16]. Most diagnostic models published to date were based on datasets from China and included complex features that had to be extracted through blood tests and imaging scans [16]. In this work, we devised a prediction model which was based solely on self-reported information, and as such it could be easily deployed and used instantly in other countries.

Our study has several additional limitations. First, our data is biased by Israel's MOH ever changing testing policy, such that at some point all of the COVID-19 positively diagnosed participants in our study had to be eligible for a test under that policy. An ideal dataset for purposes of devising a classifier should include a large random sampling of the population, but such data coupled with symptom surveys is currently unavailable at large-scale. Accordingly, all the diagnosed responders in our study are not in the first stage of showing symptoms, but are in some time-lag after diagnosis. In addition, our study is based on self reports of willing participants and is therefore bound to suffer from some selection bias. The bias is significantly reduced in the data collected via the IVR platform, since all residents in the IVR-surveyed cities were actively contacted only

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

once, on the same day and in the same manner. In the online version of the survey we made attempts to reduce this bias by promoting it in several media outlets and by engaging leaders of underrepresented communities. Finally, we acknowledge that the ideal choice of a baseline model would be the Israeli MOH testing policy, but since it constantly changed during the pandemic and included questions that were not part of our survey we were unable to compare to it. Nevertheless, at an unknown sensitivity (as we do not know the actual number of cases), the overall predictive value (PPV) of all tests in Israel was 4.6% [3], and although the measures are not directly comparable, we believe that our model substantially outperforms this PPV.

In conclusion, our constructed model predicts COVID-19 PCR test results with high discrimination (positive predictive value (PPV) of 46.3% at a 10% sensitivity) and calibration. It also suggests that several symptoms that are currently not included in the Israeli testing policy exhibit intriguing interactions with age and should probably be integrated into revised testing policies. Overall, our approach can be utilized worldwide to direct the limited resources towards individuals who are more likely to test positive for COVID-19, leading to faster isolation of infected patients and therefore to reduced rates of virus spread.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Figures and Tables

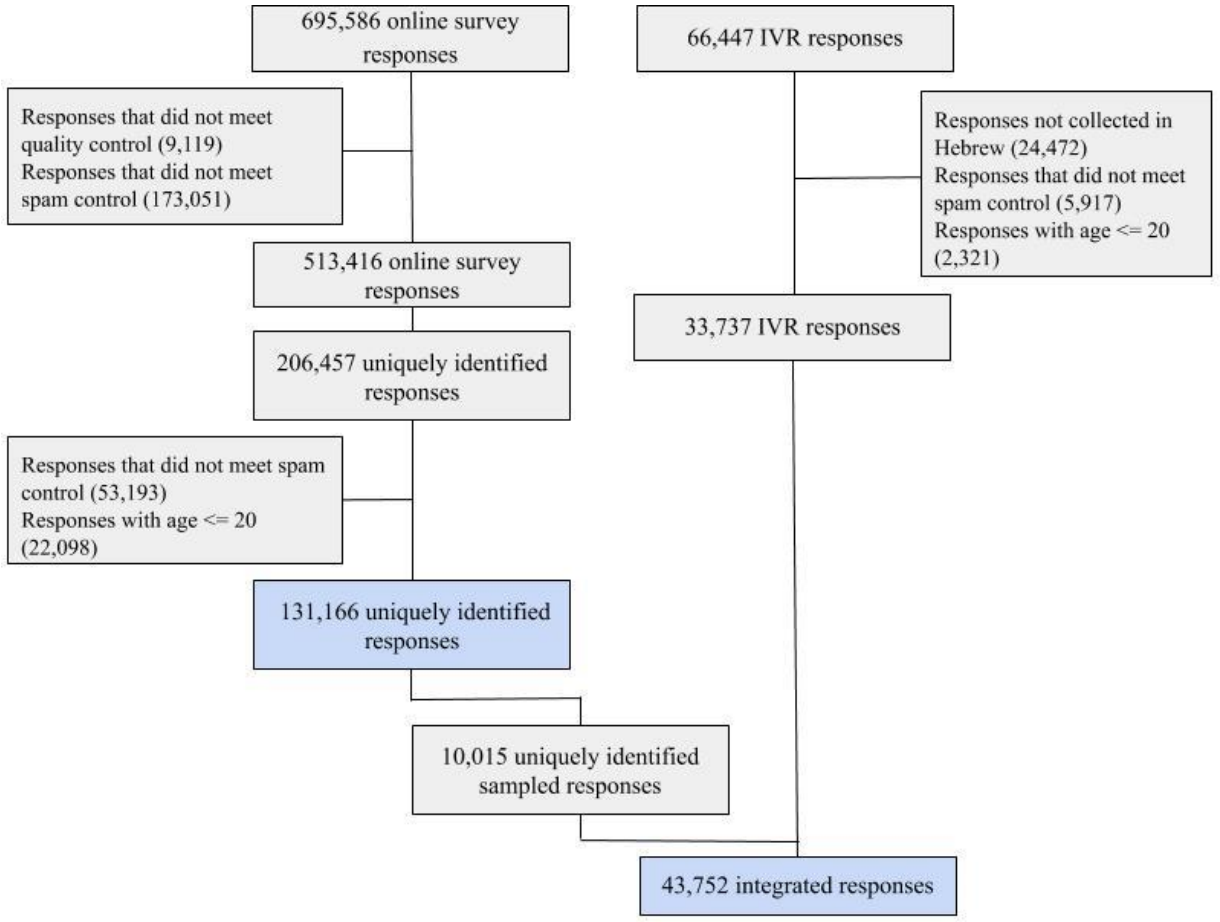



**Figure 1. Study population flow chart.** Numbers represent recorded responses. Blue colored boxes show reponses which were used in extended features model (top) and primary model (bottom) constructions.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

**Table 1. Baseline characteristics of the primary model population**

| **Characteristic, mean (SD) or %** | **All individuals n=43,752 (100%)** | **IVR version n=33,737 (77.11%)** | **Online version n=10,015 (22.89%)** | **COVID-19 undiagnosed n=43,254 (98.862%)** | **COVID-19 diagnosed n=498 (1.138%)** |
|---|---|---|---|---|---|
| Age in years | 44.941 (15.499) | 43.897 (15.244) | 48.460 (15.831) | 44.894 (15.47) | 49.076 (17.363) |
| Gender - male | 23,630 (54.009%) | 19,151 (56.766%) | 4,479 (44.723%) | 23,339 (53.958%) | 291 (58.434%) |
| COVID-19 diagnosed | 498 (1.138%) | 384 (1.138%) | 114 (1.138%) | 0 (0.0%) | 498 (100.0%) |
| Prior medical conditions | 8,070 (18.943%) | 5,176 (15.884%) | 2,894 (28.897%) | 7,946 (18.861%) | 124 (26.271%) |
| **Symptoms** | | | | | |
| Feel well | 41,661 (95.221%) | 32,132 (95.243%) | 9,529 (95.147%) | 41,217 (95.291%) | 444 (89.157%) |
| Sore throat | 1,507 (3.445%) | 1,141 (3.382%) | 366 (3.655%) | 1,422 (3.288%) | 85 (17.068%) |
| Cough | 2,138 (4.887%) | 1,459 (4.325%) | 679 (6.78%) | 1,984 (4.587%) | 154 (30.924%) |
| Shortness of breath | 576 (1.317%) | 443 (1.313%) | 133 (1.328%) | 505 (1.168%) | 71 (14.257%) |
| Loss of taste or smell | 605 (1.388%) | 545 (1.624%) | 60 (0.599%) | 469 (1.088%) | 136 (27.812%) |
| Fever (body temperature above 38 °C) | 77 (0.176%) | 53 (0.157%) | 24 (0.24%) | 64 (0.148%) | 13 (2.61%) |

