














BMJ Open Trends and symptoms of SARS-CoV-2 infection: a longitudinal study on an Alpine population representative sample

Giulia Barbieri ^{1,2}, Massimo Pizzato ³, Martin Gögele ¹,
Daniele Giardiello ¹, Christian X Weichenberger ¹, Luisa Foco ¹,
Daniele Bottigliengo ¹, Cinzia Bertelli ³, Laura Barin ¹,
Rebecca Lundin ¹, Peter P Pramstaller ¹, Cristian Pattaro ¹,
Roberto Melotti ¹

To cite: Barbieri G, Pizzato M, Gögele M, *et al.* Trends and symptoms of SARS-CoV-2 infection: a longitudinal study on an Alpine population representative sample. *BMJ Open* 2023;**13**:e072650. doi:10.1136/bmjopen-2023-072650

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2023-072650>).

PPP, CP and RM contributed equally.

Received 10 February 2023
Accepted 18 May 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Ms Giulia Barbieri;
giulia.barbieri@eurac.edu and
Dr Cristian Pattaro;
cristian.pattaro@eurac.edu

ABSTRACT

Objectives The continuous monitoring of SARS-CoV-2 infection waves and the emergence of novel pathogens pose a challenge for effective public health surveillance strategies based on diagnostics. Longitudinal population representative studies on incident events and symptoms of SARS-CoV-2 infection are scarce. We aimed at describing the evolution of the COVID-19 pandemic during 2020 and 2021 through regular monitoring of self-reported symptoms in an Alpine community sample.

Design To this purpose, we designed a longitudinal population representative study, the Cooperative Health Research in South Tyrol COVID-19 study.

Participants and outcome measures A sample of 845 participants was retrospectively investigated for active and past infections with swab and blood tests, by August 2020, allowing adjusted cumulative incidence estimation. Of them, 700 participants without previous infection or vaccination were followed up monthly until July 2021 for first-time infection and symptom self-reporting: COVID-19 anamnesis, social contacts, lifestyle and sociodemographic data were assessed remotely through digital questionnaires. Temporal symptom trajectories and infection rates were modelled through longitudinal clustering and dynamic correlation analysis. Negative binomial regression and random forest analysis assessed the relative importance of symptoms.

Results At baseline, the cumulative incidence of SARS-CoV-2 infection was 1.10% (95% CI 0.51%, 2.10%). Symptom trajectories mimicked both self-reported and confirmed cases of incident infections. Cluster analysis identified two groups of high-frequency and low-frequency symptoms. Symptoms like fever and loss of smell fell in the low-frequency cluster. Symptoms most discriminative of test positivity (loss of smell, fatigue and joint-muscle aches) confirmed prior evidence.

Conclusions Regular symptom tracking from population representative samples is an effective screening tool auxiliary to laboratory diagnostics for novel pathogens at critical times, as manifested in this study of COVID-19 patterns. Integrated surveillance systems might benefit from more direct involvement of citizens' active symptom tracking.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Longitudinal study of SARS-CoV-2 symptoms and incident cases in the absence of competing infections and vaccine prophylaxis.
- ⇒ Population-based representative sample susceptible to first-time SARS-CoV-2 infection.
- ⇒ Self-reported SARS-CoV-2 infections and symptoms may limit generalisability to most severe cases.

INTRODUCTION

After an unprecedented pathogen-led death toll in modern times,¹ the COVID-19 pandemic is shifting to an endemic phase in most territories.^{2,3} However, recrudescence of reinfections by SARS-CoV-2 novel emergent variants and high contagion rates are still a matter of public health concern, particularly towards the most vulnerable individuals. The social and economic difficulties of maintaining most non-pharmaceutical containment actions in the long term, as well as limited awareness on effective self-prevention measures, including vaccine hesitancy, warrant further attention for monitoring the spread of SARS-CoV-2 and other infectious agents.^{4,5}

Continuous monitoring and surveillance have been a challenge for health authorities throughout the COVID-19 pandemic. Routine indicators are based on the number of incident cases over the number of people tested, and on the rate of hospital admissions and deaths linked to COVID-19.⁶ These indicators mainly rely on system capacity to track infections and perform laboratory analyses, as well as effective information systems. Such indicators based on confirmed positive cases may also lack accuracy and prompt reporting

in emergent or fast-spreading situations. Lastly, diverse cultural backgrounds, adaptive healthcare policies, ever-increasing healthcare costs and the process of development for testing methods may influence case number estimation.⁷

Symptoms screening and monitoring may provide complementary and timely surveillance data, if systematically and regularly conducted in a reference population sample, and under certain conditions.^{8–12} Focusing on syndromic infections, symptoms' comprehensibility and suitability to investigate novel pathogen specificities are essential. While presymptomatic and asymptomatic infected individuals contribute to the spread of infections like SARS-CoV-2,¹³ viral shedding peaks at around the time of symptom appearance.¹⁴ Early reporting of symptoms can therefore allow prompt identification of critical areas for further prevention or containment actions. Large-scale repeated cross-sectional studies based on digital social platform anonymised data were capable to capture trends of symptoms and cases at both global and regional area levels.^{15 16} However, smaller scale ad hoc representative studies may allow participation tracking, relatively better population coverage and control of information and selection biases by detailed or linked auxiliary data. Being independent of the healthcare testing, diagnosis and referral capacity, and requiring substantially less resources, early estimate of infection spread by symptoms monitoring promises to be a better tool for the management of healthcare infection emergencies.

Population-based longitudinal studies are particularly suited to investigate the temporal patterns of incident events and symptoms in a pandemic like COVID-19.^{9 11 17–19} While planning, recruiting and accrual of evidence are time consuming, these studies have a unique and active role in shaping surveillance strategies.⁴ By the end of August 2020, we recruited a representative random sample of adult residents in the Alpine rural district of Vinschgau/Val Venosta (South Tyrol, Italy).^{20 21} As few as 4.4 cases per 10 000 district inhabitants and no active cases (16 confirmed cases overall) had been identified by the healthcare system at the time, in contrast to nearby and other European regions, where infection prevalence estimates ranged from 23.1% to 42.4% in the same period.^{22–24} At the national level, there were standard recommendations to adopt individual containment actions (eg, face masks, hands hygiene, physical distance) up until 24 October 2020, when additional public restrictions, especially school and business closures, were introduced.²⁵ A governmental decree²⁶ introduced three risk-level zones for each regional and autonomous provincial authority (including South Tyrol) in Italy, starting 6 November 2020 (online supplemental figure S1).²⁷ Each risk zone was alternatively identified for any period with the colours yellow, orange and red, orderly corresponding to increasing levels of non-pharmaceutical interventions. Since 26 April 2020, containment actions were progressively, although cautiously relaxed, yet within an emergency governance.²⁸

We characterised participants through ad hoc questionnaires, in-person molecular testing and blood sample collection at baseline, and monthly follow-up digital questionnaires from September 2020 until July 2021 (figure 1A). In this report we describe (1) the distribution of symptomatic episodes and the symptom patterns over the whole follow-up period, and (2) the dynamic relationship of specific symptoms and their aggregate patterns with incident SARS-CoV-2 infections, as captured by first ever positive swab test self-reporting (figure 1B). We discuss the potential utility of these data to complement diagnostic-based surveillance in similar emergent infectious events.

