Estimating seroprevalence of SARS-CoV-2 antibodies using three self-reported symptoms: development of a prediction model based on data from Ischgl, Austria

Jens Lehmann[1] , Johannes M. Giesinger[1], Gerhard Rumpold[2],
Wegene Borena[3], Ludwig Knabl[3], Barbara Falkensammer[3], Cornelia Ower[4],
Magdalena Sacher[5], Dorothee von Laer[3], Barbara Sperner-Unterweger[1]
and Bernhard Holzner[1]

[1]University Hospital of Psychiatry II, Medical University of Innsbruck, Anichstrasse 35, 6020, Innsbruck, Austria; [2]Department of Psychiatry and Psychotherapy, Medical University of Innsbruck, Innsbruck, Austria; [3]Institute of Virology, Medical University of Innsbruck, Innsbruck, Austria; [4]Department of Surgery, University Hospital of Trauma Surgery, Medical University of Innsbruck, Innsbruck, Austria and [5]Department of Visceral, Transplant and Thoracic Surgery, Center of Operative Medicine, Medical University of Innsbruck, Innsbruck, Austria

Abstract

We report the development of a regression model to predict the prevalence of severe acute respiratory syndrome coronavirus (SARS-CoV-2) antibodies on a population level based on self-reported symptoms. We assessed participant-reported symptoms in the past 12 weeks, as well as the presence of SARS-CoV-2 antibodies during a study conducted in April 2020 in Ischgl, Austria. We conducted multivariate binary logistic regression to predict seroprevalence in the sample. Participants ($n = 451$) were on average 47.4 years old (s.d. 16.8) and 52.5% female. SARS-CoV-2 antibodies were found in $n = 197$ (43.7%) participants. In the multivariate analysis, three significant predictors were included and the odds ratios (OR) for the most predictive categories were cough (OR 3.34, CI 1.70–6.58), gustatory/olfactory alterations (OR 13.78, CI 5.90–32.17) and limb pain (OR 2.55, CI 1.20–6.50). The area under the receiver operating characteristic curve was 0.773 (95% CI 0.727–0.820). Our regression model may be used to estimate the seroprevalence on a population level and a web application is being developed to facilitate the use of the model.

Background

The coronavirus disease 2019 (COVID-19) is a respiratory illness that emerged in December 2019 in China and has since become a global health threat [1]. The disease is caused by an infection with the severe acute respiratory syndrome coronavirus (SARS-CoV-2) [2] and is known to spread in clusters; therefore, it is essential to swiftly identify the regions in which new infections arise to stop the spread of the pandemic [3]. However, due to limited availability, SARS-CoV-2 testing is mostly performed in those who exhibit symptoms, which can bias the true prevalence rates. As another approach to determine the prevalence of SARS-CoV-2 in the general population, the constant monitoring of subjectively reported symptoms has been proposed [3]. This has been realised using smartphone apps in which users can log in and report their health, as well as a confirmed positive polymerase chain reaction (PCR) test [3].

Such self-reported symptoms may be used to quickly and economically determine an estimate of the prevalence of the disease in the population. Moreover, they may be used to guide and inform the priority of tests that are carried out, in a way that people more likely to be infected can be tested first. For a reliable estimation, two factors are important: (1) sufficient sensitivity (i.e. correct identification of positive cases) of the symptoms to correctly predict SARS-CoV-2, and (2) sufficient specificity of the symptoms to distinguish the symptoms assessed from other diseases such as the common cold or influenza.

While there are studies that use self-reported symptoms to predict a current infection with SARS-CoV-2, there are no studies that predict the presence of SARS-CoV-2 antibodies, which may be used to judge the past burden of disease in a population. In this study, we used data from a seroprevalence study in the town of Ischgl in Austria to develop a model for estimating SARS-CoV-2 antibody seroprevalence based on self-reported symptoms. The city of Ischgl was particularly impacted by the pandemic, as it is a popular ski and a party hotspot with the high number of international tourists and was hit amidst the holiday

season. Moreover, despite earlier reports of COVID-19 in Austria, a lockdown was imposed only on the 13 March.