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

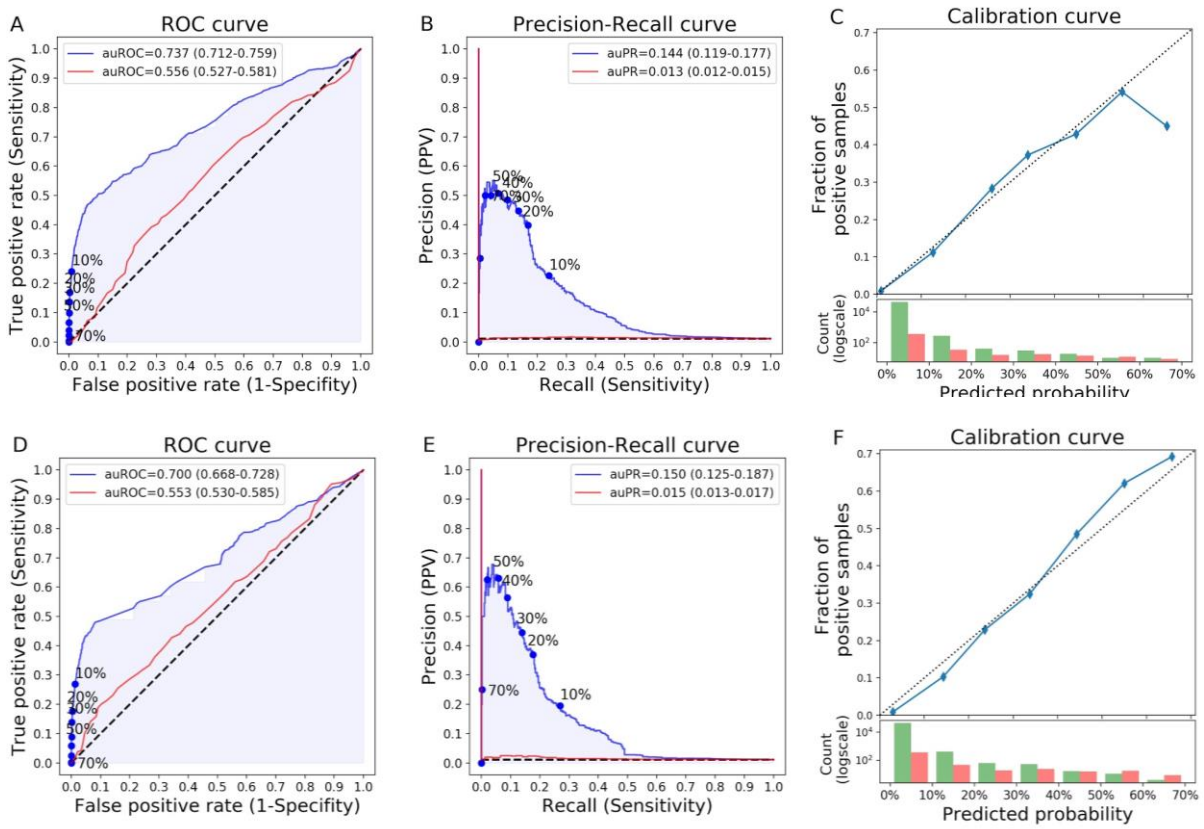



**Figure 2. Primary model performance. A-C:** Logistic Regression, **D-F:** Gradient Boosting Decision Trees. Abbreviations: auROC/auPR - area under the ROC/PR curve, ROC - Receiver-Operator-Characteristic, PR - Precision-Recall. Parenthesis - Confidence Interval. **A, D:** ROC curve of our model (blue) consisting of 9 simple questions and of the baseline model consisting of only age group and gender (red). Different decision probability thresholds are marked on the curve. **B, E:** Precision-Recall curve of our model (blue) and the baseline model (red). Different decision probability thresholds are marked on the curve. **C, F:** Calibration curve. Top: Blue dots represent deciles of predicted probabilities. Dotted diagonal line represents an ideal calibration. Bottom: Log-scaled histogram of predicted probabilities of COVID-19 undiagnosed (green) and diagnosed (red).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

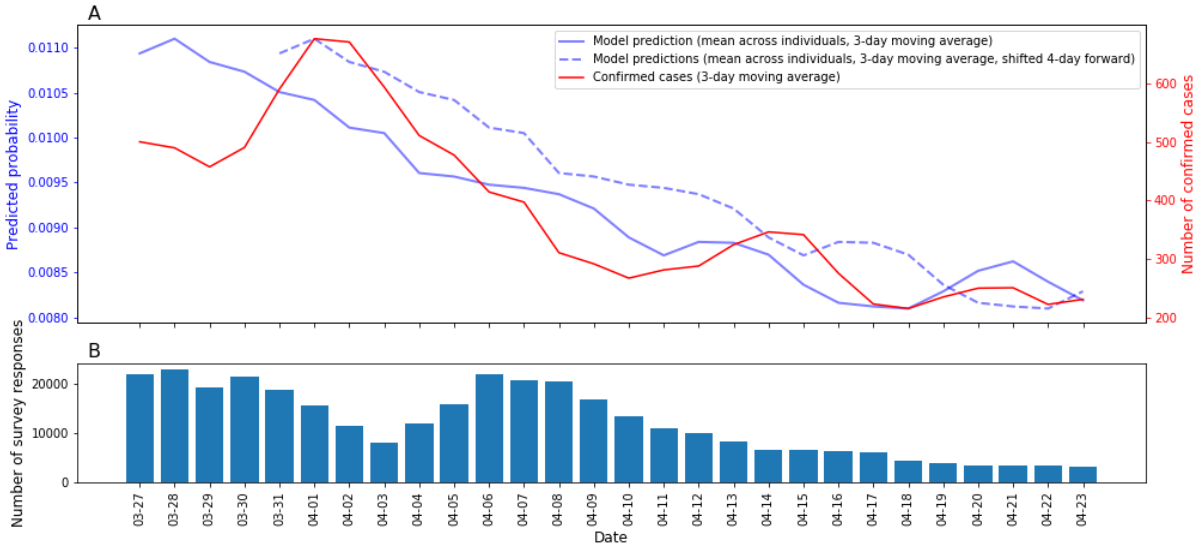



**Figure 3. Comparison of primary model predictions to new COVID-19 cases in Israel over time.**
**A:** Primary model predictions, averaged across all individuals on a 3-day running average (solid blue), and shifted 4 days forward (dotted blue), compared to the number of newly confirmed COVID-19 cases in Israel by the ministry of health (MOH), based on a 3-day running average. **B:** Number of survey responses per day.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