METHODS

Study design

We invited an age-sex stratified random sample targeting 1450 participants of the Cooperative Health Research in South Tyrol (CHRIS) study²⁹ representative of all adult residents in the district to participate in the CHRIS COVID-19 prospective investigation between 13 July and 28 August 2020.^{20 21} An online screening questionnaire covered SARS-CoV-2-related anamnesis since 1 February 2020 until participation to the baseline clinical visit for blood drawing and testing. Afterwards, repeated follow-up online questionnaires were sent to each baseline participant every 4 weeks for 1 year to monitor for SARS-CoV-2-related events.

Laboratory assessments

Laboratory assessments included a Roche Elecsys Anti-SARS-CoV-2 assay serum antibody (SAb) test to measure the level of SARS-CoV-2-specific antibodies, as a measure of past infection, and a swab PCR test to identify active SARS-CoV-2 infections.²⁰ Serum samples were also submitted to an assessment of their ability to inhibit the transduction of a lentiviral vector pseudotyped with the SARS-CoV-2 spike protein (neutralisation test). Pseudotyped vectors, transducing a gene encoding a fluorescent protein, were incubated along with scalar dilutions of serum that was then inoculated onto Huh-7 cells. Percentage of transduced fluorescent cells was quantified using the High-Content Molecular Devices ImageXpress Micro Confocal on nuclei counterstaining with Hoechst 33342 and the serum dilution associated with 50% inhibition of transduction (ID50 value) was finally estimated from each derived sigmoidal curve.

Baseline and follow-up questionnaires

The baseline questionnaire asked about participants' sociodemographic and household information, lifestyle determinants, regular therapies and chronic conditions. A detailed section was dedicated to SARS-CoV-2 anamnesis comprising previous diagnosis, occurrence of symptoms and close contacts with infected or symptomatic individuals.²⁰

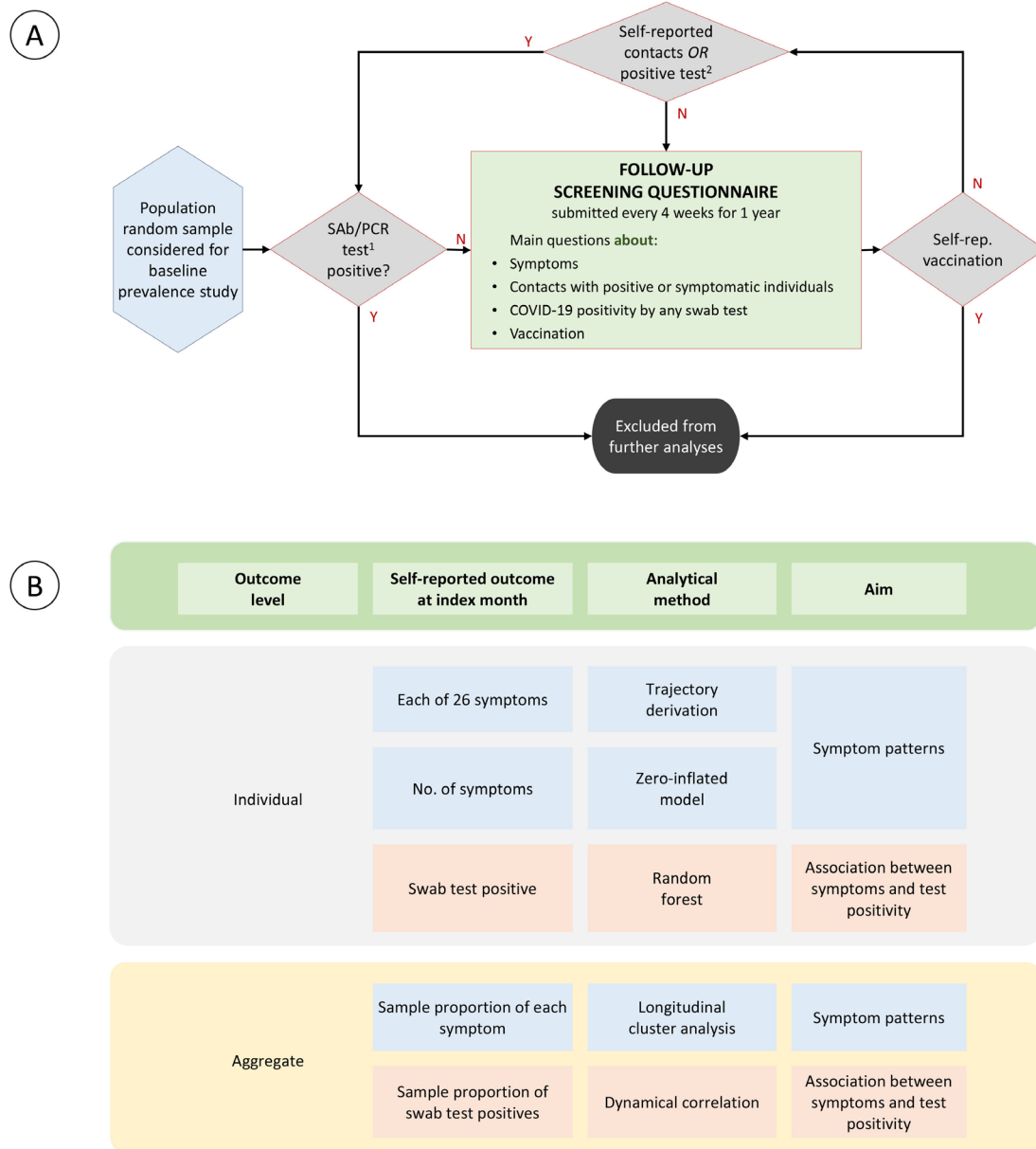


Figure 1 Graphical representation of the study design and methods. (A) Study design. (B) Methods: each aim (either pale blue or pale pink foreground) was addressed with both aggregate (pale yellow background) and individual-level (light grey background) outcome and methods. ¹Performed at the study centre. ²Self-reported contacts with positive or symptomatic individuals OR any positive swab test. SAb, serum antibody.

Follow-up questionnaires focused on SARS-CoV-2 anamnesis and additionally included ‘vaccination status’ from the beginning of the vaccination campaign (27 December 2020). At each follow-up, participants replied to a main filter question on SARS-CoV-2-related events since previous participation: on a positive reply, the participant would fill in the rest of the questionnaire, otherwise all subsequent responses were considered negative, or else missing in case of no response.^{20 21}

Longitudinal analysis framework

All individuals susceptible to first SARS-CoV-2 infection before enrolling in the follow-up were included in the longitudinal analysis. Participants’ records were prospectively integrated in the analysis to the end of the study or

until first infection or dose of vaccination to avoid misinterpretation of symptoms induced by vaccination with those of first infection with SARS-CoV-2^{30–32} and changes in individual susceptibility (figure 1A).^{33 34}