Methods

Procedure

Data collection was carried out between 21 and 27 April. All residents in Ischgl were invited via a letter to participate in the study and visit the study centre (a recreational centre in Ischgl). Participants who were not able to visit the study centre (e.g. due to quarantine) were visited at home by their general practitioner.

Participants were briefed about the study and asked to provide informed consent. A physician collected an 8 ml blood sample (EDTA) via venepuncture. Blood samples were screened for anti-SARS-CoV-2-S1-protein IgA and IgG antibodies (Euroimmun, Lübeck, Germany) as well as for anti-SARS-CoV-2-N-protein IgG (Abbott, Illinois, USA). A participant was considered seropositive if both assays (Euroimmun and Abbott) were positive. Discrepant results were tested by neutralisation assay. In addition, 110 randomly selected Euroimmun and Abbott positive samples were also tested for neutralising antibodies. Samples negative in both assays were considered negative. The full description of the laboratory testing procedure is to be published elsewhere.

Participants were handed a questionnaire on age, sex, history of confirmed infection with SARS-CoV-2 and history of clinical symptoms compatible with COVID-19 (since the 1st of February). Symptoms assessed were cough, fever, trouble breathing, sore throat, gustatory or olfactory alterations, loss of appetite, nausea, vomiting, constipation, diarrhoea, headache, fatigue, limb pain, dyspnoea and vertigo. Symptoms were rated on a four-point scale: 1 ('not at all'), 2 ('a little'), 3 ('quite a bit') and 4 ('very much'). Participants were asked to complete the questionnaire at home and send it to the Medical University of Innsbruck using a pre-stamped envelope. Participants were included in the analysis if they provided a blood sample, returned the completed questionnaire and were above 18 years of age. The study was approved by the ethics committee of the Medical University of Innsbruck (app. no. 1100/2020) and all methods were performed in accordance with the relevant guidelines and regulations. The data and materials used in this study are available online at https://osf.io/32mab/.

Statistical analysis

Univariate binary logistic regression was used to ascertain the effects of self-reported symptoms on the probability of participants having positive SARS-CoV-2 antibodies. Symptoms were entered into the regression on the four-point scale as presented to participants. In multivariable regression analysis, we used forward selection and backward elimination with likelihood ratio (LR) to determine significant ($P < 0.05$) predictors. In the forward selection process, new variables are added one by one, until adding new variables does not improve the model. In the backward selection model, starting with all variables in the model, variables are removed one by one if they do not contribute to the regression. Predictors selected by both procedures were used for the final model. The probability value for the presence of SARS-CoV-2 antibodies combining the significant predictors from the final model into a single variable was included in a receiver operating characteristic (ROC) curve to evaluate the diagnostic accuracy, sensitivity and specificity of the

regression model. We determined the optimal cut-off for prediction of positive seroprevalence based on Youden's J.

Results

All households in Ischgl were contacted. In March 2020, Ischgl had 1617 principal residents and 2353 secondary residents. Most of the secondary residents were seasonal workers, who left the county without official leave notice prior to the lockdown. In cooperation with local authorities, we estimated that ∼250 secondary residents were present on the day of the study which results in an estimated population of 1867.

The complete sample consisted of $N = 1534$ participants: $n = 1493$ participants visited the study centre, $n = 41$ participants were visited at home. Participants were excluded if they were below 18 years of age ($n = 216$), had incomplete biosamples ($n = 61$), or did not return the completed questionnaire ($n = 806$). Participants included in the analysis ($n = 451$) were on average 47.4 years old (S.D. 16.8) and 52.5% female. More than half (52.3%) believed to have had a COVID-19 infection.

Testing of blood samples taken for this study revealed that, in the raw data used for this analysis, $n = 197$ (43.7%) participants had SARS-CoV-2 antibodies. The full seroprevalence data is to be published elsewhere.

More than one-third of participants did not report any symptoms ($n = 175$, 38.8%), $n = 72$ (16.0%) reported one symptom, $n = 73$ (16.2%) reported two symptoms and $n = 131$ (29.0%) reported three or more symptoms.