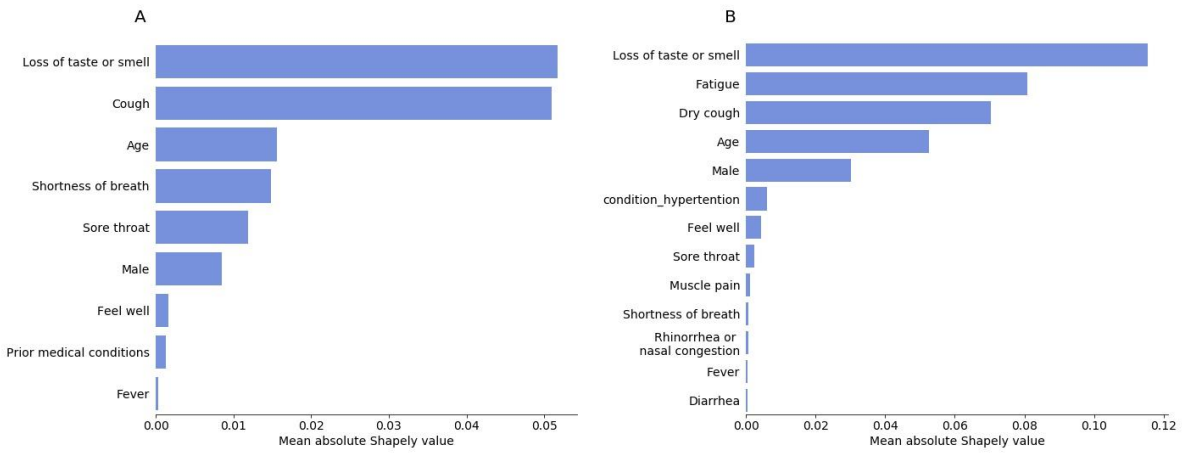



**Figure 4. Feature contribution analysis.** Mean absolute Shapley value (in units of log-odds) of **A:** the Primary model, including all features used in the model, and **B:** the Extended features model, for the 13 highest contributing features.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

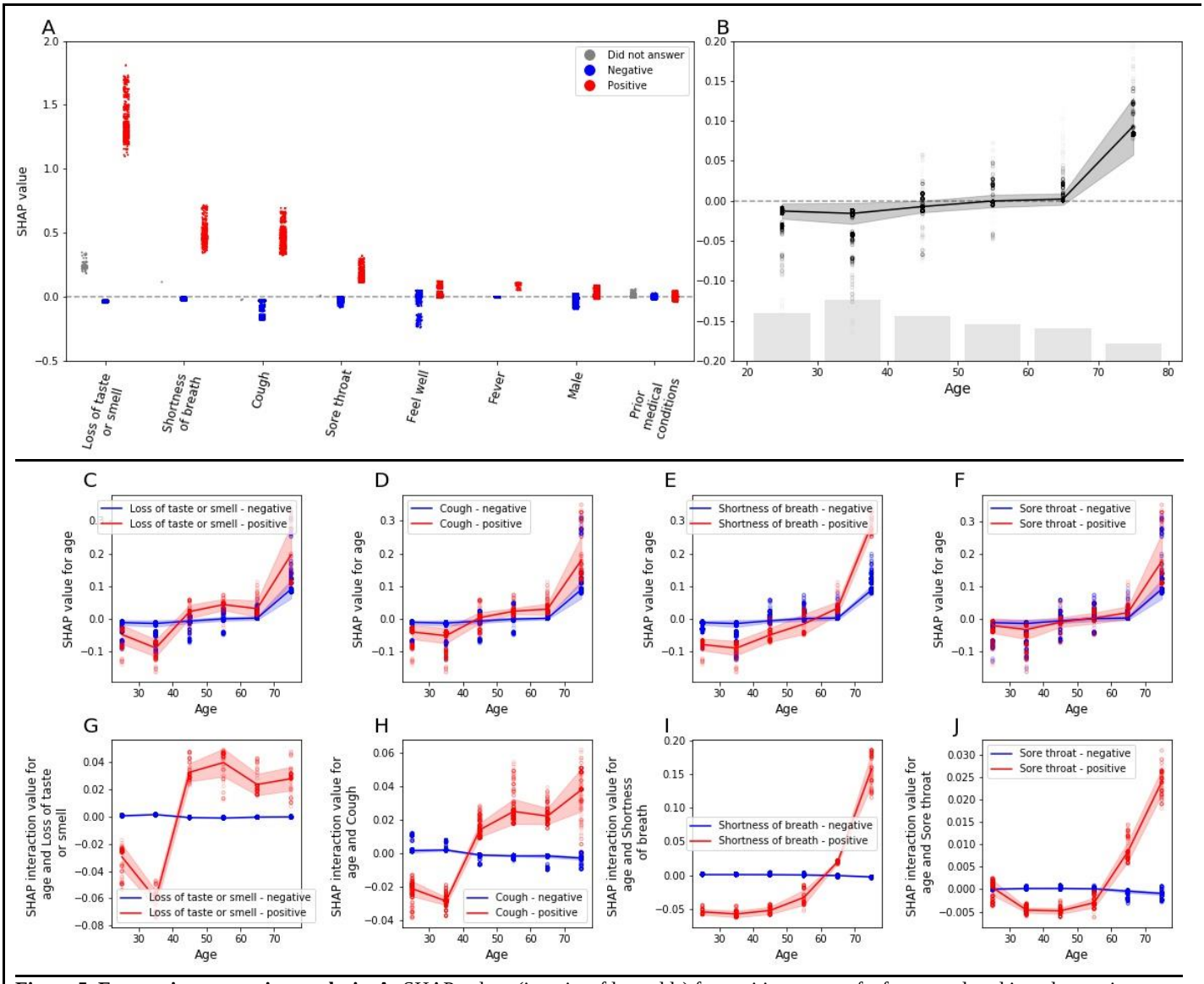



**Figure 5. Feature interpretation analysis. A:** SHAP values (in units of log-odds) for positive report of a feature colored in red, negative report of a feature colored in blue and missing answers in grey. **B:** SHAP values for age with number of responses as a histogram at the bottom. **C-F:** SHAP value for age, stratified by positive (red) and negative (blue) responses of loss of taste or smell (C), cough (D), shortness of breath (E) and sore throat (F). **G-J:** SHAP interaction values of age with positive (red) and negative (blue) responses of loss of taste or smell (G), cough (H), shortness of breath (I) and sore throat (J).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Data availability statement

Models and tables of de-identified, aggregated data are available at https://github.com/hrossman/Covid19-Survey.

