Using Survey Data to Estimate the Impact of the Omicron Variant on Vaccine Efficacy against COVID-19 Infection*

Jesús Rufino[2,1], Carlos Baquero[3,1], Davide Frey[4,1], Christin A. Glorioso[5,1], Antonio Ortega[6,1], Nina Reščič[7,1], Julian Charles Roberts[8,1], Rosa E. Lillo[9,1], Raquel Menezes[10,1], Jaya Prakash Champati[2,1], and Antonio Fernández Anta[2,1]

[1]CoronaSurveys Team, Spain
[2]IMDEA Networks Institute, Spain
[3]U. Porto & INESC TEC, Portugal
[4]Univ Rennes, IRISA, CNRS, Inria, 35042 Rennes, France
[5]Academics for the Future of Science, Inc. & U. of California San Francisco, USA
[6]U. Southern California, USA
[7]Jožef Stefan Institute, Department of Intelligent Systems, Ljubljana, Slovenia
[8]Gearu LTD, UK
[9]U. Carlos III de Madrid, Spain
[10]Centre of Mathematics of U. Minho, Portugal

Abstract

Data collected in the Global COVID-19 Trends and Impact Surveys (UMD Global CTIS), and data on variants sequencing from GISAID, are used to evaluate the impact of the Omicron variant (in South Africa and other countries) on the prevalence of COVID-19 among unvaccinated and vaccinated population, in general and discriminating by the number of doses. In South Africa, we observe that the prevalence of COVID-19 in December (with strong presence of Omicron) among the unvaccinated population is comparable to the prevalence during the previous wave (in August-September), in which Delta was the variant with the largest presence. However, among vaccinated, the prevalence of COVID-19 in December is much higher than in the previous wave. In fact, a significant reduction of the vaccine efficacy is observed from August-September to December. For instance, the efficacy drops from 0.81 to 0.30 for those vaccinated with 2 doses, and from 0.51 to 0.09 for those vaccinated with one dose. The study is then extended to other countries in which Omicron has been detected, comparing the situation in October (before Omicron) with that of December. While the reduction measured is smaller than in South Africa, we still found, for instance, an average drop in vaccine efficacy from 0.53 to 0.45 among those vaccinated with two doses. Moreover, we found a significant negative (Pearson) correlation of around -0.6 between the measured prevalence of Omicron and the vaccine efficacy.

1 Introduction

The Omicron variant of SARS-CoV-2 has seen an expressive increase since its initial classification in November 2021 [Oo21]. In South Africa it appears to have out-competed the Delta variant [Hod21] and has rapidly spread into Europe and other regions. Preliminary observations also indicate that it might spread faster and might have higher immune evasiveness than previous variants [KK21]. While vaccination still provides a level of protection against a serious disease [RHRM+21], recent results [PvSG+21, NKL+21, KST+21, LMD+21] point towards a reduced level of protection against infection, especially from 15 weeks post the second dose [ASK+21], and it is likely that the number of breakthrough infections (i.e., infections among vaccinated people) will rise with the spread of Omicron. It is also possible that the rapid spread of Omicron is not only a consequence of high transmissibility but also of immune evasiveness [LMD+21]. Some of the preliminary models [SLD+22] showed that high transmissibility in combination with high immune evasiveness could lead to a concerning health system overload [LRSC+21].

Since the spring of 2020, the University of Maryland in collaboration with Facebook has collected extensive survey data on self-reported symptoms, infection, testing behavior, and more recently, vaccination status (UMD

*This work is partially supported by grant CoronaSurveys-CM, funded by IMDEA Networks and Comunidad de Madrid, Spain, and individual donations to the CoronaSurveys Project https://coronasurveys.org.

Global CTIS) [FLS+20, The21b]. In mid December 2021, researchers used data from this survey concerning the Gauteng province in South Africa to define different combinations of symptoms that are associated with COVID-19 infection, and combined those with self-reported vaccination status to compare vaccine efficacy changes from a Delta dominant period to the current Omicron dominant period [VRAB21]. Their findings showed a measurable drop of efficacy towards infection for those vaccinated with two doses.

In this study we use self-reported confirmation of COVID-19 infection, from a subset of the UMD Global CTIS survey responses, to derive an improved proxy for COVID-19 active cases (using a Random Forest classifier) that tracks more closely the evolution of confirmed cases. We use this improved proxy for analysing prevalence and vaccine efficacy changes in South Africa as a whole, and in the Gauteng province, among those unvaccinated, partially vaccinated, and fully vaccinated. We also compute results in other countries that are currently experiencing a rise of Omicron cases, which show a significant negative correlation between the prevalence of Omicron and the vaccine efficacy.

The rest of the paper has three sections. In Section 2 the data used and the methodology applied is described. In Section 3 we describe the results obtained when applying the methodology to the data. Finally, in Section 4 we have a discussion about the implications of the results obtained.

2 Methods

2.1 Self-reported Survey Data

Since Spring 2020, the U. of Maryland (UMD) has been running a COVID-19-related survey [FLS+20, The21b] in most countries[1], in collaboration with Facebook [Fac20, KBB+20, ATMC+21]. This survey, called the *University of Maryland Social Data Science Center Global COVID-19 Trends and Impact Survey in partnership with Facebook* (UMD Global CTIS), collects more than 100,000 responses daily across the world. It asks the participants questions covering, among others aspects: symptoms, habits, testing, and vaccination status. All the participants in the CTIS have declared to be at least 18 years of age.

In this work, we use the responses to the UMD Global CTIS, to which we have access by agreement with UMD and Facebook (see Appendix D). We first curate the data by removing abnormal responses, following the approach proposed by Alvarez et al. [ÁBC+21]: We remove responses that declare to have all symptoms or that declare unusual values (greater than 100) in the quantitative questions of the survey (e.g., days of symptom duration, number of symptomatic contacts, number of people staying at the same place, etc.).

After curating the responses, the next task we face is determining whether they correspond to active cases of COVID-19. This is somewhat direct for the subset of responses that respond affirmatively to the survey question "B7: Have you been tested for COVID-19 in the past 14 days?" and then respond positively or negatively to the survey question "B8a: Did your most recent test find that you had COVID-19?" [The21a]. For this work, we assume that a participant responding affirmatively to both questions is an active case of COVID-19 (i.e., it is a *positive* case). Similarly, a participant responding affirmatively to Question B7 and negatively to Question B8a is assumed not infected with COVID-19 (i.e., *negative*). This set of classified responses constitute a *ground-truth set*, for which infection status (positive or negative) is available.

Unfortunately, this ground-truth set cannot be used directly to estimate the prevalence of COVID-19 in the overall population, because the set is usually very small and is not produced via uniform random sampling: People who have reason to believe they may be infected are more likely to be tested and therefore the ratio of positives among those tested in the latest 14 days (i.e., the *testing positive rate*, abbreviated TPR) is higher than the actual prevalence.

In order to classify the responses as positive or negative, several criteria have been proposed in the literature. In particular, we consider the following symptom-based COVID-like illness classifiers (see Appendix A for the list of symptoms collected in the survey):

- UMD CLI [FLS+20, ÁBC+21]: A response is considered to be positive if it declares fever (symptom B1_1), along with cough (symptom B1_2), or shortness of breath / difficulty breathing (symptom B1_3). Otherwise, it is negative.
- Stringent CLI [VRAB21]: A response is positive if it declares anosmia (symptom B1_10), combined with fever (B1_1), muscle pain (B1_6), or cough (B1_2). Otherwise, it is negative.
- Classic CLI [VRAB21]: A response is positive if it declares cough (B1_2), combined with fever (B1_1), muscle pain (B1_6), or anosmia (B1_10). Otherwise, it is negative.
- Broad CLI [VRAB21]: A response is positive if it declares muscle pain (B1_6), combined with fever (B1_1), cough (B1_2), or anosmia (B1_10). Otherwise, it is negative.

[1] Except in the US, where the survey has been run by CMU [Del20, SRB+21].

These methods for classifying cases as positive or negative have two main limitations. First, they do not take into account diagnostic uncertainty, e.g., the same set of symptoms might be associated with some other condition. Second, these criteria are not adaptive to possible changes in the symptoms experienced as conditions change, e.g., as vaccination rates increase or new virus variants emerge. Thus, in this work, we introduce a new machine-learning-based classifier (described in Section 2.2) where the responses of users in the ground-truth set are used to train a model, which is then used to determine the status of users outside that set (users who do not report test information). We use the random-forest technique to design this classifier and the corresponding results are labeled Random Forest in what follows.

We refer to the values obtained with each of these five classifiers (namely, Random Forest, UMD CLI, Stringent CLI, Classic CLI, and Broad CLI) as *proxy estimates* (or proxy for short). We compare each proxy estimate with the estimate of active cases obtained from the official number of cases as described by Alvarez et al. [ÁBC+21], where each new case is assumed to remain active for 10 days. These last estimates are called Confirmed. Both Confirmed and the estimates using the various proxies lead to time series with one estimated value per day.

2.2 Machine Learning Classifier: Random Forest

Each response to the survey includes a large number of questions (obviously, not all participants answer all questions). For training and inference of the Random Forest classifier, we use only questions with answers holding discrete values. From these we remove questions B7 and B8, which are only used to create the ground-truth set, as well as related questions, such as "B0: As far as you know, have you ever had coronavirus (COVID-19)?" and "B15: Do any of the following reasons describe why you were tested for COVID-19 in the past 14 days?". Finally, we do not use the questions related to vaccination, since we do not want them to influence the classification. The set of questions used can be found in Appendix B. The answers to this set of questions are "dummified" before they are used, i.e., a question with k possible answers is replaced by k binary attributes. The Random Forest model is generated with the randomForest function in R. No hyperparameter tuning is done, and the standard options of the function are used, with the exception of limiting the model to 100 trees to reduce the training time.