In the univariate analyses, the symptoms cough ($P < 0.001$), fever ($P < 0.001$), difficulties breathing ($P = 0.004$), olfactory/gustatory alterations ($P < 0.001$), loss of appetite ($P < 0.001$), nausea ($P = 0.001$), headache ($P < 0.001$), fatigue ($P < 0.001$), limb pain ($P < 0.001$), dyspnoea ($P = 0.016$) and vertigo ($P < 0.001$) were statistically significant predictors of the presence of SARS-CoV-2 antibodies ($P \geq 0.465$ for all other symptoms).

In the multivariate analyses, three predictors (cough, gustatory/olfactory alterations and limb pain) were selected in both, forward and backward selection and included in the final model (see Table 1). For these three symptoms, we show the overlap between symptoms in seropositive and seronegative participants in Figure 1.

Of the seropositive participants, $n = 175$ (88.4% of all seropositive participants) reported at least one symptom when considering all symptoms assessed; $n = 159$ participants (80.3% of all seropositive participants) reported at least one symptom when considering the three symptoms included in the model.

The predicted probability value from this multivariate model was included in a ROC analysis to determine the optimal cut-off score for this probability value to predict the presence of SARS-CoV-2 antibodies. The area under the ROC curve was 0.773 (95% CI 0.727–0.820, $P < 0.001$). The optimal cut-off was defined as the values with the highest Youden's J (predicted possibility of the presence of SARS-CoV-2 antibodies if ≥ 0.445). The sensitivity for this cut-off was 0.612, the specificity was 0.852.

Discussion

We identified cough, gustatory and/or olfactory alterations and limb pain as symptoms most predictive of SARS-CoV-2 antibodies.

Those symptoms were similar to those found in other studies on patients with COVID-19. A recent meta-analysis of commonly reported symptoms revealed that fever, cough and fatigue

Table 1. Predictor variables for the presence of SARS-CoV-2 antibodies ($n = 451$)

Predictor variable	B	Significance level	OR	95% CI for OR
Cough ('not at all')		0.003		
Cough ('a little')	0.630	0.033	1.88	1.05–3.35
Cough ('quite a bit')	1.206	<0.001	3.34	1.70–6.58
Cough ('very much')	0.159	0.686	1.17	0.54–2.53
Taste or smell alterations ('not at all')		<0.001		
Taste or smell alterations ('a little')	0.934	0.025	2.54	1.13–5.75
Taste or smell alterations ('quite a bit')	2.055	<0.001	7.81	2.90–21.00
Taste or smell alterations ('very much')	2.623	<0.001	13.78	5.90–32.17
Limb pain ('not at all')		<0.001		
Limb pain ('a little')	−0.516	0.020	0.60	0.27–1.30
Limb pain ('quite a bit')	0.571	0.113	1.77	0.88–3.59
Limb pain ('very much')	1.028	0.017	2.55	1.20–6.50
Constant	−1.228	<0.001	0.29	