## Code availability statement

Source code for the survey is available at https://github.com/hasadna/avid-covider as an open source project, and can be readily adapted to use in other countries.

## Ethics Declarations

The study protocol was approved by the Weizmann Institute of Science review board (IRB). Informed consent was waived by the IRB, as all identifying details of the participants were removed before the computational analysis. Participants were made fully aware of the way in which the data will be stored, handled and shared, which was provided to them and is in accord with the privacy and data-protection policy of the Weizmann Institute of Science (https://weizmann.ac.il/pages/privacy-policy).

## Competing Interests Statement

The authors declare no competing interests.

## Authors contribution

S.S, T.K., A.K., S.S. and H.R. conceived the project, designed and conducted the analyses, interpreted the results and wrote the manuscript. A.G., T.M., A.L., D.K., I.K., A.G. designed and conducted the survey. O.C. directed the organizational and logistic efforts. A.K., O.H., M.Z.A., N.C., A.E.Z., A.I. designed and conducted the survey as well as provided the data. B.G., D.H., V.S., R.B. and E.S. conceived, directed and supervised the project and analyses.

## Acknowledgments

We thank the following for their contributions to our efforts: Y. Landau, T. Eldar, S. Kasem, T. Bria, S. Avraham, B. Kirel, A. Terkeltaub, N. Kalkstein, E. Krupka, M. Loitsker, Y. Sela, T. Bareia, J. Shalev, A. Brik, L. Samama, S. Cohen, J. Bitton, R. Priborkin, G. Beryozkin, G. Dardyk, S. Lanzman, N. Epstein, N. B. Shizaf, E. Daian, M. Bialek, M. Suleiman, R. Maor, Z. Maor, Y. Dinur, J. Klinger, R. Zaida, M. Eden, N. Hirshman, O. Dovev, Y. Grossman and the Public Knowledge Workshop ('Hasadna').

medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## References


1. Xie J, Tong Z, Guan X, Du B, Qiu H, Slutsky AS. Critical care crisis and some recommendations during the COVID-19 epidemic in China. *Intensive Care Med*. March 2020. doi:10.1007/s00134-020-05979-7

2. Grasselli G, Pesenti A, Cecconi M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. *JAMA*. March 2020. doi:10.1001/jama.2020.4031

3. Israeli Ministry of Health. https://www.health.gov.il/English/Pages/HomePage.aspx. Accessed April 26, 2020.

4. Zhao X, Zhang B, Li P, et al. Incidence, clinical characteristics and prognostic factor of patients with COVID-19: a systematic review and meta-analysis. *medRxiv*. January 2020.

5. Rossman H, Keshet A, Shilo S, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nat Med*. April 2020. doi:10.1038/s41591-020-0857-9

6. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594. doi:10.1136/bmj.g7594

7. Dong Y, Mo X, Hu Y, et al. Epidemiological characteristics of 2143 pediatric patients with 2019 coronavirus disease in China. *Pediatrics*. March 2020. doi:10.1542/peds.2020-0702

8. de Souza TH, Nadal JA, Nogueira RJN, Pereira RM, Brandao MB. Clinical Manifestations of Children with COVID-19: a Systematic Review. *medRxiv*. April 2020. doi:10.1101/2020.04.01.20049833

9. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. New York, New York, USA: ACM Press; 2016:785-794. doi:10.1145/2939672.2939785

10. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. January 2020. doi:10.1038/s42256-019-0138-9

11. Harapan H, Itoh N, Yufika A, et al. Coronavirus disease 2019 (COVID-19): A literature review. *J Infect Public Health*. 2020;13(5):667-673. doi:10.1016/j.jiph.2020.03.019

12. Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med*. April 2020. doi:10.1056/NEJMoa2006100

13. Menni C, Valdes A, Freydin MB, et al. Loss of smell and taste in combination with other symptoms is a strong predictor of COVID-19 infection. *medRxiv*. April 2020. doi:10.1101/2020.04.05.20048421

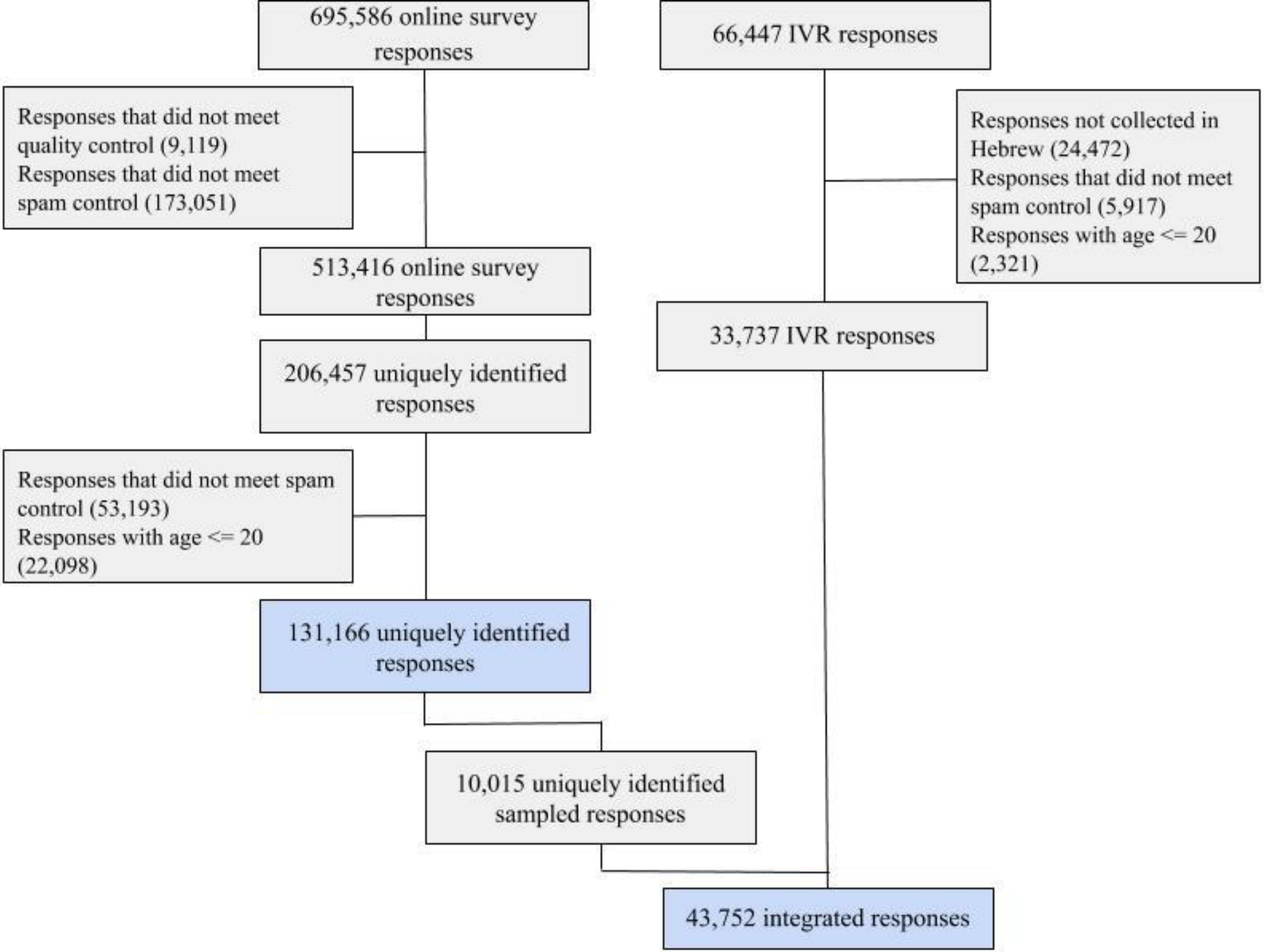
14. Yan CH, Faraji F, Prajapati DP, Boone CE, DeConde AS. Association of chemosensory dysfunction and Covid-19 in patients presenting with influenza-like symptoms. *Int Forum Allergy Rhinol*. April 2020. doi:10.1002/alr.22579