Observe that the questions in Appendix B include all symptoms, but also have many more questions, including behavioral or demographic aspects. Additionally, the Random Forest classifier can give different weights to different symptoms, while previously proposed symptom based criteria are based on determining only whether a symptom is present or not. Thus, overall the Random Forest classifier is much more versatile than the symptom-based criteria described in the previous section. Additionally, there are other aspects that make the Random Forest classifier(s) more adaptive:

- Firstly, we create different models for different countries. It is expected that different countries will have local characteristics, thus training and using the classifier with data from one same country can capture them.
- Secondly, we create not one but several models per country: one for each 3-month period. This allows the model to capture and adapt to aspects that change over time, like the level of vaccination, the surge of new variants, or the stringency measures imposed.

2.3 Evaluating the Classifiers

In order to verify whether the Random Forest classifier provides better proxy estimates than the symptoms-based classifiers, we selected a set of countries and tested the performance of each classifier in the last two quarters of 2021. To this end, we randomly divided the ground-truth set into a training and a testing set, with 70% and 30% of the responses of the ground-truth set in each subset, respectively. Table 1 shows the results for three countries that have detected Omicron in December for the periods of July-September 2021 (2021-Q3) and of October-December 2021 (2021-Q4). The classification performance metrics used are:

- Accuracy: Ratio of cases correctly classified over the size of the test set.
- Sensitivity / recall: Ratio of cases correctly classified as positive over the number of positive cases.
- Specificity: Ratio of cases correctly classified as negative over the number of negative cases.
- F-score: Harmonic mean of precision and recall, where the precision is the ratio of cases correctly classified as positive over the number of all cases classified as positive.

As can be seen in Table 1, Random Forest almost always shows the highest performance (marked in bold) among the classification methods used.

As another test, we then selected a set with the 20 countries that have the largest number of available responses in the UMD Global CTIS dataset along with South Africa. For each of these countries, the first two

Country	Quarter	Classifier	Accuracy	Sensitivity	Specificity	F-score
Argentina	2021-Q3	Random Forest	**0.85**	0.80	**0.86**	**0.61**
		UMD CLI	0.78	0.74	0.79	0.25
		Stringent CLI	0.82	**0.85**	0.82	0.44
		Classic CLI	0.81	0.67	0.83	0.48
		Broad CLI	0.80	0.64	0.82	0.45
Japan	2021-Q3	Random Forest	**0.95**	**0.81**	**0.96**	**0.51**
		UMD CLI	0.94	0.58	0.95	0.36
		Stringent CLI	**0.95**	0.77	0.95	0.39
		Classic CLI	0.93	0.44	**0.96**	0.42
		Broad CLI	0.91	0.29	0.95	0.29
South Africa	2021-Q3	Random Forest	**0.83**	0.81	**0.83**	**0.71**
		UMD CLI	0.71	0.70	0.72	0.34
		Stringent CLI	0.79	**0.87**	0.77	0.57
		Classic CLI	0.77	0.71	0.80	0.61
		Broad CLI	0.76	0.70	0.78	0.57
Argentina	2021-Q4	Random Forest	**0.90**	**0.71**	**0.91**	**0.51**
		UMD CLI	0.88	0.63	0.89	0.35
		Stringent CLI	0.88	0.70	0.89	0.37
		Classic CLI	0.86	0.48	**0.91**	0.44
		Broad CLI	0.86	0.47	0.90	0.42
Japan	2021-Q4	Random Forest	**0.97**	**0.69**	**0.97**	**0.31**
		UMD CLI	0.96	0.26	**0.97**	0.20
		Stringent CLI	**0.97**	0.59	**0.97**	0.30
		Classic CLI	0.94	0.18	**0.97**	0.22
		Broad CLI	0.93	0.11	**0.97**	0.14
South Africa	2021-Q4	Random Forest	**0.83**	0.69	**0.85**	**0.55**
		UMD CLI	0.79	0.63	0.81	0.35
		Stringent CLI	0.80	**0.74**	0.80	0.32
		Classic CLI	0.80	0.58	0.84	0.48
		Broad CLI	0.80	0.58	0.84	0.47

Table 1: Performance for three different countries in two different 3-month periods (2021-Q3: July-September 2021 and 2021-Q4: October-December 2021) of the different classifiers in the ground-truth set, when randomly divided into training (70%) and testing (30%) subsets.

			Pearson correlation with Confirmed				
Country	OWID TPR	CTIS TPR	Random Forest	UMD CLI	Stringent CLI	Classic CLI	Broad CLI
Argentina	**0.09**	0.17	**0.95**	**0.97**	**0.96**	**0.92**	**0.91**
Australia	**0.01**	**0.02**	**0.93**	0.46	0.31	-0.10	0.03
Brazil	–	0.19	**0.98**	0.03	0.82	0.36	0.46
Canada	**0.03**	**0.04**	**0.94**	0.85	0.66	0.73	0.71
France	**0.03**	**0.05**	**0.92**	0.69	0.80	0.57	0.61
Germany	**0.09**	**0.01**	**0.96**	0.88	**0.91**	0.82	0.81
Hungary	**0.08**	0.16	**0.93**	0.85	**0.95**	0.82	0.79
India	**0.02**	0.16	0.31	-0.38	-0.31	-0.71	-0.37
Italy	**0.02**	**0.03**	**0.98**	0.86	0.85	0.71	0.72
Japan	**0.05**	**0.04**	**0.93**	**0.90**	0.84	-0.17	0.67
Mexico	0.27	0.22	**0.97**	**0.99**	**0.98**	**0.95**	**0.98**
Poland	**0.08**	0.16	**0.96**	0.82	**0.97**	0.80	0.80
Romania	**0.07**	**0.09**	**0.94**	**0.96**	**0.98**	**0.96**	**0.95**
Russia	**0.05**	0.14	0.38	0.34	0.37	0.41	0.33
South Africa	0.16	0.24	**0.93**	**0.92**	0.84	**0.97**	**0.98**
Spain	**0.07**	**0.09**	**0.93**	0.82	0.79	0.48	0.52
Sweden	**0.06**	**0.05**	**0.91**	0.83	0.74	0.71	0.67
Thailand	0.20	**0.07**	0.85	0.83	**0.92**	0.84	0.77
Ukraine	0.20	0.16	**0.97**	0.87	**0.95**	**0.91**	0.89
United Kingdom	**0.04**	**0.06**	0.84	0.70	0.52	0.59	0.60
Vietnam	**0.06**	**0.02**	0.83	0.79	0.79	0.74	0.78

Table 2: Test-positivity rate (TPR) obtained from OWID and extracted from the UMD Global CTIS data for the 20 countries with largest survey data and South Africa. Values of at most 0.1 are shown in bold. The rest of columns show the Pearson correlation coefficient of each different proxy with the Confirmed time series. Correlation values of at least 0.9 are shown in bold. The time period used is Jun 18th, 2021 to Dec 31st, 2021. The estimates have been smoothed with a rolling average of 14 days.

columns of Table 2 show the official Test Positivity Rates obtained via *Our World In Data* [RMRG+20, Our21] (OWID TPR) and the corresponding survey-based estimate from the UMD Global CTIS dataset (CTIS TPR). The remaining columns show the Pearson correlation coefficient between the time series of Confirmed active cases (computed based on data from Johns Hopkins University [Joh20] as described by Alvarez et al. [ÁBC+21]) and that of each of the candidate proxies in the period June 18th, 2021[2] to December 31st, 2021. All time series have one value per day, which is the average of the latest 14 days.

We can make two observations from Table 2. First, Random Forest turns out to be the candidate proxy that exhibits the highest correlation values in most countries. Second, 17 out of the 21 countries exhibit low TPR (≤ 0.1) values in at least one of the first two columns (either official or survey-based TPR), and 11 out of the 21 exhibit low values in both columns, with 7 having values no higher than 0.05[3]. This suggests that such countries tend to keep the case count relatively under control and report data somewhat correctly. We can thus interpret the high correlation between the Random Forest proxy and the Confirmed time series as a sign that this proxy constitutes the most promising option among the five proxies considered.

2.4 Prevalence and Efficacy Estimation

As mentioned, each classifier will be used to determine whether survey responses correspond to positive or negative cases. Hence, the prevalence of COVID-19 estimated by a given classifier is the ratio between the number of positive cases over the total number of responses. Then, we consider four subsets of responses:

- Unvaccinated: Participants that respond negatively to the question "V1: Have you had a COVID-19 vaccination?"
- Vaccinated: Participants that respond positively to Question V1.
- Vaccinated with 1 dose: Participants that respond positively to Question V1 and declare having received 1 dose in Question "V2: How many COVID-19 vaccinations have you received?"

[2]Start of the first period considered in [VRAB21].
[3]The WHO considers countries to have the epidemic under control when their TPR is below 0.05 [W+20].

- Vaccinated with 2 doses: Participants that respond positively to Question V1 and declare having received 2 doses in Question V2.

Unfortunately, from the questions in the UMD Global CTIS it is not possible to know whether those with one dose are fully vaccinated, i.e., they have received a one-dose vaccine, or they simply received only the first dose of a two-dose vaccination. Similarly, it is not possible to know whether the participant received a booster shot.

For each of these subsets, the prevalence of COVID-19 is computed as the fraction of responses classified as positive among the responses that report a given vaccination status. For each proxy we also estimate the *vaccine efficacy* (V_E) against illness as in [VRAB21], based on the estimates of prevalence among unvaccinated (P_U) and vaccinated (P_V):

$$V_E = 1 - \frac{P_V}{P_U}.$$

The confidence intervals of this metric are obtained using the Katz-log Method [AB15]. Since we have three subsets of vaccinated participants, we compute the vaccine efficacy for the subsets Vaccinated, Vaccinated with 1 dose, and Vaccinated with 2 doses.