Variable definitions

At baseline, a positive result to either the PCR or SAb test defined a SARS-CoV-2-positive case. Follow-up questionnaire completion dates were specific to each participant at every 4-week wave. For any self-reported symptomatic episode, the date at symptom onset was asked. However, for any self-reported swab test, the date of testing was not asked. To match symptom-onset dates with symptom-free dates from independent respondents, we defined the ‘index month of reporting’ as either the month of

the first symptomatic episode or, had no symptoms been reported, the month corresponding to ‘shifting’ the actual questionnaire completion date by 15 days backwards (ie, to the midpoint of each questionnaire reference period). In case of duplicate dates of symptom onset in successive questionnaires, the reported event with the most symptoms was retained in the analysis. For the longitudinal analyses, we defined 12 time points (t) corresponding to the months from August 2020 until July 2021. At each time point, a participant could report positivity to any of 26 symptoms. We defined a dichotomous variable ‘occurrence of symptom j ’ as s_{ij} that was 1 if the i th individual reported symptom j at time t , 0 otherwise. We calculated the total number of symptoms reported by each participant at each time point, referred to as the ‘count of symptoms’ (S_{it}). Presence of SARS-CoV-2 infection was defined at each time point as either ‘positive’ (positivity to any swab test, henceforth T_{it}^+) or ‘negative’ (T_{it}^-), based on self-reported questionnaire information. At each index month, the incidence of each symptom j was defined as the proportion of symptomatic cases for that symptom, and the incidence of T^+ as the proportion of positive cases, both estimated among all retained questionnaire respondents for that month. T_{it}^+ , s_{ij} and S_{it} were used as outcome measures in the individual-level analyses, whereas incidences of symptoms and positive cases per month were used as outcome measures in aggregate analyses.

Statistical analyses

Population-calibrated cumulative incidence of SARS-CoV-2 at baseline was estimated using the Clopper-Pearson method for extreme proportions, including sampling weights to account for non-response.²⁰ Pairwise association testing between infection status at baseline (only nine positive participants were detected) and each of 64 personal characteristics (described elsewhere²⁰) or ID50 values used non-parametric statistics, as described in the online supplemental material.

The number of symptoms reported by respondents at each time point (S_{it}) displayed excess of zeros and overdispersion. To model the temporal evolution of this variable, we fitted a weighted zero-inflated negative binomial mixed model with random intercept (reflecting the individual), using the index month as a fixed effect categorical predictor and adjusting for sex and decade of age. For these analyses we used the ‘glmmTMB’ R package v1.1.4. To assess which individual symptoms were mostly associated with self-reported T^+ , we conducted a random forest analysis, generating 200 classification trees, using Gini impurity node splitting. SEs were estimated based on 100 bootstrap samples, with a forest of 20 trees each. Each tree randomly included 8 out of 25 (~33%) of all the available predictors.³⁵ We set the minimum node size at five observations, which cast the tree depth. The discriminatory ability of each symptom was defined as the average decline in prediction accuracy on the out-of-bag samples, when excluding that symptom from the model.³⁵

The relative increase in the misclassification rate (MR) was quantified and named as the relative change (RC).³⁶ For these analyses we used the *hrf* function from the R package ‘htree’ v2.0.0.³⁷

The individual symptom time trajectories were modelled on incidence data via longitudinal cluster analysis based on the k-means method.³⁸ The optimal partition was obtained through an iterative process, according to the Calinski-Harabasz criterion,³⁹ using the R package ‘kml’ v2.4.1.⁴⁰

Finally, we conducted a dynamic correlation analysis^{41 42} to assess the aggregation among symptom trajectories and the association between symptoms occurrence and incidence of self-reported positive swab tests (T^+). Since the number of symptoms was large compared with the number of reported occurrences across time points, regularisation was applied to obtain a shrinkage estimate of the correlation matrix. This analysis was conducted with the *dyn.cor* function of the R package ‘longitudinal’ v1.1.13.⁴³

All statistical analyses were run in the R environment v4.1.1.

Patient and public involvement

Participants of the study were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS

Characteristics of the study sample

We enrolled 845 participants at baseline (participation: 58%; females: 51.6%; age—years, median: 50, range: 20–94). By 28 August 2020, nine participants tested positive to SARS-CoV-2 infection (eight by SAb test and one by PCR test), corresponding to an adjusted cumulative incidence of 1.10% (95% CI 0.51%, 2.10%). Baseline positivity to infection was associated with self-report of a previous positive SAb test, having been isolated because of suspected or confirmed SARS-CoV-2 infection and specific symptoms such as loss of taste and loss of smell (online supplemental table S1). Participants with any positive test at baseline had higher antibody neutralisation levels (online supplemental table S1). In those with both tests negative, neutralisation levels were not associated with any characteristic except for the municipality of residence (online supplemental table S2).

After excluding the nine baseline positive participants, 134 participants who did not fill any follow-up questionnaires and two who had been vaccinated before participation into follow-up, 700 participants were available for longitudinal analyses. Their characteristics were similar to the baseline sample (online supplemental table S3). Each participant filled in a median of nine follow-up questionnaires (range: 1–13; IQR: 5–11). Based on self-report, throughout the study period, 200 (28.6%) individuals never undertook any swab test, 194 (27.7%) underwent one test only and 306 (43.7%) had ≥ 2 tests.

Table 1 Characteristics of the 700 study participants during the follow-up period by index month and according to symptoms and swab test reporting.

Index month	Susceptible participants (n)	Characteristics		Self-reported swab test n (%)	Any symptoms			
		Males (%)	Age Median (IQR)		n (%)	Self-reported T ⁺ n (%)	Median (IQR)	Self-reported T ⁺ n (%)
August	446	48.9	48.6 (37.3–60.2)	72 (16.1)	17 (3.8)	0 (0.0)	2 (1.0–3.0)	0 (0.0)
September	478	49.4	48.3 (36.8–59.8)	28 (5.9)	19 (4.0)	0 (0.0)	2 (1.0–4.5)	0 (0.0)
October	508	50.4	48.3 (37.1–60.1)	88 (17.4)	55 (10.8)	17 (30.9)	3 (1.5–6.5)	18 (3.6)
November	516	48.4	48.6 (36.8–59.0)	291 (56.4)	63 (12.2)	26 (41.9)	3 (1.0–7.0)	29 (5.6)
December	520	46.7	48.4 (37.2–59.2)	115 (22.1)	40 (7.7)	11 (28.2)	3 (2.0–4.2)	14 (2.7)
January	470	49.6	48.4 (37.6–60.2)	104 (22.1)	44 (9.4)	17 (38.6)	3.5 (2.0–5.0)	20 (4.3)
February	425	50.4	49.9 (37.4–61.1)	129 (30.4)	53 (12.5)	22 (42.3)	3 (2.0–6.0)	23 (5.5)
March	411	47.9	48.9 (37.3–60.7)	94 (23.0)	18 (4.4)	5 (29.4)	2.5 (1.2–6.0)	5 (1.2)
April	385	46.8	47.2 (36.3–56.6)	83 (21.6)	10 (2.6)	1 (10.0)	4 (2.2–7.8)	1 (0.3)
May	295	48.8	45.7 (35.3–58.6)	81 (27.6)	7 (2.4)	1 (14.3)	2 (2.0–4.5)	1 (0.3)
June	259	42.9	49.6 (37.5–64.4)	27 (10.5)	2 (0.8)	0 (0.0)	3 (2.0–4.0)	1 (0.4)
July	185	40	48.3 (38.7–58.4)	11 (6.0)	3 (1.6)	0 (0.0)	2 (1.5–2.5)	0 (0.0)
Overall	700	48.4	50 (37.4–62.2)	500 (71.4)	228 (32.6)	94 (13.4)*	–	99 (14.1)

*Over the whole time frame, 94 individuals reported to be simultaneously symptomatic and positive, representing 96.9% of all positive cases (n=99) and 41.2% of all symptomatic cases (n=228).

