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

Running head: Estimating SARS-CoV-2 seroprevalence in Ischgl

Authors:

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, Bernhard Holzner ¹

Affiliations:

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 University Hospital of Trauma Surgery, Department of Surgery, Medical University of Innsbruck, Austria

⁵ Department of Visceral, Transplant and Thoracic Surgery, Center of Operative Medicine, Medical University of Innsbruck, Innsbruck, Austria

* Corresponding author:

Jens Lehmann

jens.lehmann@i-med.ac.at

Anichstraße 35

6020 Innsbruck, Austria

Acknowledgements: We thank all participants for their valuable time and effort. We thank the state of Tyrol for partly funding the study.

Declarations

Sources of funding:

The project was partially funded by the state Tyrol.

Research ethics and informed consent:

Participants were briefed about the study and signed an informed consent. The study was approved by the ethics committee of the Medical University of Innsbruck (app. number 1100/2020).

Data availability:

The complete epidemiological data to the study are currently undergoing publication elsewhere. The data used in this study are available at <https://osf.io/32mab>.

Conflict of interest:

Bernhard Holzner and Gerhard Rumpold hold intellectual property rights of the software tool CHES. All remaining authors declare that they have no conflict of interest.

Authors' contributions:

JL analyzed the data and wrote the manuscript. JMG analyzed the data and wrote the manuscript. BH, GR, DVL, WB, and LK, conceived the study, analyzed the data, and assisted in writing the manuscript. BSU conceived the study and assisted in writing the manuscript. BF, CO, and MS assisted in generating the data and writing the manuscript. All authors read and approved the final manuscript.

Abstract

We report the development of a regression model to predict prevalence of SARS-CoV-2 antibodies on a population level based on self-reported symptoms. We assessed participant-reported symptoms in the past twelve weeks, as well as 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 (SD 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: 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 AUC was 0.773 (95% CI: 0.727-0.820). Our regression model may be used to estimate seroprevalence on a population-level and a web application is being developed to facilitate use of the model.

Keywords:

COVID-19; SARS-CoV-2; antibodies; symptom; anosmia; dysgeusia

Background

The coronavirus disease (COVID-19) is a respiratory illness that emerged in December 2019 in China and has since become a global health threat [1]. COVID-19 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 [2]. 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 [2].

Such self-reported symptoms may be used to quickly and economically determine an estimate of the prevalence in the population. 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.

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.

Methods

Procedure

Data collection was carried out between April 21st and 27th. All residents in Ischgl were invited via a letter to participate in the study and visit the study center (a recreational center in Ischgl). Participants that were not able to visit the study center (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 venipuncture. 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 neutralization assay. In addition, 110 randomly selected Euroimmun and Abbott positive samples were also tested for neutralizing 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 February 1th). Symptoms assessed were cough, fever, troubles breathing, sore throat, gustatory or olfactory alterations, loss of appetite, nausea, vomiting, constipation, diarrhea, headache, fatigue, limb pain, dyspnea, and vertigo. Symptoms were rated on a four-point scale: 1 ('not at all'), 2 ('a little'), 3 ('quite a bit'), 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. number 1100/2020).

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. For regression analysis, we used forward selection and backward elimination with likelihood ratio (LR) to determine significant ($p < 0.05$) predictors. Predictors selected by both procedures were used for the final model. The probability value for 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. The complete sample consisted of N=1534 participants: n=1493 participants visited the study center, n=41 participants were visited at home. Participants were excluded if they were below 18 years of age (n=216) if they had incomplete bio samples (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 (SD 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 a 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 ($p<0.001$), dyspnea ($p=0.016$), and vertigo ($p<0.001$) were statistically significant predictors of 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.

[insert **Table 1** Predictor variables for presence of SARS-CoV-2 antibodies (n=451)]

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 presence of SARS-CoV-2 antibodies. The AUC 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 presence of SARS-CoV-2 antibodies if greater than or equal to 0.445). The sensitivity for this cut-off was 0.612, the specificity was 0.852.

[insert **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 ($\chi^2(1)=51.554$, $p < 0.001$).

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 were amongst the most prominent symptoms [3]. However, not all symptoms differentiate well against other respiratory diseases [4]. In our study, the symptom 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 [5,6]. Our results confirm that this symptom may be useful for screening of COVID-19.

Compared to a similar model presented by Menni et al [7], 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. 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) [8]. Our sample might include more mild or atypical cases in which slightly different symptomatology predicts past infection. Our model may be used to estimate the seroprevalence in different settings and samples. By simply asking participants three questions on 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, [9]) for large surveys or epidemiological contexts is currently undergoing final testing. Figure 2 displays how such a procedure could be done.

[insert **Fig. 2** different applications of the regression model for the cost effective and swift estimation of SARS-CoV-2 seroprevalence]

A key strength of our study is that testing for SARS-CoV-2 seroprevalence was carried out regardless of symptomatology or indication. Consequently, our results and the model we provide are mostly unbiased by selective testing and may offer reliable estimates of seroprevalence.

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.

202 While the diagnostic accuracy of our regression model is valuable for determining
203 population-level seroprevalence, predictions for individuals are, like for most other prediction
204 models [8], not recommended. Another caveat of this study is that the model has not yet
205 been validated in an independent dataset. Finally, we retrospectively assessed symptom
206 occurrence in a time frame between February and April. Consequently, the interval between
207 symptom occurrence and questionnaire assessment was not constant but potentially differed
208 up to 12 weeks between participants.

209 **Conclusion**

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

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020 Feb 15;395(10223):470–3.
2. Rossman H, Keshet A, Shilo S, Gavrieli A, Bauman T, Cohen O, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nat Med*. 2020 May;26(5):634–8.
3. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, et al. 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*. 2020;15(6):e0234765.
4. Pascarella G, Strumia A, Piliiego C, Bruno F, Del Buono R, Costa F, et al. COVID-19 diagnosis and management: a comprehensive review. *J Intern Med*. 2020;288(2):192–206.
5. Dawson P, Rabold EM, Laws RL, Connors EE, Gharpure R, Yin S, et al. Loss of Taste and Smell as Distinguishing Symptoms of COVID-19. *Clin Infect Dis*. 2020 Jun 21;
6. 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*. 2020;10(7):806–13.
7. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med*. 2020;26(7):1037–40.
8. Wynants L, Calster BV, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020 Apr 7;369.
9. Holzner B, Giesinger JM, Pinggera J, Zugul S, Schöpf F, Oberguggenberger AS, et al. The Computer-based Health Evaluation Software (CHES): a software for electronic patient-reported outcome monitoring. *BMC Med Inform Decis Mak*. 2012 Nov 9;12:126.