Date	% Delta	% Omicron	# samples
2021-06-14	45.23	0.00	1101
2021-06-28	78.09	0.00	1661
2021-07-12	88.90	0.00	2226
2021-07-26	94.30	0.00	1667
2021-08-09	95.19	0.00	1601
2021-08-23	97.58	0.00	1242
2021-09-06	97.01	0.00	1269
2021-09-20	95.77	0.00	923
2021-10-04	93.57	0.00	513
2021-10-18	93.56	0.00	450
2021-11-01	95.67	0.48	208
2021-11-15	69.30	20.18	114
2021-11-29	13.08	85.00	780
2021-12-13	0.92	95.92	980
2021-12-27	0.00	93.85	65

Table 3: Percentage of sequenced virus samples belonging to Delta and Omicron in South Africa from June 1st to December 31st of 2021. The third column presents the total number of samples reported on the corresponding date.

2.5 Time Periods of Interest

2.5.1 South Africa

The main objective of this work is to evaluate the change in vaccine efficacy due to the Omicron variant. To this end, we evaluate the decrease in vaccine efficacy in South Africa from mid-June 2021 until the end of 2021. Moreover, to ensure that we have sufficient data for our estimates, we concentrate on three time periods in 2021, each lasting about a month, two dominated by the Delta variant: i) June 18 to July 18, 2021, which is the period considered in [VRAB21], and ii) August 9 to September 6, 2021; and one dominated by Omicron: December 1st to 31st, 2021[4] (see Table 3). In addition to considering South Africa as a whole, we also study the Gauteng province, which is among the most affected by Omicron in the country.

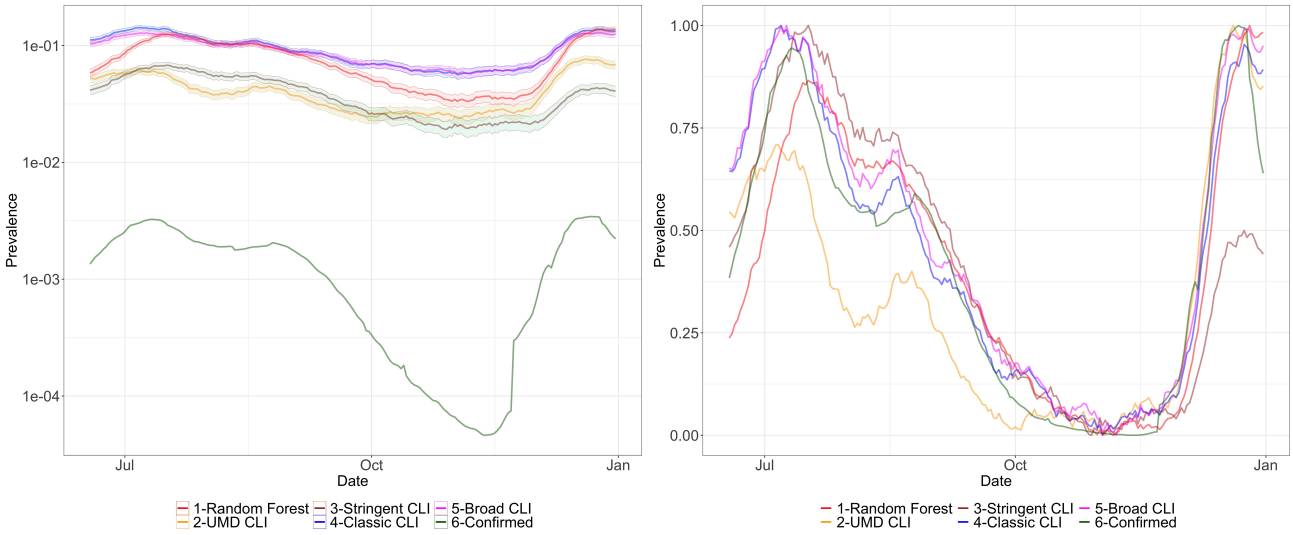
2.5.2 World

Beyond South Africa, we study the 50 countries for which the UMD Global CTIS has the largest amount of data. We compute for all of them the vaccine efficacy in two periods.

- Period 1: The month of October (in which Omicron was still not present).
- Period 2: The month of December (in which Omicron was present).

A computed efficacy value is only considered if it is non-negative, both prevalences P_V and P_U are at least 0.01, and the number of samples used to compute them is at least 1000. We only consider further the countries with at least one efficacy value in Period 2.

[4]The information on variant presence is obtained from [Our21], which extracts it from [EBM17] via [Hod21].



(a) Direct (b) Normalized

Figure 1: Prevalence in South Africa obtained with the different proxies, smoothed with a rolling average of 14 days from June 18th to December 31st, 2021. In the left plot we have the actual ratio (note that the y axis is in logarithmic scale). On the right plot all curves are normalized so the smallest value is 0 and the largest value is 1.

We have observed that the information on prevalence of Omicron is available [Our21] with a significant delay. Hence, most countries do not report relevant presence of Omicron until the second half of December 2021. For that reason, we consider the prevalence of Omicron reported in Period 3: from December 15th, 2021 to January 7th, 2022[5]. Furthermore, among the countries mentioned above, in order to have a reasonable estimate of the prevalence of the Omicron variant, we consider only countries whose data is based on sequencing at least 30 virus samples. We say that these are the countries with *presence of Omicron* and use their estimated Omicron prevalence in Period 3 in some of our results.

For all countries with presence of Omicron, we compare the estimated vaccination efficacy using Random Forest among all three vaccination groups and for both periods. For this, we adopt simple statistical methods, such as correlation analysis.

3 Results

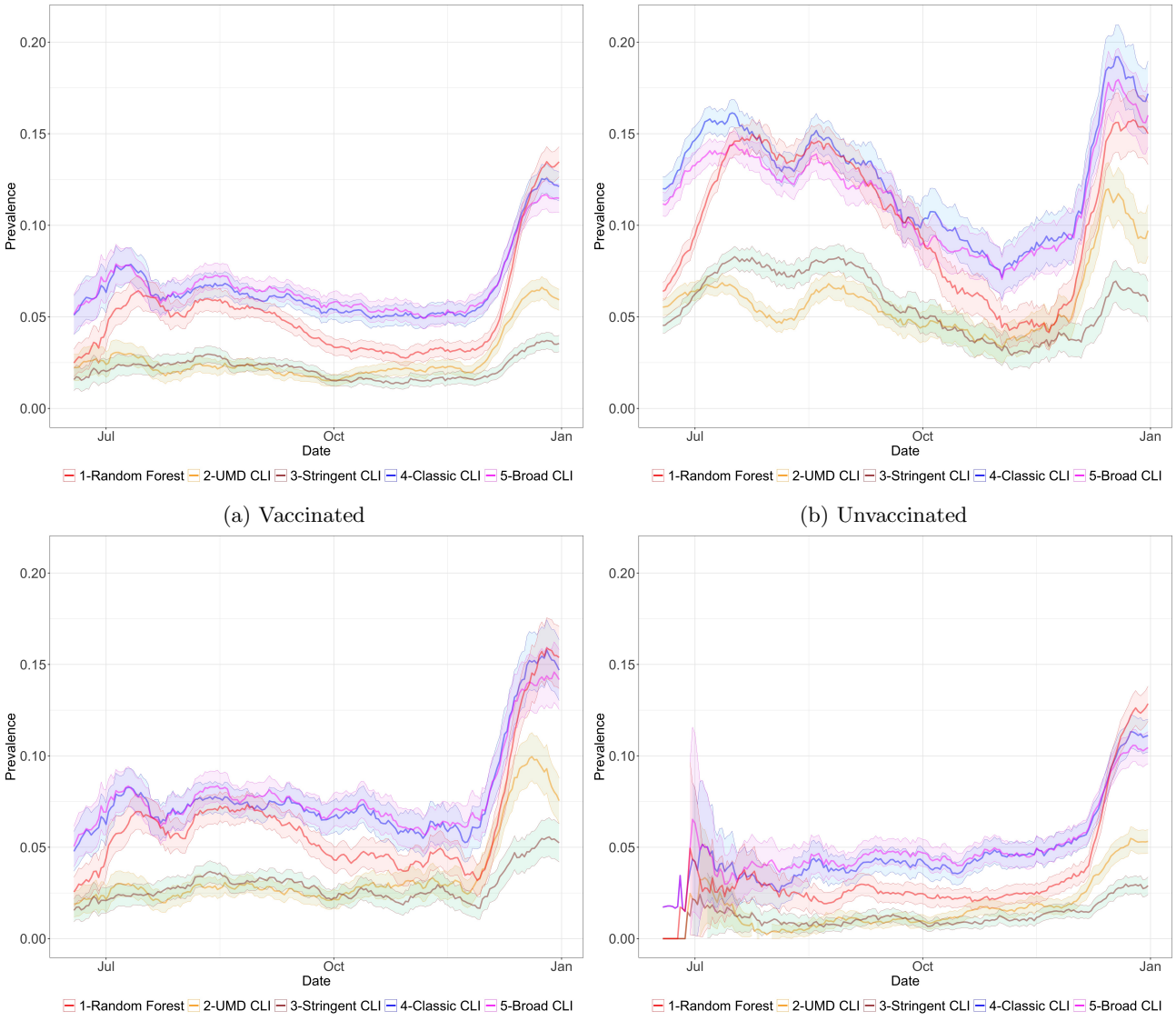
3.1 Prevalence and Vaccination Efficacy in South Africa

Figures 1a and 1b show the prevalence of COVID-19 in South Africa in the period June 18th to December 31st, 2021, with the different proxies. The direct approach of Figure 1a shows a gap from the estimate Confirmed derived from the official number of cases to the other proxies. This gap can be explained by a combination of under-detection in the official number of cases (in South Africa the test-positivity rate is above 15%, as seen in Table 2) and the presence of a background of symptoms that never goes to zero. Figure 1b shows that if each curve is independently normalized to the unit scale all proxies closely track the evolution of the official number of cases Confirmed.