The maximum observed number of tests per person was 8. 472 participants (67.4%) never reported any symptoms, 158 (22.6%) reported one symptomatic episode and 70 (10.0%) reported more than one (table 1).

Symptom patterns and test positivity

The months from October 2020 to February 2021 were characterised by a higher probability of reporting symptoms than August. October, November and January were the months with the highest number of symptoms per participant, among those reporting symptoms, compared with August (figure 2, online supplemental table S4). Random forest analysis (MR=0.013) identified loss of smell (RC=15.3%) as the most predictive symptom of a positive swab test, followed by fatigue or tiredness (RC=14.5%), joint or muscle pain (RC=12.8%), headache (RC=8.4%), fever (RC=5.2%) and loss of taste (RC=4.8%; figure 3).

Longitudinal cluster analysis of symptom patterns

Figure 4A shows the distribution of each symptom incidence over time and the overlapping pattern of self-reported positive swab tests. Two clusters best explained the aggregate trajectories of symptoms, which were characterised by high-frequency and low-frequency symptoms, respectively (figure 4B). The time series of the positive swab tests was best reflected by symptoms included in the high-frequency cluster: cold, joint or muscle pain, fatigue or tiredness, sore throat or hoarseness and headache. Alternative analytical solutions that forced symptom aggregations in more than two clusters produced similar

mean trajectories, with no major changes in the symptom aggregation (online supplemental figure S2).

Dynamic correlation analysis

We observed a generally large and positive dynamic correlation among the symptom trajectories (figure 4; online supplemental figure S3), especially among symptoms included in the high-frequency cluster (dynamic correlation coefficient, r , between 0.72 and 0.85). Joint or muscle pain was the most correlated symptom with headache ($r=0.85$) and fatigue or tiredness ($r=0.83$). Almost all observed symptoms were highly correlated with self-reported T⁺ ($r \geq 0.60$), except for abdominal pain ($r=0.40$) and otitis ($r=0.45$). The symptoms most correlated with T⁺ were loss of smell ($r=0.85$), joint or muscle pain ($r=0.85$) and headache ($r=0.85$).

DISCUSSION

This study provides a moving picture of COVID-19 pandemic dynamics since inception over its hardest hitting phases to date in the Val Venosta/Vinschgau district (South Tyrol, Italy). By summer 2020, the resident population was nearly naïve to SARS-CoV-2 infection by official figures, likely thanks to the absence of super spreader events and the rapid application of strict nationwide containment measures.^{27 44} While still modest in absolute size, our data indicate a cumulative rate of infection in the Val Venosta/Vinschgau district at 11 per 1000 inhabitants, which roughly corresponds to 25-fold

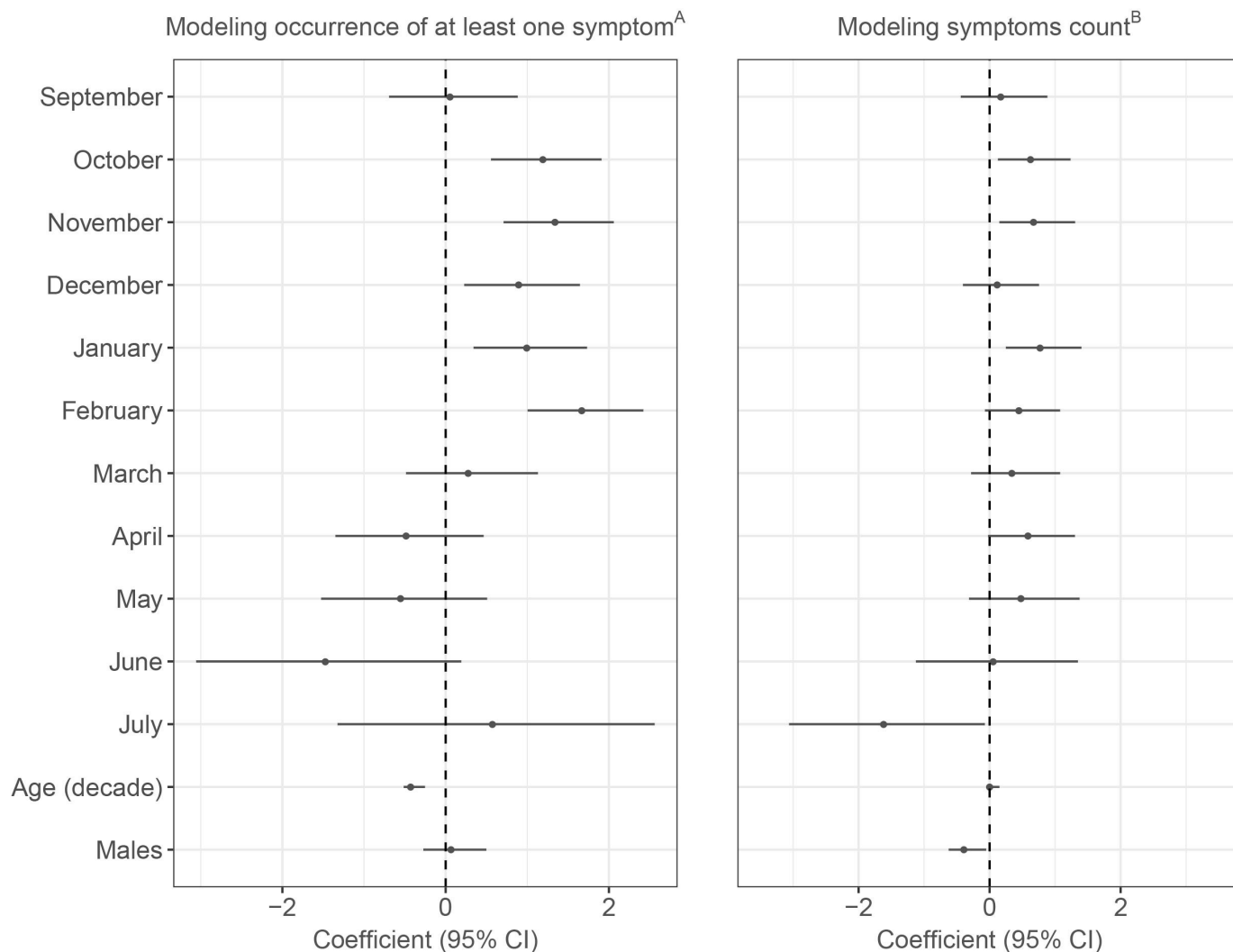


Figure 2 Results of the zero-inflated negative binomial model. ^AModelling the probability of a symptomatic episode with any number of symptoms (reversed log odds from the original model). ^BModelling the expected number of symptoms conditional on a symptomatic episode.

the figure derived from officially confirmed cases. This evidence confirms the large number of unidentified positive individuals that had characterised the beginning of the pandemic and the hidden ongoing viral shedding.⁴⁵ Regardless, the observed incidence was very low compared with other areas of central Europe and Northern Italy that, at the same time, had already been more severely affected by the novel coronavirus.^{22–24 45} This situation provided the opportunity to observe incident cases of infection prospectively since fall 2020 in a nearly naïve susceptible population.