Note. The model was statistically significant: χ^2 (9) = 123.851. $P < 0.001$ in predicting cases with SARS-CoV-2 antibodies. The model explained 33.6% of the variance (Nagelkerke R^2); *B*, unstandardised β; *OR*, odds ratio; *CI*, confidence interval.
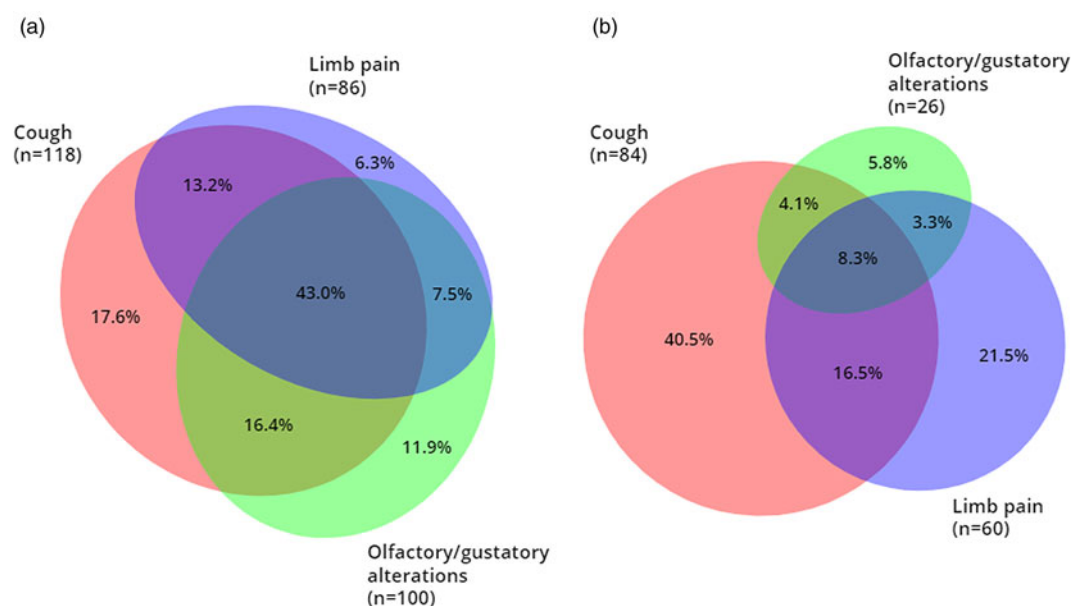
Fig. 1. Euler diagrams with overlap between significant symptoms of the regression.
Note. (a) All seropositive participants with at least one of the three symptoms ($n = 159$, 80.7%), (b) All seronegative participants with at least one of the three symptoms ($n = 121$, 47.6%). Of the seropositive participants, $n = 38$ (19.3%) reported none of the three symptoms, compared to $n = 133$ (52.4%) of seronegative participants ($\chi^2(1) = 51.554$, $P < 0.001$).

were amongst the most prominent symptoms [4]. However, not all symptoms differentiate well against other respiratory diseases [5]. In our study, the symptoms most predictive of SARS-CoV-2 were alterations in taste or smell. This conforms with other studies in which the alterations in (or sometimes loss of) smell or taste were a key symptom to distinguish patients with flu-like symptoms from patients with COVID-19 [6, 7]. Our results confirm that this symptom may be useful for screening for SARS-CoV-2 antibodies.

Compared to a similar model presented by Menni *et al.* [8], neither loss of appetite nor general fatigue increased the diagnostic accuracy of our multivariate model. Rather, limb pain was included as a predictor in our model. Those differences might be traced back to differences in sampling and testing approaches. While we tested for antibodies in the whole population, regardless of indication, most other prediction studies have samples that were tested due to an indication (e.g. experiencing symptoms or exposure) [9]. Our sample might include more mild or atypical cases in which slightly different symptomatology predicts past infection. Moreover, the model presented by Menni *et al.* [8] predicts current infection with SARS-CoV-2 and not the prevalence of antibodies. Currently, there are no studies that
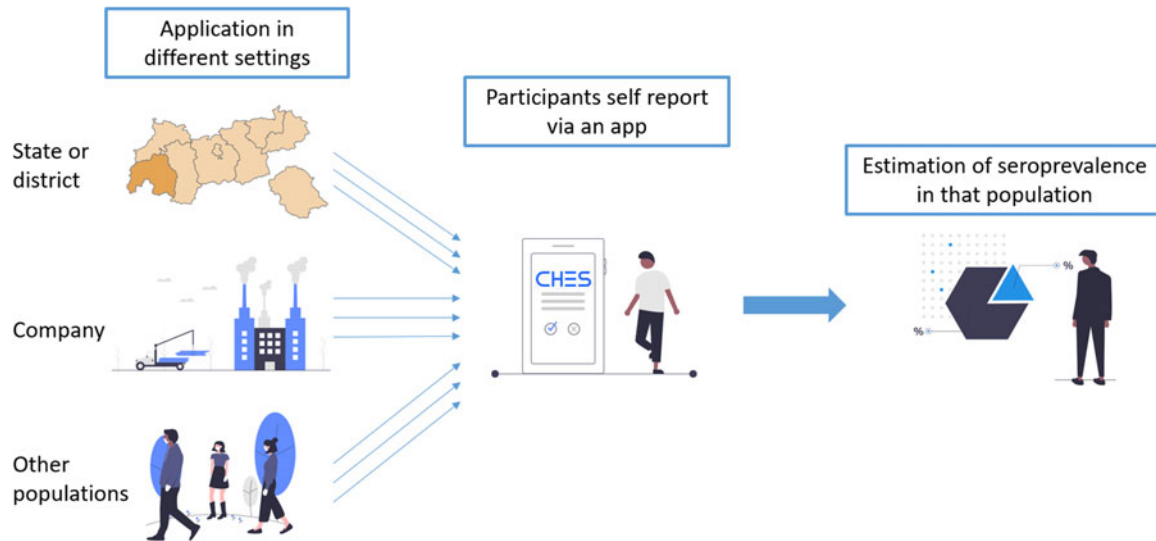
Fig. 2. Different applications of the regression model for the cost-effective and swift estimation of SARS-CoV-2 seroprevalence.