In Figures 2a, 2b, 2c and 2d we show the COVID-19 prevalence in South Africa among Vaccinated, Unvaccinated, Vaccinated with 1 dose and Vaccinated with 2 doses with the diffferent proxies. We can observe that the UMD CLI and Stringent CLI proxies show a low infection prevalence in the period July-September and the month of December when compared with the Random Forest proxy. This is possibly because UMD CLI and Stringent CLI have a fixed combination of symptoms that did not capture well the new variants Delta and Omicron, while the Random Forest classifier is trained on a 3-month period and can adapt to these changes. On the other hand, Classic CLI and Broad CLI show a high prevalence in the period October-November, when the official data was showing that the number of cases was very low, possibly because of existing symptoms in the population not related to COVID-19.

Focusing on the Random Forest proxy, and in Vaccinated (2a) versus Unvaccinated (2b) prevalence, we can observe that although in the unvaccinated population we see a similar magnitude across the two waves (August-September and December) we see that in the Vaccinated group there is a much higher rate of prevalence in the

[5]Our World In Data [Our21] stopped sharing the variant data on January 10th, 2022, upon GISAID request.



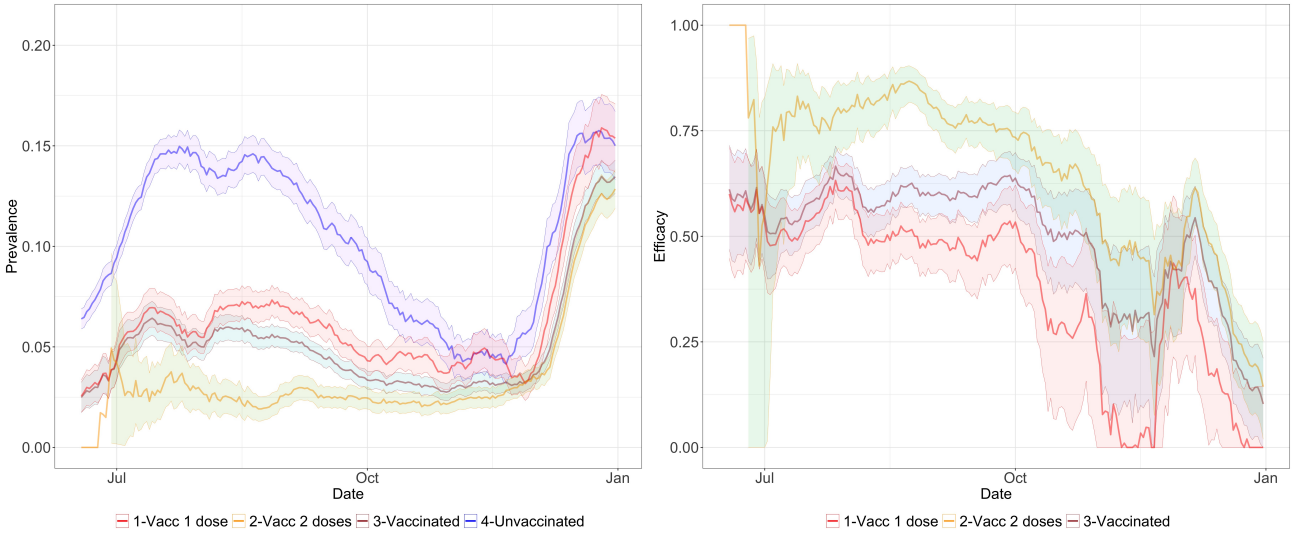
(a) Vaccinated

(b) Unvaccinated

(c) Vaccinated with 1 dose

(d) Vaccinated with 2 doses

Figure 2: Prevalence in South Africa among Vaccinated, Unvaccinated, Vaccinated with 1 dose, and Vaccinated with 2 doses, with different proxies.



(a) Prevalence

(b) Vaccination efficacy

Figure 3: (a) Prevalence and (b) vaccination efficacy in South Africa among people with different levels of vaccination, estimated with Random Forest.

Method	Jun-Jul Efficacy [95%CI]	Aug-Sep Efficacy [95%CI]	Dec Efficacy [95%CI]
	Vaccinated		
Random Forest	0.54 [0.48,0.59]	0.62 [0.58,0.65]	0.24 [0.17,0.30]
UMD CLI	0.60 [0.53,0.66]	0.66 [0.61,0.70]	0.46 [0.39,0.51]
Stringent CLI	0.69 [0.63,0.74]	0.70 [0.66,0.73]	0.48 [0.40,0.55]
Classic CLI	0.55 [0.50,0.59]	0.56 [0.52,0.59]	0.38 [0.33,0.43]
Broad CLI	0.50 [0.44,0.54]	0.49 [0.44,0.52]	0.36 [0.30,0.41]
	Vaccinated with one dose		
Random Forest	0.50 [0.44,0.56]	0.51 [0.46,0.55]	0.09 [0.00,0.18]
UMD CLI	0.61 [0.54,0.68]	0.56 [0.50,0.62]	0.21 [0.09,0.31]
Stringent CLI	0.67 [0.61,0.73]	0.60 [0.54,0.65]	0.23 [0.07,0.36]
Classic CLI	0.53 [0.47,0.57]	0.47 [0.42,0.51]	0.21 [0.13,0.28]
Broad CLI	0.46 [0.40,0.52]	0.39 [0.34,0.44]	0.18 [0.09,0.26]
	Vaccinated with two doses		
Random Forest	0.76 [0.64,0.84]	0.81 [0.78,0.84]	0.30 [0.23,0.36]
UMD CLI	0.75 [0.57,0.86]	0.85 [0.79,0.88]	0.56 [0.50,0.61]
Stringent CLI	0.82 [0.66,0.90]	0.88 [0.84,0.91]	0.59 [0.51,0.65]
Classic CLI	0.77 [0.66,0.84]	0.71 [0.67,0.75]	0.45 [0.40,0.49]
Broad CLI	0.75 [0.63,0.83]	0.66 [0.61,0.71]	0.43 [0.37,0.48]

Table 4: Vaccine efficacy in South Africa calculated for three time periods: June 18th to July 18th (Jun-Jul), August 9th to September 6th (Aug-Sep), and December 1st to 31st (Dec).

Method	Jun-Jul Efficacy [95%CI]	Aug-Sep Efficacy [95%CI]	Dec Efficacy [95%CI]
	Vaccinated		
Random Forest	0.43 [0.33,0.51]	0.62 [0.54,0.69]	0.30 [0.18,0.40]
UMD CLI	0.58 [0.44,0.68]	0.63 [0.51,0.73]	0.52 [0.41,0.61]
Stringent CLI	0.64 [0.53,0.72]	0.70 [0.61,0.78]	0.57 [0.43,0.67]
Classic CLI	0.50 [0.42,0.58]	0.51 [0.42,0.59]	0.48 [0.39,0.55]
Broad CLI	0.49 [0.39,0.57]	0.41 [0.31,0.50]	0.45 [0.35,0.53]
	Vaccinated with one dose		
Random Forest	0.40 [0.28,0.49]	0.54 [0.44,0.63]	0.14 [0.00,0.30]
UMD CLI	0.60 [0.46,0.71]	0.58 [0.42,0.70]	0.38 [0.18,0.53]
Stringent CLI	0.62 [0.49,0.71]	0.61 [0.47,0.71]	0.39 [0.13,0.57]
Classic CLI	0.47 [0.37,0.56]	0.47 [0.36,0.56]	0.35 [0.20,0.46]
Broad CLI	0.44 [0.33,0.53]	0.34 [0.20,0.45]	0.29 [0.14,0.42]
	Vaccinated with two doses		
Random Forest	0.62 [0.36,0.78]	0.77 [0.67,0.85]	0.36 [0.24,0.46]
UMD CLI	0.69 [0.27,0.87]	0.73 [0.54,0.84]	0.57 [0.45,0.66]
Stringent CLI	0.85 [0.55,0.95]	0.88 [0.76,0.94]	0.65 [0.51,0.74]
Classic CLI	0.79 [0.59,0.90]	0.58 [0.44,0.68]	0.53 [0.44,0.60]
Broad CLI	0.80 [0.59,0.91]	0.54 [0.39,0.65]	0.50 [0.41,0.58]

Table 5: Vaccine efficacy in the Gauteng province of South Africa calculated for three time periods: June 18th to July 18th (Jun-Jul), August 9th to September 6th (Aug-Sep), and December 1st to 31st (Dec).

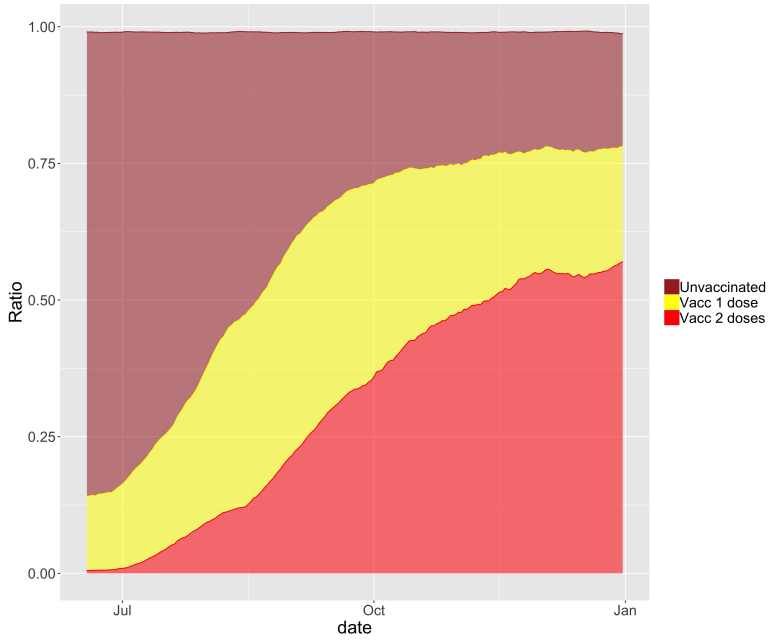


Figure 4: Evolution of the vaccination in South Africa as ratio of the population, estimated from the UMD Global CTIS data. A small fraction of responses that declared being vaccinated without reporting the number of doses are not presented for clarity. The values are from June 18th to December 31st, 2021, smoothed with a rolling average of 14 days.