Among participants susceptible to first infection, we observed two peaks of self-reported incident positive cases, corresponding to November 2020 and February 2021, that is, before widespread vaccine availability. These peaks coincided with the peaks of officially recorded confirmed cases, intensive care unit admissions and deaths for the whole South Tyrol region (online supplemental figure S1).⁴⁶ This suggests that, during the study period, in the Val Venosta/Vinschgau district the contagion pattern was

similar to other areas under the same risk zone mitigation strategies.^{27 47} Also, the pattern of symptomatic episodes and number of reported symptoms closely resembled the pandemic pattern. While the peaks of symptomatic episodes were identified in October 2020 and February 2021, the load of symptoms for any given episode peaked in October/November 2020 and January 2021. All symptoms followed similar time trajectories, mimicking self-reported positive swab tests as well as the official figures of the pandemic dynamic (compare figure 4A and online supplemental figure S1). This pattern was likely the result of a lack of competing infectious diseases at the same time frame⁴⁸ favoured by the adoption of strict isolation rules and containment actions, which varied modestly in response to the pandemic spread over the course of the study period (online supplemental figure S1).²⁷

The five most discriminatory symptoms of infection were, in order of relevance, loss of smell, fatigue or tiredness, joint or muscle pain, headache and fever. This confirms the previous reports that identified a similar set

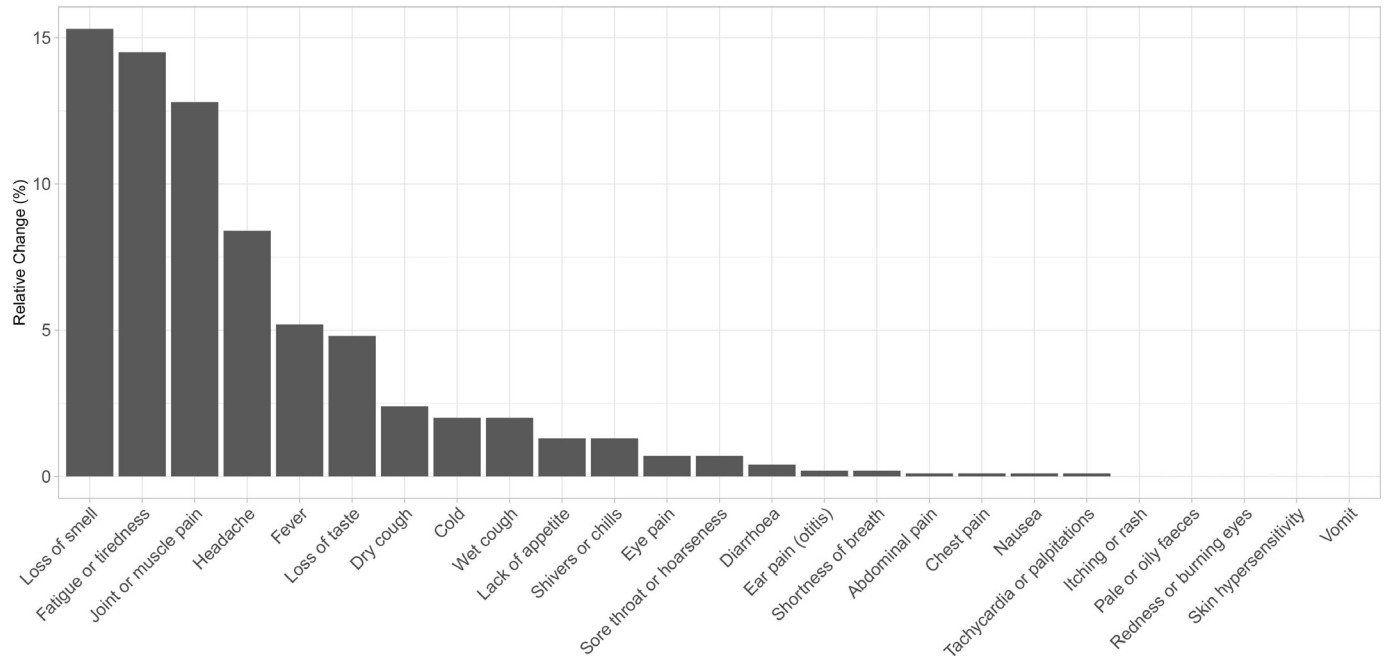


Figure 3 Discriminatory capacity of symptoms derived from random forest analysis. The relative change represents the decrease in accuracy in the discriminatory ability of the model if the symptom is not included.

of symptoms to predict extant infections.^{10 12 49} In our analysis, loss of taste had a small discriminatory capacity when accounting for loss of smell, following evidence of cellular mechanisms of infection acting at the smell receptor level.⁵⁰

On an aggregate level, our best fitting longitudinal cluster analysis split the symptoms into two separate

patterns of high-frequency versus low-frequency symptoms. The two clusters followed similar trajectories with no substantial crossover, suggesting limited role for emerging variants on specific symptom frequency. Such evidence is also supported by the estimation of the dominant variants circulating during the study period.⁵¹ The most prevalent variants were emerging mixed SARS-CoV-2 strains until