predict past infection with SARS-CoV-2 via self-reported symptoms and potential differences in prediction models should be subject to further investigation.

Our model may be used to estimate the seroprevalence in different settings and samples. By simply asking participants three questions about past symptoms, it is possible to reliably estimate the share of participants with a previous or recent SARS-CoV-2 infection. This may be used in population-based studies (e.g., assessing the seroprevalence in a certain area, in a company, or a social group). A web application building on the code basis of the Computer-Based Health System (CHES [10]) for large surveys or epidemiological contexts is currently undergoing final testing. Figure 2 displays how such a procedure could be done.

Strengths and limitations

A key strength of our study is that testing for SARS-CoV-2 seroprevalence was carried out regardless of symptomatology or indication.

Compared to approaches testing directly for SARS-CoV-2, namely PCR, our estimation model is based on antibody seroprevalence data from assays. This means that it can more accurately estimate the burden of current, and especially past disease in a sample. Another benefit lies in the economic and fast measurement; our approach can be used when testing capacity is limited or it is not feasible to conduct broad-scale sample testing.

While the diagnostic accuracy of our regression model is valuable for determining population-level seroprevalence, predictions for individuals are, like for most other prediction models [9], not recommended. Another caveat of this study is that the model has not yet been validated in an independent data set.

Also, due to the study design, it is not possible to determine whether the self-reported symptoms occurred prior to, during or after the infection with SARS-CoV-2 (i.e. were related or unrelated to COVID-19). The retrospective time frame in which we assessed symptoms was between February and April. Consequently, the interval between symptom occurrence and

questionnaire assessment was not constant, but potentially differed up to 12 weeks between the participants.

Finally, our study was conducted during the first wave of the pandemic. Research on COVID-19 is a rapidly evolving field and future studies will be needed to validate our results.

Conclusion

The model developed in our study may allow easy and economic evaluations of seroprevalence of SARS-CoV-2 antibodies at the population level.

References

1. **Wang C *et al.*** (2020) A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473.
2. **Hu B *et al.*** (2021) Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* **19**, 141–154.

3. **Rossman H** *et al.* (2020) A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nature Medicine* **26**, 634–638.

4. **Grant MC** *et al.* (2020) The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): a systematic review and meta-analysis of 148 studies from 9 countries. *PLoS ONE* **15**, e0234765.

5. **Pascarella G** *et al.* (2020) COVID-19 diagnosis and management: a comprehensive review. *Journal of Internal Medicine* **288**, 192–206.

6. **Dawson P** *et al.* (2020) Loss of taste and smell as distinguishing symptoms of COVID-19. *Clinical Infectious Diseases* **72**, 682–685.

7. **Yan CH** *et al.* (2020) Association of chemosensory dysfunction and COVID-19 in patients presenting with influenza-like symptoms. *International Forum of Allergy & Rhinology* **10**, 806–813.

8. **Menni C** *et al.* (2020) Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine* **26**, 1037–1040.

9. **Wynants L** *et al.* (2020) Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* **369**, m1328.

10. **Holzner B,** *et al.* (2012) The Computer-based Health Evaluation Software (CHES): a software for electronic patient-reported outcome monitoring. *BMC Medical Informatics and Decision Making* **12**, 126.