	Prevalence		Vaccination efficacy	
Vaccination status	October	December	October	December
Vaccinated 2 doses	0.02 [0.01,0.02]	0.03 [0.03,0.04]	0.53 [0.49,0.58]	0.45 [0.39,0.50]
Vaccinated	0.02 [0.01,0.03]	0.04 [0.03,0.04]	0.49 [0.45,0.52]	0.43 [0.37,0.48]
Vaccinated 1 dose	0.03 [0.02,0.04]	0.05 [0.04,0.06]	0.34 [0.22,0.45]	0.32 [0.23,0.41]
Unvaccinated	0.04 [0.03,0.05]	0.06 [0.05,0.07]	–	–

Table 6: Prevalence of COVID-19 and vaccine efficacy (with 95% confidence interval) in the countries with presence of Omicron in the periods of October and December 2021.

	Correlation coefficient	P-value
Prevalence omicron vs vaccination efficacy	-0.680301	0.000354
Prevalence omicron vs vacc. efficacy 1 dose	-0.564977	0.035274
Prevalence omicron vs vacc. efficacy 2 doses	-0.628936	0.001306

Table 7: Relationship between prevalence of Omicron and vaccine efficacy in the countries with presence of Omicron.

December wave. This hints at a decrease of vaccine efficacy towards infection with the introduction of Omicron, as we will show next.

Figure 3a shows the prevalence in South Africa estimated with Random Forest across the reported vaccination states. Here we confirm the observation that in the December wave there was a disproportionate increase of infections in the vaccinated groups (Vaccinated, Vaccinated with 1 dose and Vaccinated with 2 doses). We also observe that, as expected, subjects vaccinated with two doses show higher protection that those reporting only one dose (with Vaccinated somewhere in between since it combines both groups).

As for vaccination efficacy, Figure 3b shows the estimates for South Africa, again with Random Forest. While the data in October-November has lower quality due to the reduced number of cases in that country, we can clearly observe the reduction of vaccine efficacy, towards infection, when contrasting the August-September period to the December period when Omicron dominates. Table 4 quantifies the measurements of estimated efficacy for the three periods of interest and for the five classifiers. We also provide a similar analysis in Table 5 with data restricted to the Gauteng province.

Figure 4 shows an area plot, estimated from the UMD Global CTIS data, of the proportion of vaccinated with 1 dose, Vaccinated with 2 doses, and Unvaccinated from June 18th until December 31st, 2021. As can be seen, the ratio of the population vaccinated is low at the beginning of this interval, especially with two doses. Then, we can see a high increase in Vaccinated between July and October. We point out that in each time point of this plot the proportions are provided by a different set of surveys respondents, and it still closely captures the increase of vaccination.

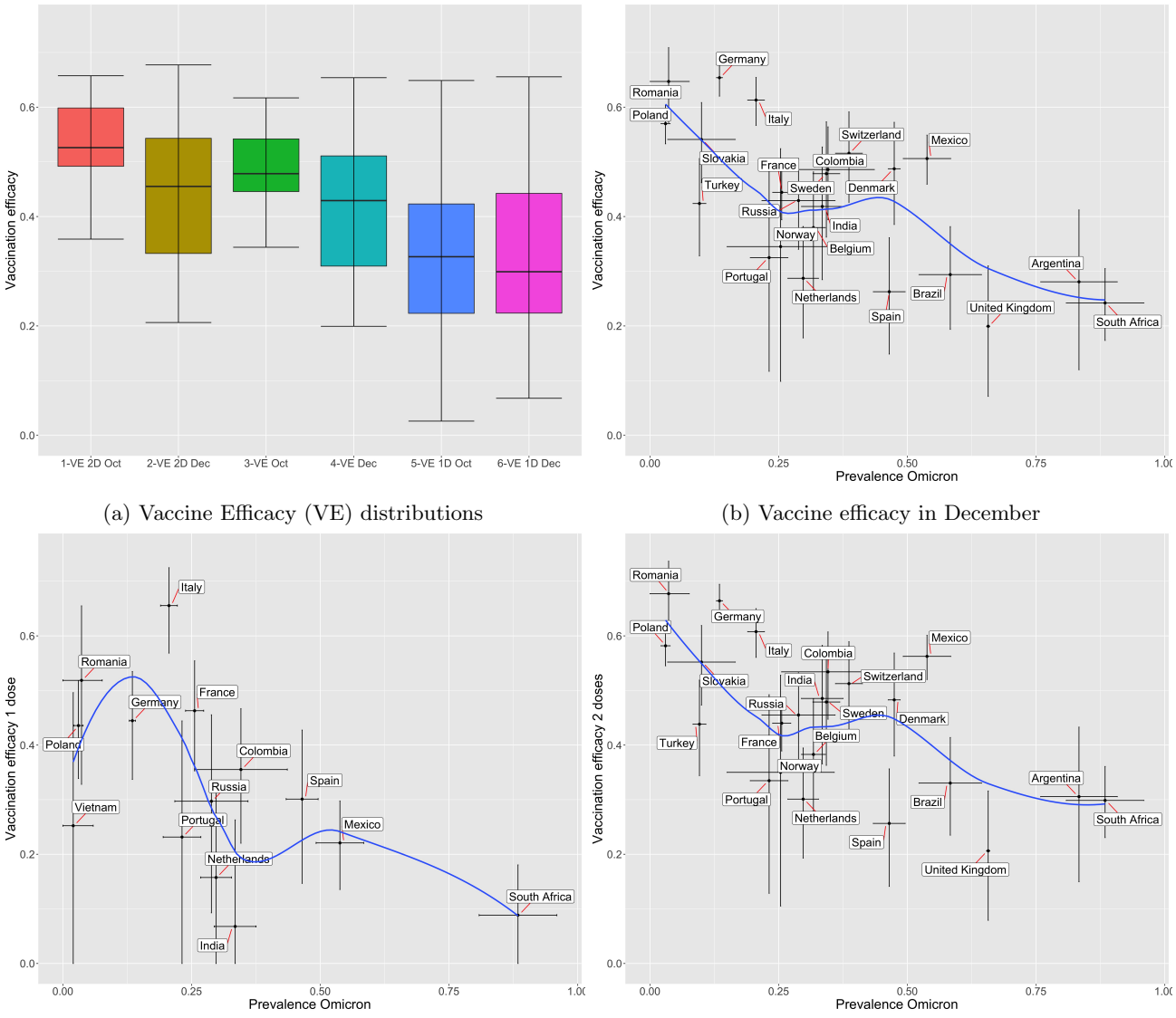
3.2 Prevalence and Vaccination Efficacy in the World

From the analysis of the 50 countries with the largest amount of data in the CTIS plus *presence of Omicron* and a calculated efficacy value, as defined in Section 2.5.2, we obtain a set of 24 countries. In Table 8 (in Appendix C) we show, for reference, the level of vaccination in these countries[6]. The next two tables, Table 9 and 10, present the estimates of virus prevalence in the same countries in the periods of October and December, and also estimates of vaccination efficacy towards infection.

Both prevalence estimates and the derived efficacy estimates are obtained by the Random Forest classifier and shown with 95% confidence intervals. When data is insufficient to meet the defined selection criteria (c.f. Section 2.5.2), it is omitted and replaced by "–". Both tables are presented alphabetically by country name and also share a column depicting the most recent data on Omicron prevalence among all virus samples. While Table 9 focuses on the data from individuals that declared their overall vaccination status (using groups Vaccinated, Unvaccinated), Table 10 makes a more detailed characterization by considering the number of doses declared (groups Vaccinated with 1 dose, Vaccinated with 2 doses, Unvaccinated). We also observe that there is less data on individuals with only one dose, since this is a transient state in the vaccination sequence. The full information on sample sizes can be consulted in Appendix C in Tables 11 and 12.

Figure 5a shows three pairs of box plots. Each pair allows comparing vaccine efficacy in October and December when considering data from the selected countries. Table 6 presents the average corresponding to each boxplot, with the 95% confidence interval. We observe that although results are inconclusive for Vaccinated with 1 dose, there is a clear decrease of overall efficacy when considering Vaccinated and Vaccinated with 2 doses.

[6]Vaccination data is obtained from [Our21, MROO+21].



(a) Vaccine Efficacy (VE) distributions

(b) Vaccine efficacy in December

(c) Efficacy 1 dose in December

(d) Efficacy 2 doses in December

Figure 5: Analysis of vaccine efficacy towards preventing infection: Sub-figure (a) shows distributions of efficacy in October and December, for the countries with presence of Omicron (as defined in Section 2.5.2); Sub-figures (b,c,d) show vaccination efficacy versus Omicron prevalence in the same set of countries, depending on vaccination status. For each country the 95% confidence intervals of the two values are shown as black lines. The blue line is the Loess curve fitting of the data.

The next three figures, Figures 5b, 5c and 5d, allow us to see a clear trend when plotting efficacy against the most recent relative level of Omicron presence in each selected country. For each case, we present a smoothed line, in blue, depicting a clear decreasing trend. Table 7 presents estimates for the correlation coefficient (using Pearson correlation) together with the corresponding p-value, which confirms its statistical significance for the usual $\alpha = 5\%$.