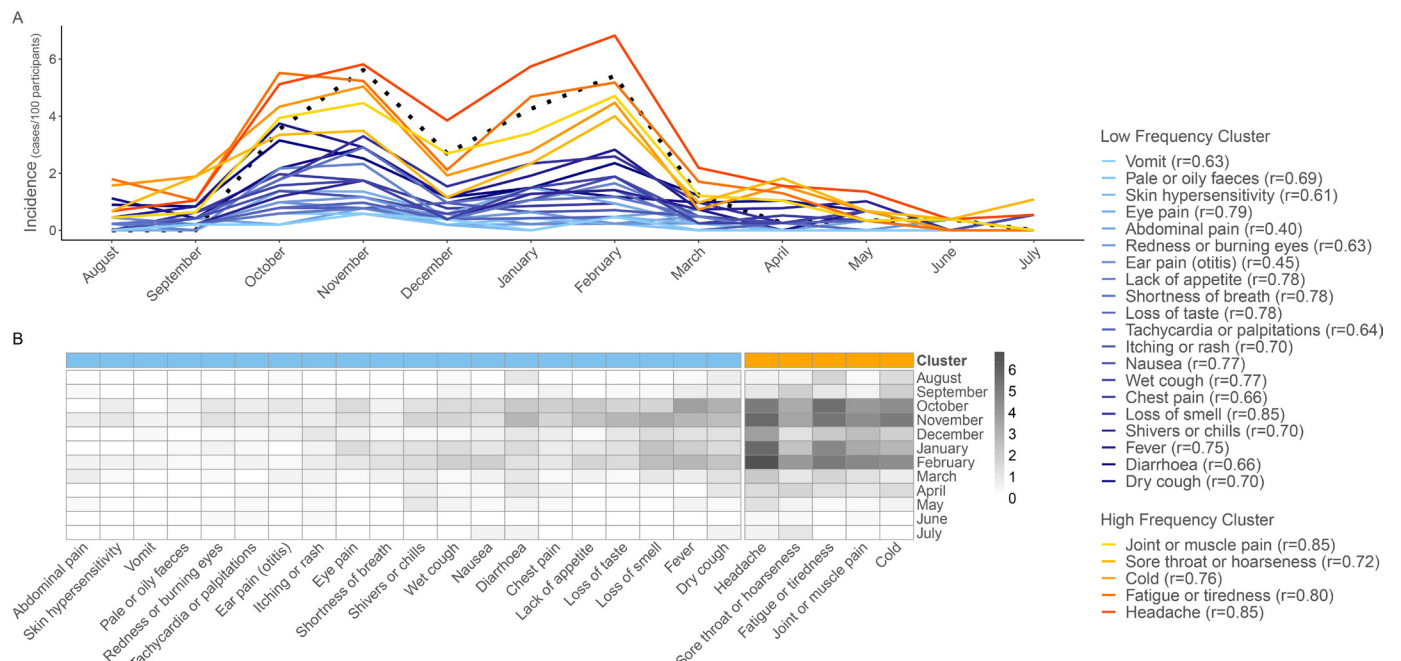


Figure 4 Results of the cluster analysis on the incidence of symptoms over time. (A) Trajectories of symptoms (coloured plain lines) and self-reported T⁺ (black dotted line). The dynamic correlation between the trajectory of each symptom and self-reported T⁺ is included within brackets in the legend keys. (A, B) Two clusters of symptoms are distinguished by warm (high frequency) and cold (low frequency) colours, respectively. (B) The heatmap in grey scale represents incidence of each symptom across time and reflects y-axis incidence in panel A.

the end of 2020. Subsequently, there was a steep uptake of the Alpha variant, which then left predominance to the Delta variant, by July 2021. To our knowledge, none of these variants provided strong scientific evidence for diverse symptom patterns among SARS-CoV-2-infected individuals.⁵²

Fever and loss of smell fell in the low-frequency cluster. In the high-frequency cluster, the five most frequent symptoms were fatigue or tiredness, joint or muscle pain, headache, cold and sore throat or hoarseness. Fatigue or tiredness, joint or muscle pain, headache and loss of smell showed very high dynamic correlation with swab test positivity. These results need careful interpretation due to possible effects of seasonality, reporting bias and mediating pathways. For example, the link between headache, the most reported symptom, and the peaks of infection could be partially mediated through psychosocial distress (eg, confusion, worrisome attitudes) or isolation and indoor confinement,^{53–55} as corroborated in our results by limited discriminatory capacity. While fever and loss of smell are familiar and recognisable symptoms, their lower frequency and good discriminatory capacity of infection support their high specificity to COVID-19 in a relatively low prevalence and protected context. Nevertheless, the occurrence of infection may not imply those symptoms.^{22 56} Finally, joint or muscle pain and fatigue or tiredness are generic symptoms that often occur with ordinary influenza-like illness and in combination with other symptoms. In our analyses, these symptoms maintained both high frequency and discriminatory capacity, as observed elsewhere.^{9 12}

Our study is a rare account of a population-based prospective observational study attempting to map incident diagnoses and apply symptom manifestation to COVID-19 screening over a long period from the start of the pandemic. For example, a similar cohort study conducted in Lübeck (Germany) investigated patterns of infection through PCR test and measured antibody response from March 2020 until February 2021, mimicking the official figures of infection.¹⁷ A deluge of questionnaires screened for symptoms and self-reported positive tests over the 2020 late-spring/summer period of low incidence and two extra examinations covered the pandemic expansion phase in November 2020 and February 2021. Our manageable sample size and confined catchment area was privileged with regular, frequent and comprehensive assessments of symptoms and incident cases, which extended through the whole evolution of the pandemic in Italy until its next temporal dampening by July 2021. Over this whole period, a nuanced description and grouping of symptom trajectories reflected the trend of incident cases.

The patterns of symptoms and cases of infection described in our study may complement those of ad hoc monitoring systems based on confirmed positive cases to inform effective surveillance strategies. Routine surveillance systems of acute respiratory infections rely on a network of sentinel general practices and testing

laboratories subject to voluntary participation, which are often adequate to monitor nationwide seasonal patterns of ordinary respiratory infections. However, the uncontrollable spread of SARS-CoV-2 revealed suboptimal surveillance frameworks for the detection of novel pathogens of concern and for the timely identification of outbreaks in restricted areas.⁵⁷ Population-based reports like ours show that remote digital technology applications may be employed to screen and trace symptomatic human infections on a reference or largely compliant population sample during a health emergency.^{10 17 49} Real-time electronic data sharing between citizens seeking assistance and general practices or emergency departments in an overarching global surveillance network could be an efficient and effective model for epidemiological surveillance. Extant and adaptable privacy preserving norms and secured technologies would now allow to make these data flow a critical and ordinary source for both individual care and preventative public health actions.

To our knowledge, this is the first population-based longitudinal study conducted in Europe able to trace the pattern of incident cases both retrospectively and prospectively over an 18-month long course of the COVID-19 pandemic since inception and match data-driven trajectories of symptom clusters. An additional strength of this study was the calibration of the study sample to be representative of the adult population of a wide rural area, which was free of confirmed positive active cases at the time of initial recruitment and susceptible to primary SARS-CoV-2 infection. Our carefully designed 4-week follow-up witnessed high compliance of study participants, as manifest by the large amount of follow-up questionnaires completed, and may have both limited the recall bias and increased the precision of temporal allocation of symptoms and other events.

Several limitations should also be considered. The sample size was limited. However, this was calibrated to estimate a cumulative incidence between 0.01% and 1.1% with a confidence level of 99.0%.²⁰ Given our final estimate corresponded to the upper bound of such an interval, the collected sample size provided sufficient power for reliable cumulative incidence estimates. Symptoms like fever, cough and shortness of breath may have been partly obscured as possible indicators of more severe infections, as suggested by the low frequency reported.^{9 11} Accordingly, we cannot exclude selection bias concerning the most severely affected and distressed participants. Next, we acknowledge the possibility of missing marginalised or less digitally literate individuals, especially in relation to using an online screening questionnaire. However, participants who filled at least one questionnaire beyond baseline were approximately 83%.²⁰ Moreover, according to a recent local survey, 81% of South Tyrolean families comprising members in the age range of 16–74 reported having access to the internet from home.^{20 58 59} A peculiarity of our study was that we followed up participants who were susceptible of first infection. When participants reported a positive test or having received a dose of

vaccine, they were systematically excluded from further online screening. This might have prevented the observation of the evolution of symptoms over the course of infection on some respondents who might happen to be positive close to the time of infection and questionnaire response. Hitherto, the observational period was unique in several aspects: no major competing illness events, evolving and dynamic containment norms, developing testing capacities and techniques, no prior widespread vaccination prophylaxis and ultimately a mass testing campaign conducted in late November 2020 in the whole of South Tyrol, which allowed the identification of many hidden positive cases in a short time window.⁶⁰ While these aspects may impair the generalisability of our findings, the symptomatic patterns and reported cases matched the dynamics of the pandemic from official data of a larger area, suggesting that widespread control measures could balance out areas at different levels of risk. Lastly, symptoms and incident cases of infections were self-reported. However, self-reported symptoms have shown greater breadth than electronic medical record-extracted symptoms.⁴⁹ Moreover, another report on the same cohort²¹ showed that if participants were to experience symptoms they would generally be tested for COVID-19, limiting the possibility of testing bias in our results.