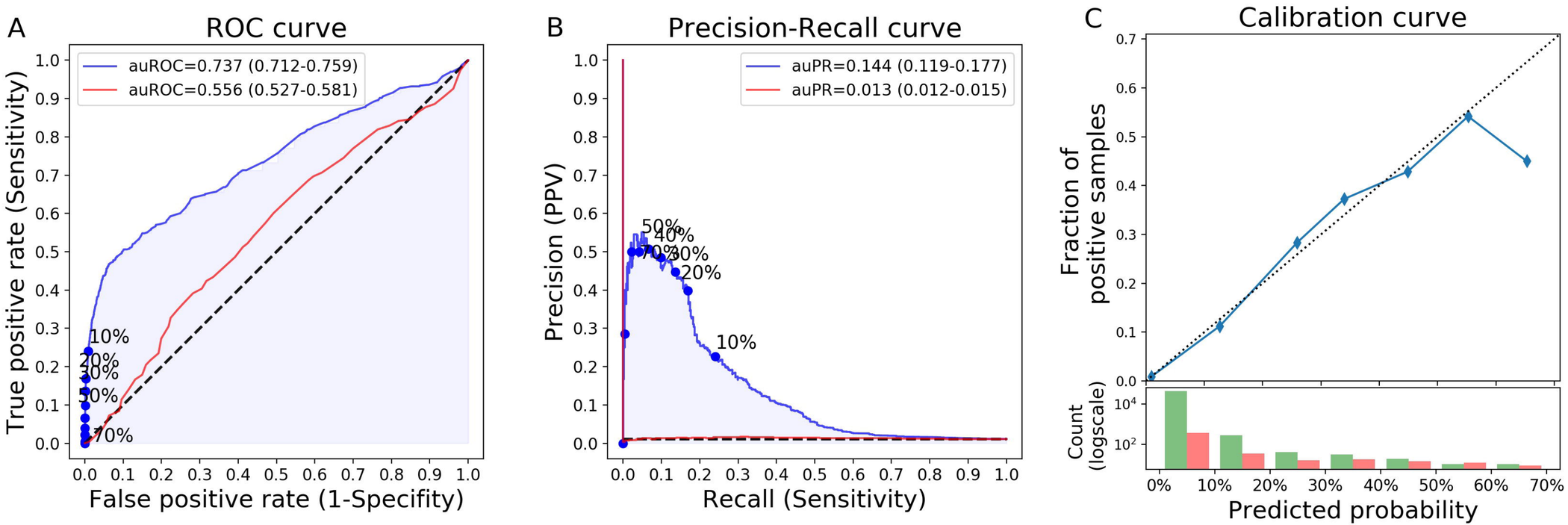
15. Lechien JR, Chiesa-Estomba CM, De Siati DR, et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. *Eur Arch Otorhinolaryngol*. April 2020. doi:10.1007/s00405-020-05965-1

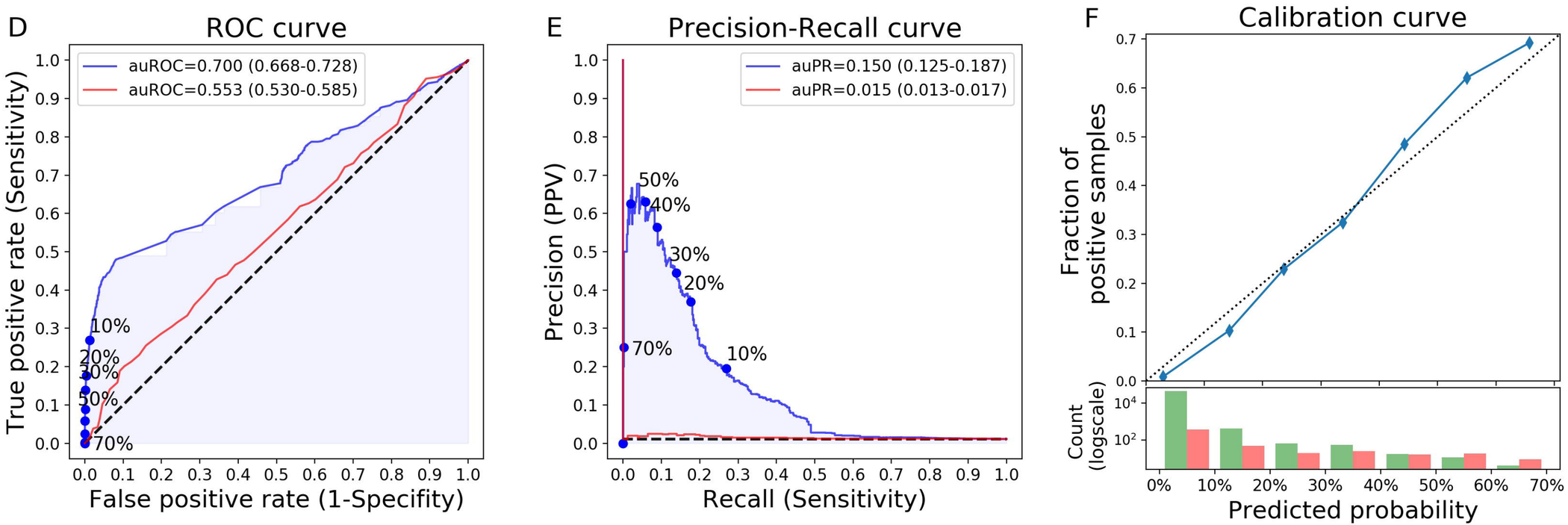
16. Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:m1328. doi:10.1136/bmj.m1328