4 Discussion

After its surge in South Africa, the Omicron variant is increasing in prevalence in other countries. Although it is still unclear if this variant is associated to a milder disease [KBPC+21] several studies have raised concerns over the decrease of vaccine effectiveness against infection [PvSG+21, NKL+21, KST+21, LMD+21] and this can lead to a wider spread of the virus even in countries with a high vaccination uptake. While we have observed that Omicron reduces the efficacy of vaccines, new studies show that T cells may remain effective with this new variant [AQM22].

Daily participatory symptom surveillance, with widespread deployment in most world countries along the last couple of years, has the potential to offer a new instrument for assessing both global and local trends in health status. While limited in assessing the ground truth, due to the smaller control over the sample design and the need to preserve anonymity, we believe that the vast number of daily survey responses can compensate some of these factors. In this study, we developed a method to adapt and calibrate against the reported SARS-CoV-2 infection status the selection of symptoms, and other covariates from the survey, along different time periods and locations. This was shown to provide a better proxy for assessing the trend in infections and more closely track the official reported cases, in particular in those countries that had a strong surveillance and consistent test positivity rates.

Using this improved classifier we complemented earlier results [VRAB21] that used traditional fixed combinations of symptoms, and updated the analysis for South Africa showing the observed decrease in vaccine efficacy when contrasting a Delta-dominated period (August-September 2021) with the recent Omicron-dominated period (December 2021). We confirmed the presence of a measurable drop in vaccine efficacy from 0.62 (with 95% confidence interval $[0.58, 0.65]$) in the Delta period to 0.24 (95% CI $[0.17, 0.30]$) in the Omicron period in the whole country ($0.62[0.54, 0.69]$ to $0.30[0.18, 0.40]$ in the Gauteng province). In addition, we confirmed that having two doses of vaccine confers better protection than one dose, both in Delta ($0.81[0.78, 0.84]$ versus $0.51[0.46, 0.55]$) and Omicron ($0.30[0.23, 0.36]$ versus $0.09[0.00, 0.18]$) dominated periods. However, we have no data on the status of respondents with regard to a possible booster dose.

By January 7th, 2022, there were a limited number of candidate countries exhibiting both a high prevalence of Omicron and a high level of sequencing data supporting it. Nevertheless, we extend the analysis to these countries and show the observed changes in efficacy when comparing the months of October (pre-Omicron) with December (with partial presence of Omicron). Although these results should be confirmed once the level of Omicron becomes more dominant in many countries, we have observed a significant level of correlation of around and beyond -0.6 between vaccine efficacy (with either one or two doses) and the prevalence of Omicron. We must also keep clear that this reduction of efficacy is towards infection, and while it does have impact on transmission it does not imply a reduction of vaccine efficacy in protection against serious disease, hospitalization and death.

There are several assumptions that frame our analysis. We assume that UMD Global CTIS answers provide a sample of the population that is interchangeable among the Delta and Omicron dominated periods. Additionally, we did not take into account possible effects from waning immunity and vaccine boost shots, however by considering several different countries we have a mix of different vaccination timings.

References

[AB15] Ken Aho and R Terry Bowyer. Confidence intervals for ratios of proportions: implications for selection ratios. *Methods in Ecology and Evolution*, 6(2):121–132, 2015.

[ÁBC+21] Javier Álvarez, Carlos Baquero, Elisa Cabana, Jaya Prakash Champati, Antonio Fernández Anta, Davide Frey, Augusto Garcia-Agundez, Chryssis Georgiou, Mathieu Goessens, Harold Hernández, Rosa Lillo, Raquel Menezes, Raúl Moreno, Nicolas Nicolaou, Oluwasegun Ojo, Antonio Ortega, Estrella Rausell, Jesús Rufino, Efstathios Stavrakis, Govind Jeevan, and Christin Glorioso. Estimating Active Cases of COVID-19. *medRxiv*, 2021.

[AQM22] Syed Faraz Ahmed, Ahmed Abdul Quadeer, and Matthew R. McKay. SARS-CoV-2 T Cell Responses Elicited by COVID-19 Vaccines or Infection Are Expected to Remain Robust against Omicron. *Viruses*, 14(1), 2022.

[ASK+21] Nick Andrews, Julia Stowe, Freja Kirsebom, Samuel Toffa, Tim Rickeard, Eileen Gallagher, Charlotte Gower, Meaghan Kall, Natalie Groves, Anne-Marie O'Connell, et al. Effectiveness of covid-19 vaccines against the omicron (b. 1.1. 529) variant of concern. *medRxiv*, 2021.

[ATMC+21] Christina M Astley, Gaurav Tuli, Kimberly A Mc Cord, Emily L Cohn, Benjamin Rader, Tanner J Varrelman, Samantha L Chiu, Xiaoyi Deng, Kathleen Stewart, Tamer H Farag, et al. Global monitoring of the impact of the COVID-19 pandemic through online surveys sampled from the Facebook user base. *Proceedings of the National Academy of Sciences*, 118(51), 2021.

[Del20] Delphi Group, CMU. COVIDcast. https://delphi.cmu.edu/covidcast/, 2020. Accessed: 2021-06-02.

[EBM17] Stefan Elbe and Gemma Buckland-Merrett. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global challenges*, 1(1):33–46, 2017.

[Fac20] Facebook Data for Good. COVID-19 symptom survey – request for data access. https://dataforgood.fb.com/docs/covid-19-symptom-survey-request-for-data-access/, 2020. Accessed: 2021-01-24.

[FLS+20] Junchuan Fan, Yao Li, Kathleen Stewart, Anil R. Kommareddy, Adrianne Bradford, Samantha Chiu, Frauke Kreuter, Neta Barkay, Alyssa Bilinski, Brian Kim, Roee Eliat, Tal Galili, Daniel Haimovich, Sarah LaRocca, Stanley Presser, Katherine Morris, Joshua A Salomon, Elizabeth A. Stuart, Ryan Tibshirani, Tali Alterman Barash, Curtiss Cobb, Andres Garcia, Andi Gros, Ahmed Isa, Alex Kaess, Faisal Karim, Ofir Eretz Kedosha, Shelly Matskel, Roee Melamed, Amey Patankar, Irit Rutenberg, Tal Salmona, and David Vannette. Covid-19 world symptom survey data api. https://covidmap.umd.edu/api.html, 2020.

[Hod21] Emma B. Hodcroft. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. https://covariants.org/, 2021. Accessed: 2022-01-10.

[Joh20] Johns Hopkins University & Medicine. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu, 2020. Accessed: 2021-06-02.

[KBB+20] Frauke Kreuter, Neta Barkay, Alyssa Bilinski, Adrianne Bradford, Samantha Chiu, Roee Eliat, Junchuan Fan, Tal Galili, Daniel Haimovich, Brian Kim, et al. Partnering with a global platform to inform research and public policy making. *Survey Research Methods*, 14(2):159–163, 2020.

[KBPC+21] Matt J. Keeling, Ellen Brooks-Pollock, Rob Challen, Leon Danon, Louise Dyson, Julia R. Gog, Laura Guzmán Rincón, Edward M. Hill, Lorenzo Pellis, Jonathan M. Read, and Michael J. Tildesley. Short-term projections based on early omicron variant dynamics in england. *medRxiv*, 2021.

[KK21] Salim S Abdool Karim and Quarraisha Abdool Karim. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *The Lancet*, 398(10317):2126–2128, 2021.

[KST+21] David S Khoury, Megan Steain, James Triccas, Alex Sigal, Miles Philip Davenport, and Deborah Cromer. Analysis: A meta-analysis of early results to predict vaccine efficacy against omicron. *medRxiv*, 2021.

[LMD+21] Frederik Plesner Lyngse, Laust Hvas Mortensen, Matthew J. Denwood, Lasse Engbo Christiansen, Camilla Holten Møller, Robert Leo Skov, Katja Spiess, Anders Fomsgaard, Ria Lassauniere, Morten Rasmussen, Marc Stegger, Claus Nielsen, Raphael Niklaus Sieber, Arieh Sierra Cohen, Frederik Trier Møller, Maria Overvad, Kåre Mølbak, Tyra Grove Krause, and Carsten Thure Kirkeby. Sars-cov-2 omicron voc transmission in danish households. *medRxiv*, 2021.

[LRSC+21] Epke A Le Rutte, Andrew J Shattock, Nakul Chitnis, Sherrie L Kelly, and Melissa A Penny. Assessing impact of omicron on sars-cov-2 dynamics and public health burden. *medRxiv*, 2021.

[MROO+21] Edouard Mathieu, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao. A global database of COVID-19 vaccinations. *Nature human behaviour*, pages 1–7, 2021.

[NKL+21] Ital Nemet, Limor Kliker, Yaniv Lustig, Neta S Zuckerman, Oran Erster, Carmit Cohen, Yitshak Kreiss, Sharon Alroy-Preis, Gili Regev-Yochay, Ella Mendelson, et al. Third bnt162b2 vaccination neutralization of sars-cov-2 omicron infection. *medRxiv*, 2021.

[Oo21] World Health Organization and 26 November 2021 others. Classification of omicron (b.1.1.529): Sars-cov-2 variant of concern. https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern, 2021.

[Our21] Our World in Data. Data on COVID-19 (coronavirus) by Our World in Data. https://covid.ourworldindata.org/, 2021. Accessed: 2022-01-07.

[PvSG+21] Juliet RC Pulliam, Cari van Schalkwyk, Nevashan Govender, Anne von Gottberg, Cheryl Cohen, Michelle J Groome, Jonathan Dushoff, Koleka Mlisana, and Harry Moultrie. Increased risk of sars-cov-2 reinfection associated with emergence of the omicron variant in south africa. *MedRxiv*, 2021.