In conclusion, regular remote symptom tracking is feasible in the context of an emergent pandemic situation, such as that of COVID-19, and can closely characterise the temporal pattern of infection. Such an approach would be logistically and economically advantageous, and more sustainable than long-term, large-scale testing and tracing alternatives. Surveillance systems should broaden and integrate their infrastructure to involve multiple participatory units and include symptom reporting in their templates through citizens' direct involvement via digital technology or other means.

Author affiliations

¹Institute for Biomedicine (affiliated to the University of Lübeck), Eurac Research, Bolzano, Italy

²Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy

³Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy

Acknowledgements We are grateful to all CHRIS COVID-19 study participants. We thank the collaborators of both Eurac Research and the Healthcare System of the Autonomous Province of Bolzano for field and laboratory operations, the representatives, and staff of the local administrative authorities for support and all ancillary volunteers who made the study possible. The authors thank the Department of Innovation, Research and University of the Autonomous Province of Bozen/Bolzano for covering the Open Access publication costs.

Contributors CP, PPP, RM and MG conceived the CHRIS COVID-19 study. CP, RM and GB conceived the investigation topic object of the present work. GB, MG, RM, DG and CXW performed all statistical analyses. MP and CB performed the neutralising antibody analyses. GB, MP, MG, DG, CXW, DB, CP and RM interpreted the results. GB, RM and CP drafted the manuscript. MP, MG, DG, CXW, LF, DB, CB, LB, RL and PPP contributed to the critical revision and editing of the manuscript. GB acts as guarantor for this study. All authors approved the manuscript.

Funding The CHRIS COVID-19 study was supported by Eurac Research, the South Tyrolean Health Authority, and the Department of Innovation, Research and

University of the Autonomous Province of Bolzano. The present research was conducted within the project 'PACE: Partnership to Accelerate COVID-19 rEsearch in South Tyrol', funded by the Department of Innovation, Research and University of the Autonomous Province of Bolzano within the 2019–2021 Research Program (unique project code: D52F20000770003). This work was supported by a grant from Fondazione VRT/CARITRO to MP.

Disclaimer Funding sources did not have any role in the research conduction, writing of the manuscript and decision to submit it for publication.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The CHRIS COVID-19 study was authorised by the Ethics Committee of the Healthcare System of the Autonomous Province of Bolzano/Bozen with deliberation number 53-2020 on 27 May and 22 July 2020.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. CHRIS study data can be requested for research purposes by submitting a dedicated request to the CHRIS Access Committee. Please contact access.request.biomedicine@eurac.edu for more information on the process.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Giulia Barbieri <http://orcid.org/0000-0002-3531-0725>
 Massimo Pizzato <http://orcid.org/0000-0002-9491-1483>
 Martin Gögele <http://orcid.org/0000-0003-4785-4348>
 Daniele Giardiello <http://orcid.org/0000-0002-9005-9430>
 Christian X Weichenberger <http://orcid.org/0000-0002-2176-0274>
 Luisa Foco <http://orcid.org/0000-0002-6694-3322>
 Daniele Bottigliengo <http://orcid.org/0000-0002-8802-3834>
 Cinzia Bertelli <http://orcid.org/0000-0002-9443-1957>
 Laura Barin <http://orcid.org/0000-0001-8139-4686>
 Rebecca Lundin <http://orcid.org/0000-0002-8380-4834>
 Peter P Pramstaller <http://orcid.org/0000-0002-9831-8302>
 Cristian Pattaro <http://orcid.org/0000-0002-4119-0109>
 Roberto Melotti <http://orcid.org/0000-0003-4115-317X>

REFERENCES

- 1 Wang H, Paulson KR, Pease SA, *et al*. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *Lancet* 2022;399:1513–36.
- 2 Emanuel EJ, Osterholm M, Gounder CR. A national strategy for the "new normal" of life with COVID. *JAMA* 2022;327:211:211–2..
- 3 Lavine JS, Bjornstad ON, Antia R. Immunological characteristics govern the transition of COVID-19 to Endemicity. *Science* 2021;371:741–5.
- 4 Suk JE, Pharris A, Beauté J, *et al*. Public health considerations for Transitioning beyond the acute phase of the COVID-19 pandemic in the EU/ EEA. *Euro Surveill* 2022;27:2200155.
- 5 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20:533–4.
- 6 Country overview report: week 38 2022. n.d. Available: <https://covid19-country-overviews.ecdc.europa.eu/index.html#key-indicators-week-38>