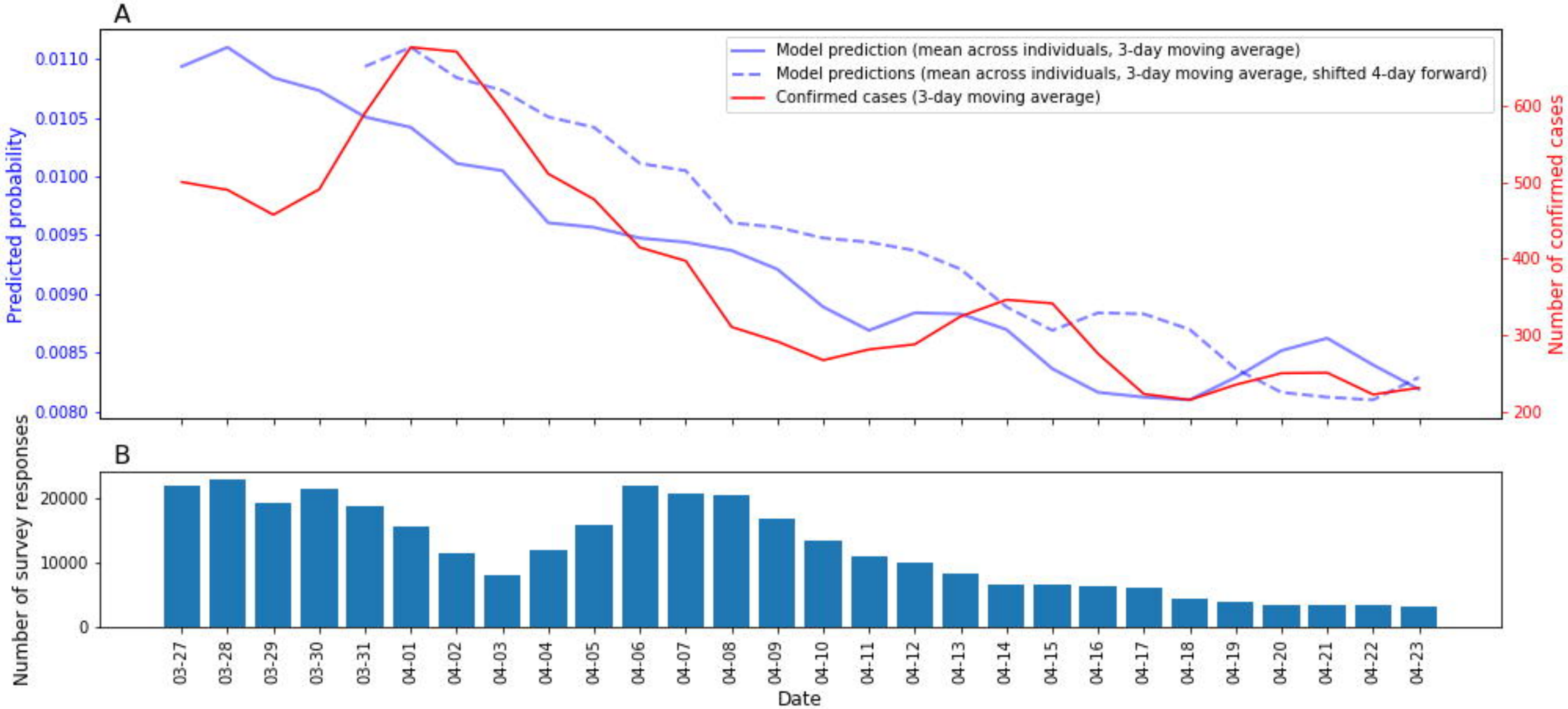
medRxiv preprint doi: https://doi.org/10.1101/2020.05.18.20105569; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

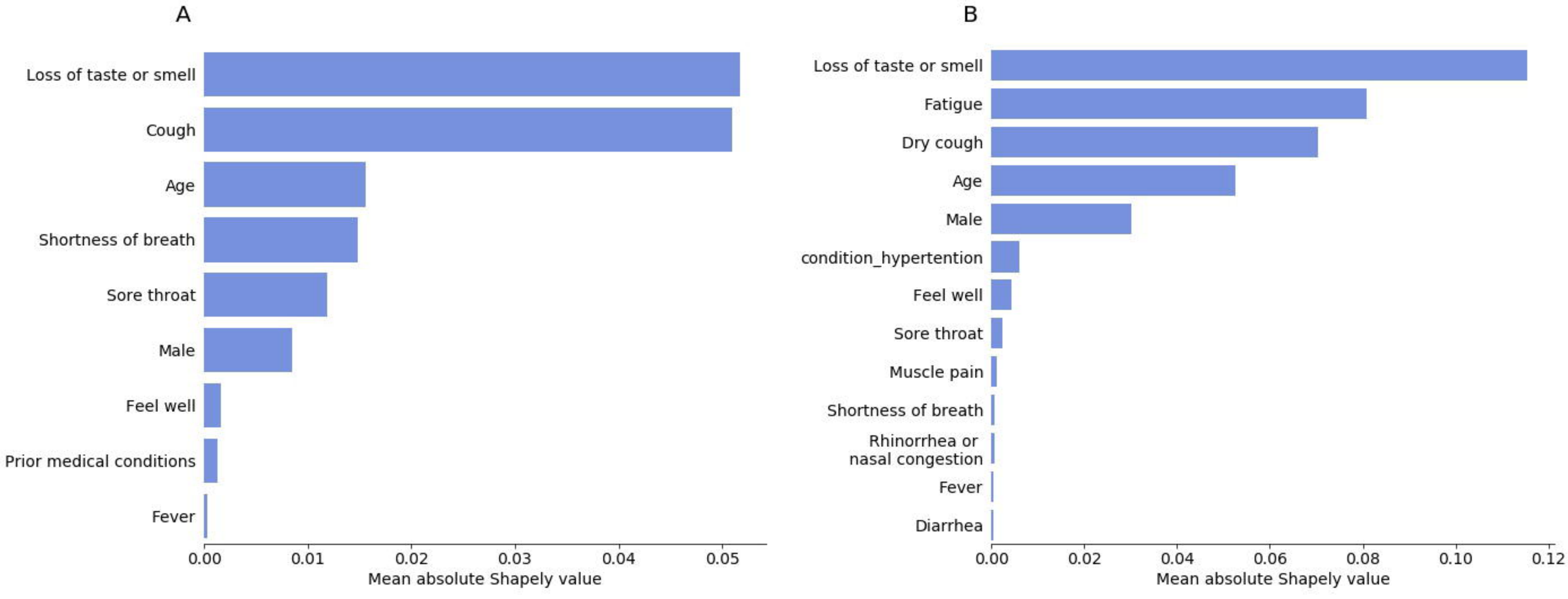
17. CBS Site. https://www.cbs.gov.il/en/pages/default.aspx. Accessed April 26, 2020.