[RHRM+21] Victoria Rotshild, Bruria Hirsh-Raccah, Ian Miskin, Mordechai Muszkat, and Ilan Matok. Comparing the clinical efficacy of covid-19 vaccines: a systematic review and network meta-analysis. *Scientific Reports*, 11, 2021.

[RMRG+20] Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, and Max Roser. Coronavirus Pandemic (COVID-19). *Our World in Data*, 2020. https://ourworldindata.org/coronavirus.

[SLD+22] Andrew J. Shattock, Epke A. Le Rutte, Robert P. Dünner, Swapnoleena Sen, Sherrie L. Kelly, Nakul Chitnis, and Melissa A. Penny. Impact of vaccination and non-pharmaceutical interventions on sars-cov-2 dynamics in switzerland. *Epidemics*, 38:100535, 2022.

[SRB+21] Joshua A Salomon, Alex Reinhart, Alyssa Bilinski, Eu Jing Chua, Wichada La Motte-Kerr, Minttu M Rönn, Marissa B Reitsma, Katherine A Morris, Sarah LaRocca, Tamer H Farag, et al. The us covid-19 trends and impact survey: Continuous real-time measurement of covid-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proceedings of the National Academy of Sciences*, 118(51), 2021.

[The21a] The University of Maryland Social Data Science Center. COVID19_symptom_survey_intl_V11_noneu. https://covidmap.umd.edu/document/COVID19_symptom_survey_intl_V11_0723.pdf, 2021. Accessed: 2022-01-10.

[The21b] The University of Maryland Social Data Science Center. The University of Maryland Social Data Science Center Global COVID-19 Trends and Impact Survey in partnership with Facebook. https://covidmap.umd.edu/, 2021. Accessed: 2022-01-10.

[VRAB21] Tanner J Varrelman, Benjamin M Rader, Christina M Astley, and John S Brownstein. Syndromic surveillance-based estimates of vaccine efficacy against covid-like illness from emerging omicron and covid-19 variants. *medRxiv*, 2021.

[W+20] World Health Organization et al. Public health criteria to adjust public health and social measures in the context of covid-19: annex to considerations in adjusting public health and social measures in the context of covid-19, 12 may 2020. Technical report, World Health Organization, 2020.

A List of Symptoms

In the UMD Global CTIS the following question is asked: "B1 In the last 24 hours, have you had any of the following?" [The21a]. The following is the list of possible answers (non exclusive):

- Fever (B1_1).
- Cough (B1_2).
- Difficulty breathing (B1_3).
- Fatigue (B1_4).
- Stuffy or runny nose (B1_5).
- Aches or muscle pain (B1_6).
- Sore throat (B1_7).
- Chest pain (B1_8).

- Nausea (B1_9).
- Loss of smell or taste (B1_10).
- Headache (B1_12).
- Chills (B1_13).

B Questions Used for the Machine Learning Model

The following is the list of survey questions whose answers are used to create the Random Forest models, and to classify with them the responses: B1_1, B1_2, B1_3, B1_4, B1_5, B1_6, B1_7, B1_8, B1_9, B1_10, B1_11, B1_12, B1_13, B1_14, B1b_x1, B1b_x2, B1b_x3, B1b_x4, B1b_x5, B1b_x6, B1b_x7, B1b_x8, B1b_x9, B1b_x10, B1b_x11, B1b_x12, B1b_x13, B1b_x14, B3, B5, B6, B9, B10, B11, B12_1, B12_2, B12_3, B12_4, B12_5, B12_6, B13_1, B13_2, B13_3, B13_4, B13_5, B13_6, B13_7, B14_1, B14_2, B14_3, B14_4, B14_5, C0_1, C0_2, C0_3, C0_4, C0_5, C0_6, C1_m, C2, C3, C5, C6, C7, C8, C9, C9a, C12, C13_1, C13_2, C13_3, C13_4, C13_5, C13_6, C14, D1, D2, D3, D4, D5, D6_1, D6_2, D6_3, D7, D8, D9, D10, E2, E3, E4, E7, H1, H2, H3.

The questions removed are B0, B7, B8, B15, and all the questions related to vaccination (V-questions).

C Countries with Omicron Prevalence

Table 8 shows basic official vaccination data on December 31st, 2021, of these countries. Tables 9 and 10 show the COVID-19 prevalence and the vaccine efficacy in October and December in the countries with presence of Omicron as defined in Section 2.5.2.

Country	% doses/pop	% pop vacc	% pop fully vacc	% pop booster	Vacc start date
Argentina	167.98	83.76	71.61	12.22	2020-12-29
Belgium	186.28	76.65	75.70	37.59	2020-12-28
Brazil	154.81	77.66	67.03	12.42	2021-01-17
Colombia	126.19	74.81	55.25	6.49	2021-02-17
Denmark	208.57	82.65	78.43	48.30	2021-02-05
France	183.78	78.61	73.48	33.28	2020-12-27
Germany	178.84	73.62	70.61	38.87	2020-12-27
India	103.98	60.69	43.29	0.00	2021-01-16
Italy	184.28	80.14	74.11	32.52	2020-12-27
Mexico	114.24	62.89	55.87	0.00	2020-12-24
Netherlands	162.18	77.54	71.18	18.50	2021-01-09
Norway	178.68	78.41	71.76	28.52	2020-12-08
Poland	124.32	57.34	55.68	18.16	2020-12-28
Portugal	190.72	91.47	89.53	29.44	2020-12-27
Romania	82.86	28.64	40.87	0.00	2020-12-27
Russia	100.31	50.60	45.76	5.06	2020-12-15
Slovakia	111.09	50.13	47.61	16.33	2021-01-11
South Africa	46.47	31.49	26.37	0.00	2021-02-18
Spain	178.69	84.85	81.01	29.40	2021-01-04
Sweden	172.96	76.14	72.68	0.00	2021-01-03
Switzerland	158.90	68.56	66.88	24.99	2020-12-21
Turkey	154.80	66.92	60.68	27.19	2021-01-14
United Kingdom	195.45	75.93	69.54	49.98	2021-01-10
Vietnam	153.75	79.00	69.71	0.00	2021-03-08

Table 8: Information about vaccination on December 31st, 2021, in the countries with presence of Omicron (as defined in Section 2.5.2).

D Ethical Declaration

The Ethics Board of IMDEA Networks Institute gave ethical approval for this work on 2021/07/05. IMDEA Networks has signed Data Use Agreements with Facebook, Carnegie Mellon University (CMU) and the Univer-

Country	% Prevalence Omicron	Prevalence Oct	Prevalence Dec	Vac efficacy Oct	Vac efficacy Dec
Argentina	0.83 [0.76,0.91]	0.02 [0.01,0.02]	0.03 [0.03,0.03]	0.48 [0.35,0.58]	0.28 [0.12,0.41]
Belgium	0.32 [0.29,0.34]	0.02 [0.02,0.02]	0.05 [0.05,0.05]	0.53 [0.39,0.64]	0.38 [0.26,0.48]
Brazil	0.58 [0.52,0.64]	0.03 [0.03,0.03]	0.03 [0.02,0.03]	0.43 [0.37,0.49]	0.29 [0.19,0.38]
Colombia	0.35 [0.26,0.44]	0.03 [0.03,0.03]	0.03 [0.03,0.03]	0.55 [0.49,0.61]	0.49 [0.39,0.56]
Denmark	0.47 [0.46,0.49]	0.01 [0.01,0.01]	0.05 [0.05,0.05]	–	0.49 [0.39,0.57]
France	0.26 [0.24,0.27]	0.01 [0.01,0.01]	0.03 [0.03,0.03]	–	0.44 [0.39,0.49]
Germany	0.13 [0.13,0.14]	0.01 [0.01,0.01]	0.02 [0.02,0.02]	–	0.65 [0.62,0.68]
India	0.33 [0.29,0.38]	0.04 [0.04,0.04]	0.03 [0.03,0.03]	0.44 [0.35,0.52]	0.42 [0.28,0.53]
Italy	0.21 [0.19,0.22]	0.01 [0.01,0.01]	0.02 [0.02,0.02]	–	0.61 [0.57,0.65]
Mexico	0.54 [0.49,0.58]	0.05 [0.05,0.05]	0.04 [0.04,0.04]	0.57 [0.54,0.59]	0.51 [0.46,0.55]
Netherlands	0.30 [0.27,0.33]	0.02 [0.02,0.02]	0.05 [0.04,0.05]	0.36 [0.20,0.49]	0.29 [0.18,0.38]
Norway	0.25 [0.15,0.36]	0.01 [0.01,0.01]	0.03 [0.02,0.03]	–	0.35 [0.10,0.52]
Poland	0.03 [0.02,0.04]	0.03 [0.03,0.04]	0.07 [0.06,0.07]	0.50 [0.42,0.56]	0.57 [0.53,0.60]
Portugal	0.23 [0.19,0.27]	0.01 [0.01,0.01]	0.03 [0.03,0.03]	–	0.32 [0.12,0.48]
Romania	0.04 [0.00,0.08]	0.06 [0.06,0.06]	0.02 [0.02,0.02]	0.59 [0.56,0.62]	0.65 [0.57,0.71]
Russia	0.29 [0.22,0.36]	0.04 [0.04,0.05]	0.03 [0.02,0.03]	0.45 [0.39,0.50]	0.43 [0.34,0.51]
Slovakia	0.10 [0.03,0.17]	0.03 [0.03,0.03]	0.06 [0.05,0.06]	0.47 [0.32,0.59]	0.54 [0.46,0.61]
South Africa	0.88 [0.81,0.96]	0.04 [0.04,0.04]	0.12 [0.12,0.13]	0.50 [0.41,0.57]	0.24 [0.17,0.30]
Spain	0.46 [0.43,0.50]	0.01 [0.01,0.02]	0.05 [0.05,0.06]	0.62 [0.50,0.70]	0.26 [0.15,0.36]
Sweden	0.34 [0.32,0.37]	0.01 [0.00,0.01]	0.02 [0.02,0.02]	–	0.48 [0.36,0.57]
Switzerland	0.39 [0.36,0.41]	0.01 [0.01,0.01]	0.04 [0.04,0.04]	–	0.52 [0.43,0.59]
Turkey	0.10 [0.08,0.11]	0.05 [0.05,0.06]	0.05 [0.05,0.05]	0.45 [0.38,0.51]	0.42 [0.33,0.51]
United Kingdom	0.66 [0.65,0.66]	0.03 [0.03,0.03]	0.05 [0.04,0.05]	0.34 [0.22,0.45]	0.20 [0.07,0.31]
Vietnam	0.02 [0.00,0.06]	0.01 [0.01,0.01]	0.03 [0.03,0.03]	–	–

Table 9: Prevalence of Omicron in COVID-19 and vaccination efficacy in the countries with presence of Omicron (as defined in Section 2.5.2).

sity of Maryland (UMD) to access their data, specifically UMD project 1587016-3 entitled C-SPEC: Symptom Survey: COVID-19 and CMU project STUDY2020_00000162 entitled ILI Community-Surveillance Study.