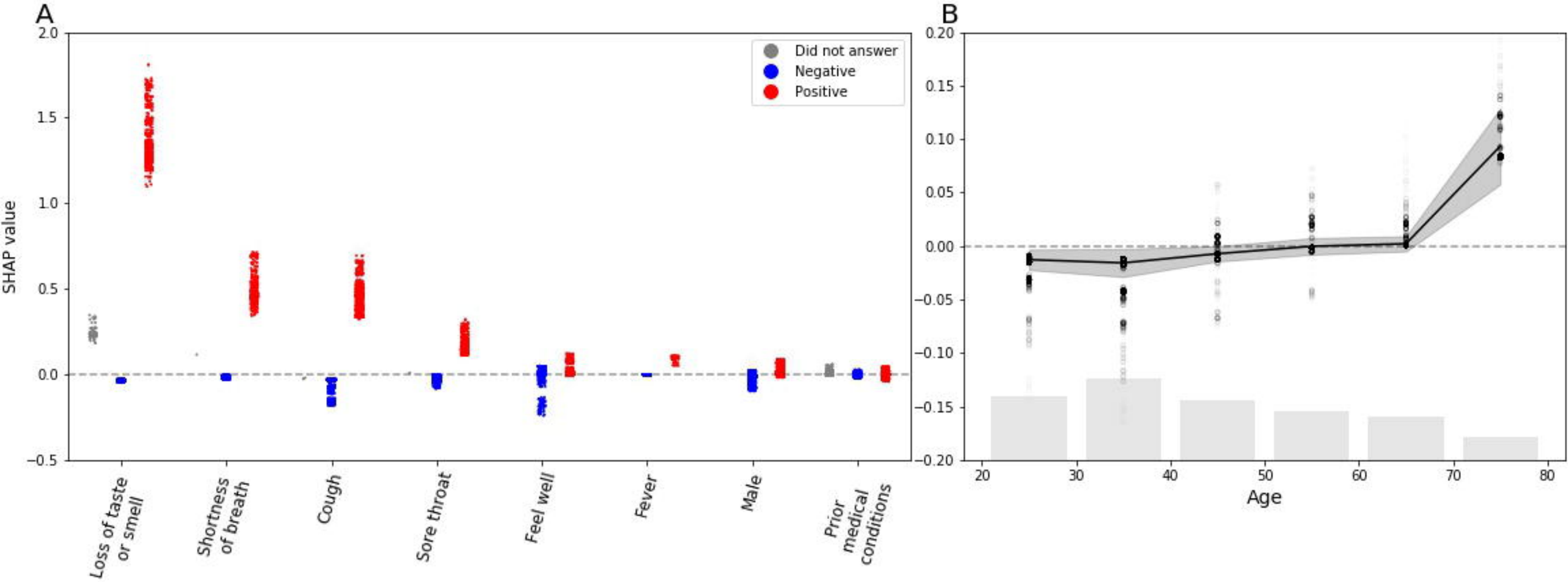

695,586 online survey responses
Responses that did not meet quality control (9,119)
Responses that did not meet spam control (173,051)
513,416 online survey responses
206,457 uniquely identified responses
Responses that did not meet spam control (53,193)
Responses with age <= 20 (22,098)
131,166 uniquely identified responses
10,015 uniquely identified sampled responses
66,447 IVR responses
Responses not collected in Hebrew (24,472)
Responses that did not meet spam control (5,917)
Responses with age <= 20 (2,321)
33,737 IVR responses
43,752 integrated responses

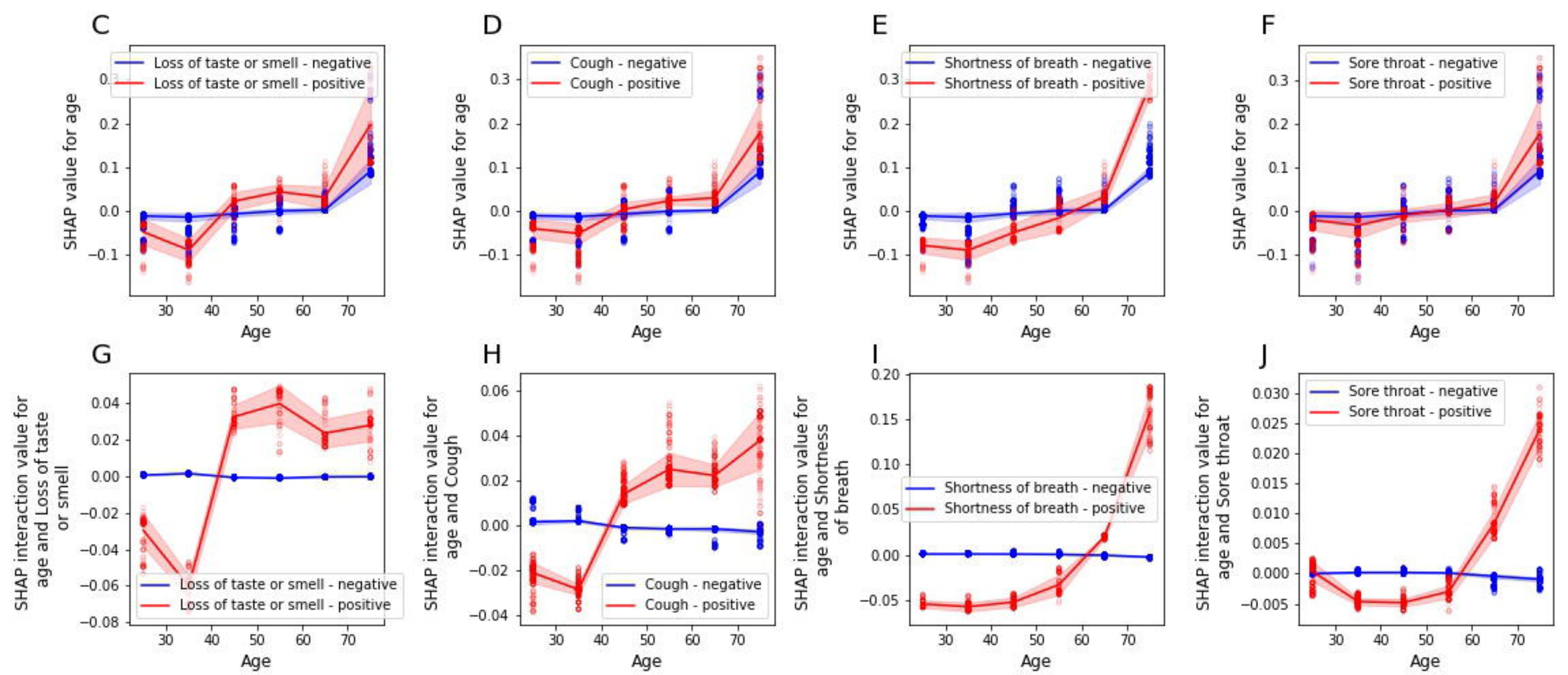

A
ROC curve
auROC=0.737 (0.712-0.759)
auROC=0.556 (0.527-0.581)
True positive rate (Sensitivity)
False positive rate (1-Specifity)
10%
20%
30%
50%
70%
B
Precision-Recall curve
auPR=0.144 (0.119-0.177)
auPR=0.013 (0.012-0.015)
Precision (PPV)
Recall (Sensitivity)
50%
40%
70%
30%
20%
10%
C
Calibration curve
Fraction of positive samples
Count (logscale)
Predicted probability

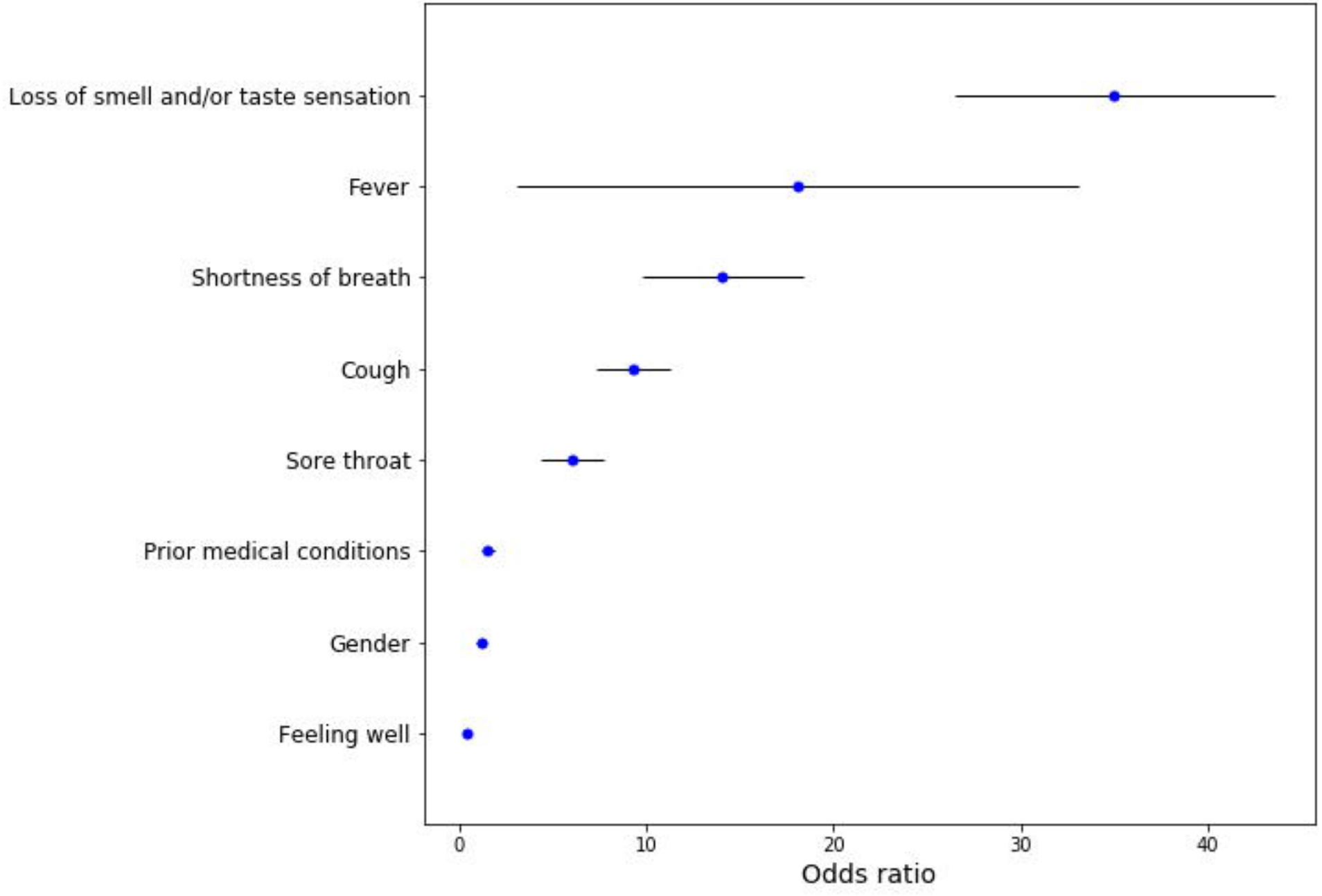

D
ROC curve
auROC=0.700 (0.668-0.728)
auROC=0.553 (0.530-0.585)
True positive rate (Sensitivity)
False positive rate (1-Specifity)
10%
20%
30%
50%
70%
E
Precision-Recall curve
auPR=0.150 (0.125-0.187)
auPR=0.015 (0.013-0.017)
Precision (PPV)
Recall (Sensitivity)
50%
40%
30%
20%
10%
70%
F
Calibration curve
Fraction of positive samples
Count (logscale)
Predicted probability


A
Model prediction (mean across individuals, 3-day moving average)
Model predictions (mean across individuals, 3-day moving average, shifted 4-day forward)
Confirmed cases (3-day moving average)
Predicted probability
0.0110
0.0105
0.0100
0.0095
0.0090
0.0085
0.0080
Number of confirmed cases
600
500
400
300
200
B
Number of survey responses
20000
10000
0
03-27
03-28
03-29
03-30
03-31
04-01
04-02
04-03
04-04
04-05
04-06
04-07
04-08
04-09
04-10
04-11
04-12
04-13
04-14
04-15
04-16
04-17
04-18
04-19
04-20
04-21
04-22
04-23
Date


A
Loss of taste or smell
Cough
Age
Shortness of breath
Sore throat
Male
Feel well
Prior medical conditions
Fever
0.00
0.01
0.02
0.03
0.04
0.05
Mean absolute Shapely value
B
Loss of taste or smell
Fatigue
Dry cough
Age
Male
condition_hypertention
Feel well
Sore throat
Muscle pain
Shortness of breath
Rhinorrhea or nasal congestion
Fever
Diarrhea
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Mean absolute Shapely value


A
B
SHAP value
Did not answer
Negative
Positive
Loss of taste or smell
Shortness of breath
Cough
Sore throat
Feel well
Fever
Male
Prior medical conditions
Age


C
Loss of taste or smell - negative
Loss of taste or smell - positive
SHAP value for age
Age
D
Cough - negative
Cough - positive
SHAP value for age
Age
E
Shortness of breath - negative
Shortness of breath - positive
SHAP value for age
Age
F
Sore throat - negative
Sore throat - positive
SHAP value for age
Age
G
SHAP interaction value for age and Loss of taste or smell
Loss of taste or smell - negative
Loss of taste or smell - positive
Age
H
SHAP interaction value for age and Cough
Cough - negative
Cough - positive
Age
I
SHAP interaction value for age and Shortness of breath
Shortness of breath - negative
Shortness of breath - positive
Age
J
SHAP interaction value for age and Sore throat
Sore throat - negative
Sore throat - positive
Age


Loss of smell and/or taste sensation
Fever
Shortness of breath
Cough
Sore throat
Prior medical conditions
Gender
Feeling well
0
10
20
30
40
Odds ratio