E Data Availability

The data presented in this paper and some of the programs used to process it are openly accessible at `https://github.com/GCGImdea/coronasurveys/tree/master/papers/omicron_efficacy_paper_medRxiv`.

Country	% Prevalence Omicron	Vac 1 dose efficacy Oct	Vac 1 dose efficacy Dec	Vac 2 doses efficacy Oct	Vac 2 doses efficacy Dec
Argentina	0.83 [0.76,0.91]	0.03 [0.00,0.27]	–	0.53 [0.41,0.62]	0.31 [0.15,0.43]
Belgium	0.32 [0.29,0.34]	–	–	0.55 [0.41,0.65]	0.38 [0.26,0.48]
Brazil	0.58 [0.52,0.64]	0.20 [0.11,0.28]	–	0.50 [0.44,0.55]	0.33 [0.23,0.41]
Colombia	0.35 [0.26,0.44]	0.44 [0.35,0.53]	0.36 [0.22,0.47]	0.61 [0.55,0.67]	0.53 [0.45,0.61]
Denmark	0.47 [0.46,0.49]	–	–	–	0.48 [0.38,0.57]
France	0.26 [0.24,0.27]	–	0.46 [0.35,0.55]	–	0.44 [0.39,0.49]
Germany	0.13 [0.13,0.14]	–	0.44 [0.34,0.53]	–	0.66 [0.63,0.69]
India	0.33 [0.29,0.38]	0.19 [0.05,0.31]	0.07 [0.00,0.26]	0.54 [0.47,0.61]	0.49 [0.37,0.58]
Italy	0.21 [0.19,0.22]	–	0.66 [0.57,0.72]	–	0.61 [0.56,0.65]
Mexico	0.54 [0.49,0.58]	0.36 [0.32,0.40]	0.22 [0.14,0.30]	0.66 [0.63,0.68]	0.56 [0.52,0.60]
Netherlands	0.30 [0.27,0.33]	–	0.16 [0.00,0.33]	0.41 [0.26,0.53]	0.30 [0.19,0.39]
Norway	0.25 [0.15,0.36]	–	–	–	0.35 [0.11,0.53]
Poland	0.03 [0.02,0.04]	0.31 [0.13,0.45]	0.44 [0.34,0.52]	0.52 [0.45,0.58]	0.58 [0.55,0.62]
Portugal	0.23 [0.19,0.27]	–	0.23 [0.00,0.44]	–	0.33 [0.13,0.49]
Romania	0.04 [0.00,0.08]	0.65 [0.59,0.70]	0.52 [0.33,0.65]	0.58 [0.55,0.61]	0.68 [0.60,0.74]
Russia	0.29 [0.22,0.36]	0.55 [0.43,0.64]	0.30 [0.09,0.46]	0.44 [0.38,0.50]	0.46 [0.37,0.53]
Slovakia	0.10 [0.03,0.17]	–	–	0.50 [0.35,0.61]	0.55 [0.47,0.62]
South Africa	0.88 [0.81,0.96]	0.29 [0.15,0.40]	0.09 [0.00,0.18]	0.64 [0.56,0.70]	0.30 [0.23,0.36]
Spain	0.46 [0.43,0.50]	0.34 [0.09,0.52]	0.30 [0.15,0.43]	0.66 [0.55,0.74]	0.26 [0.14,0.36]
Sweden	0.34 [0.32,0.37]	–	–	–	0.48 [0.36,0.57]
Switzerland	0.39 [0.36,0.41]	–	–	–	0.51 [0.42,0.59]
Turkey	0.10 [0.08,0.11]	–	–	0.49 [0.42,0.55]	0.44 [0.34,0.52]
United Kingdom	0.66 [0.65,0.66]	–	–	0.36 [0.24,0.46]	0.21 [0.08,0.32]
Vietnam	0.02 [0.00,0.06]	–	0.25 [0.00,0.50]	–	–

Table 10: Prevalence of Omicron and vaccination efficacy with one and two doses in the countries with presence of Omicron (as defined in Section 2.5.2). The prevalence of Omicron is replicated from Table 9 for easy reference.

Country	Total Oct	Total Dec	Unvac Oct	Unvac Dec	Vac Oct	Vac Dec	Vac 1D Oct	Vac 1D Dec	Vac 2D Oct	Vac 2D Dec
Argentina	44509	48807	3077	2778	40276	44590	3704	1884	36115	41783
Belgium	16448	18373	1687	1718	14266	16004	747	463	13327	15269
Brazil	198423	162402	9428	6552	183859	151114	38885	8680	142594	139517
Colombia	34859	33883	5437	2734	28457	30197	9979	7514	18034	22137
Denmark	19591	27284	917	1206	18279	25472	212	217	17781	24684
France	82767	111041	10234	11593	67393	95663	6369	4708	60218	89139
Germany	89348	110359	12601	11868	71980	95530	6655	5490	64611	88548
India	76675	68155	4076	2631	63803	60076	16798	7344	45967	51622
Italy	98712	112754	7023	6095	89120	103305	9066	5108	78852	96124
Mexico	139967	118861	12063	6472	119471	109330	35960	17776	82321	90162
Netherlands	27505	30803	3804	3380	23001	26621	2175	2025	20397	24087
Norway	16746	21862	935	1010	15536	20404	389	304	14980	19724
Poland	30295	38001	5318	6105	23924	30578	2327	2499	21236	27603
Portugal	22758	29352	1299	1368	21017	27340	3470	3172	17180	23631
Romania	45123	24638	11038	4917	32558	19022	4477	2451	27594	16192
Russia	35186	30037	12301	9001	21680	19884	2845	2819	18573	16779
Slovakia	9567	11323	1987	2208	7382	8841	306	487	6989	8215
South Africa	18308	19492	4149	4006	12805	14753	5009	4138	7624	10423
Spain	33455	51568	2035	2625	30652	47444	3814	3574	26453	43223
Sweden	53564	57823	3001	3200	49564	53544	699	443	48380	52348
Switzerland	14863	16755	2906	2617	11585	13742	886	676	10541	12824
Turkey	27159	22854	3238	2307	23033	19844	1473	729	21015	18561
United Kingdom	41812	47072	3080	3174	37421	42421	925	770	36109	41122
Vietnam	48955	39105	8043	1116	37073	36097	17325	3241	19233	32246

Table 11: Number of survey responses used in each period from the countries with presence of Omicron (as defined in Section 2.5.2), for each level of vaccination.

Country	Pos Oct	Pos Dec	Unvac Oct	Unvac Dec	Vac Oct	Vac Dec	Vac 1D Oct	Vac 1D Dec	Vac 2D Oct	Vac 2D Dec
Argentina	715	1302	87	99	594	1143	102	90	484	1034
Belgium	364	912	69	130	274	751	25	31	248	713
Brazil	5111	4066	405	224	4486	3648	1334	355	3072	3194
Colombia	1013	1103	285	158	666	897	291	280	364	596
Denmark	232	1405	24	116	196	1256	5	16	186	1228
France	703	3452	149	596	486	2733	102	130	377	2566
Germany	619	2253	155	580	428	1616	52	149	373	1453
India	2899	2231	186	93	1629	1235	623	242	958	939
Italy	558	2610	120	329	394	2158	67	95	322	2035
Mexico	6881	4747	1201	485	5167	4047	2287	1038	2808	2956
Netherlands	487	1441	95	210	367	1179	60	106	299	1046
Norway	147	569	15	39	127	516	10	17	116	495
Poland	1039	2504	298	749	676	1614	90	173	572	1416
Portugal	170	821	17	55	142	742	28	98	112	632
Romania	2579	448	1109	175	1335	239	158	42	1158	186
Russia	1550	775	752	318	727	401	79	70	633	323
Slovakia	276	635	89	216	174	397	14	36	157	360
South Africa	695	2348	249	599	388	1672	214	564	167	1093
Spain	468	2776	65	186	375	2479	80	177	290	2277
Sweden	297	1037	48	103	234	899	8	16	225	878
Switzerland	170	639	61	175	102	445	10	21	90	418
Turkey	1479	1143	288	181	1125	897	136	57	962	818
United Kingdom	1321	2168	141	180	1124	1926	53	59	1060	1851
Vietnam	364	1271	58	35	251	1141	95	76	152	1043

Table 12: Number of survey responses classified as positive by Random Forest in each period from the countries with presence of Omicron (as defined in Section 2.5.2), for each level of vaccination.