- 7 Beauté J, Adlhoch C, Bundle N, *et al.* Testing indicators to monitor the COVID-19 pandemic. *Lancet Infect Dis* 2021;21:1344–5.
- 8 Saad NJ, Moek F, Steitz F, *et al.* A longitudinal study on symptom duration and 60-day clinical course in non-hospitalised COVID-19 cases in. *Euro Surveill* 2021;26:2001757.
- 9 Eythorsson E, Helgason D, Ingvarsson RF, *et al.* Clinical spectrum of Coronavirus disease 2019 in Iceland: population based cohort study. *BMJ* 2020;371:m4529.
- 10 Menni C, Valdes AM, Freidin MB, *et al.* Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;26:1037–40.
- 11 Sudre CH, Lee KA, Lochlainn MN, *et al.* Symptom clusters in COVID-19: a potential clinical prediction tool from the COVID symptom study App. *Sci Adv* 2021;7:eabd4177.
- 12 Elliott J, Whitaker M, Bodinier B, *et al.* Predictive symptoms for COVID-19 in the community: REACT-1 study of over 1 million people. *PLoS Med* 2021;18:e1003777.
- 13 Buitrago-Garcia D, Ipekci AM, Heron L, *et al.* Occurrence and transmission potential of asymptomatic and Presymptomatic SARS-Cov-2 infections: update of a living systematic review and meta-analysis. *PLoS Med* 2022;19:e1003987.
- 14 He X, Lau EHY, Wu P, *et al.* Temporal Dynamics in viral shedding and Transmissibility of COVID-19. *Nat Med* 2020;26:672–5.
- 15 Astley CM, Tuli G, Mc Cord KA, *et al.* Global monitoring of the impact of the COVID-19 pandemic through Online surveys sampled from the Facebook user base. *Proc Natl Acad Sci U S A* 2021;118:e2111455118.
- 16 Salomon JA, Reinhart A, Bilinski A, *et al.* The US COVID-19 trends and impact survey: continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proc Natl Acad Sci U S A* 2021;118:e2111454118.
- 17 Klein C, Borsche M, Balck A, *et al.* One-year surveillance of SARS-Cov-2 transmission of the ELISA cohort: A model for population-based monitoring of infection risk. *Sci Adv* 2022;8:eabm5016.
- 18 Ridgway JP, Tideman S, Wright B, *et al.* Rates of COVID-19 among Unvaccinated adults with prior COVID-19. *JAMA Netw Open* 2022;5:e227650.
- 19 Pulliam JRC, van Schalkwyk C, Govender N, *et al.* Increased risk of SARS-Cov-2 Reinfection associated with emergence of Omicron in South Africa. *Science* 2022;376:eabn4947.
- 20 Pattaro C, Barbieri G, Foco L, *et al.* Prospective Epidemiological, molecular, and genetic characterization of a novel Coronavirus disease in the Val Venosta/Vinschgau: the CHRIS COVID-19 study protocol. *Pathog Glob Health* 2022;116:128–36.
- 21 Giardiello D, Melotti R, Barbieri G, *et al.* Determinants of SARS-Cov-2 Nasopharyngeal testing in a rural community sample susceptible of first infection: the CHRIS COVID-19 study. *Pathog Glob Health* 2023;0:1–10.
- 22 Melotti R, Scaggiante F, Falciani M, *et al.* Prevalence and determinants of serum antibodies to SARS-Cov-2 in the general population of the Gardena Valley. *Epidemiol Infect* 2021;149:e194.
- 23 Knabl L, Mitra T, Kimpel J, *et al.* High SARS-Cov-2 Seroprevalence in children and adults in the Austrian ski resort of Ischgl. *Commun Med (Lond)* 2021;1:4.
- 24 Stefanelli P, Bella A, Fedele G, *et al.* Prevalence of SARS-Cov-2 IgG antibodies in an area of northeastern Italy with a high incidence of COVID-19 cases: a population-based study. *Clin Microbiol Infect* 2021;27:633.
- 25 Presidenza del Consiglio dei Ministri. Decreto del Presidente del Consiglio Dei Ministri 24 Ottobre 2020. n.d. Available: <https://www.gazzettaufficiale.it/eli/id/2020/10/25/20A05861/SG>
- 26 Presidenza del Consiglio dei Ministri. Decreto del Presidente del Consiglio Dei Ministri 3 Novembre 2020. n.d. Available: <https://www.gazzettaufficiale.it/eli/id/2020/11/04/20A06109/sg>
- 27 Pelagatti M, Maranzano P. Assessing the effectiveness of the Italian risk-zones policy during the second wave of COVID-19. *Health Policy* 2021;125:1188–99.
- 28 Presidente della Repubblica. Decreto-Legge 22 Aprile 2021, N. 52. n.d. Available: <https://www.gazzettaufficiale.it/eli/id/2021/04/22/21G00064/sg>
- 29 Pattaro C, Gögele M, Mascalzoni D, *et al.* The cooperative health research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J Transl Med* 2015;13:348.
- 30 Rimmel A. COVID vaccines and safety: what the research says. *Nature* 2021;590:538–40.
- 31 Alhazmi A, Alamer E, Daws D, *et al.* Evaluation of side effects associated with COVID-19 vaccines in Saudi Arabia. *Vaccines (Basel)* 2021;9:674.
- 32 Klugar M, Riad A, Mekhemar M, *et al.* Side effects of mRNA-based and viral vector-based COVID-19 vaccines among German Healthcare workers. *Biology (Basel)* 2021;10:752.
- 33 Polack FP, Thomas SJ, Kitchin N, *et al.* Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med* 2020;383:2603–15.
- 34 Ssentongo P, Ssentongo AE, Voleti N, *et al.* SARS-Cov-2 vaccine effectiveness against infection, symptomatic and severe COVID-19: a systematic review and meta-analysis. *BMC Infect Dis* 2022;22:439.
- 35 Probst P, Wright MN, Boulesteix A-L. Hyperparameters and tuning strategies for random forest. *WIREs Data Mining Knowl Discov* 2019;9.
- 36 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- 37 Sexton J. Historical tree ensembles for longitudinal data. 2018. Available: <https://CRAN.R-project.org/package=htree>
- 38 Genolini C, Falissard B. Kml: K-means for longitudinal data. *Comput Stat* 2010;25:317–28.
- 39 Calinski T, Harabasz J. A Dendrite method for cluster analysis. *Comm in Stats - Theory & Methods* 1974;3:1–27.
- 40 Genolini C, Alacoque X, Sentenac M, *et al.* Kml and Kml3D: R packages to cluster longitudinal data. *J Stat Softw* 2015;65:1–34.
- 41 Schäfer J, Strimmer K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional Genomics. *Stat Appl Genet Mol Biol* 2005;4:Article32.
- 42 Oppen-rhein R, Strimmer K. Using Regularized dynamic correlation to infer Gene dependency networks from time-series Microarray data. In proceedings of the 4TH International workshop on computational systems biology (WCSB 2006). 2006:12–3.
- 43 Strimmer RO-R and K. Longitudinal: analysis of multiple time course data. 2021. Available: <https://CRAN.R-project.org/package=longitudinal>
- 44 Sun K, Wang W, Gao L, *et al.* Transmission Heterogeneities, Kinetics, and Controllability of SARS-Cov-2. *Science* 2021;371:eabe2424.
- 45 Lavezzo E, Franchin E, Ciavarella C, *et al.* Suppression of a SARS-Cov-2 outbreak in the Italian municipality of VO. *Nature* 2020;584:425–9.
- 46 Pelagatti M, Rigon T. Report Dati COVID-19 in Italia. n.d. Available: <https://bicocca-datalab.shinyapps.io/covid19/>
- 47 Bonifazi G, Lista L, Menasce D, *et al.* A study on the possible merits of using symptomatic cases to trace the development of the COVID-19 pandemic. *Eur Phys J Plus* 2021;136:481.
- 48 Rapporto Influnet. n.d. Available: <https://w3.iss.it/site/rmi/influnet/rapporto/rapportoinflunet.aspx>
- 49 Mizrahi B, Shilo S, Rossman H, *et al.* Longitudinal symptom Dynamics of COVID-19 infection. *Nat Commun* 2020;11:6208.
- 50 Spinato G, Fabbris C, Polesel J, *et al.* Alterations in smell or taste in mildly symptomatic outpatients with SARS-Cov-2 infection. *JAMA* 2020;323:2089–90.
- 51 Khare S, Gurry C, Freitas L, *et al.* GISAIID's role in pandemic response. *China CDC Wkly* 2021;3:1049–51.
- 52 Schulze H, Bayer W. Changes in symptoms experienced by SARS-Cov-2-infected individuals – from the first wave to the Omicron variant. *Front Virol* 2022;2.
- 53 Feng L-S, Dong Z-J, Yan R-Y, *et al.* Psychological distress in the shadow of the COVID-19 pandemic: preliminary development of an assessment scale. *Psychiatry Res* 2020;291:113202.
- 54 Invitto S, Romano D, Garbarini F, *et al.* Major stress-related symptoms during the Lockdown: A study by the Italian society of Psychophysiology and cognitive Neuroscience. *Front Public Health* 2021;9:711132.
- 55 Patrono A, Renzetti S, Manco A, *et al.* COVID-19 aftermath: exploring the mental health emergency among students at a northern Italian University. *IJERPH* 2022;19:8587.
- 56 Poletti P, Tirani M, Cereda D, *et al.* Association of age with likelihood of developing symptoms and critical disease among close contacts exposed to patients with confirmed SARS-Cov-2 infection in Italy. *JAMA Netw Open* 2021;4:e211085.
- 57 Cobb NL, Collier S, Attia EF, *et al.* Global influenza surveillance systems to detect the spread of influenza-negative influenza-like illness during the COVID-19 pandemic: time series Outlier analyses from 2015–2020. *PLoS Med* 2022;19:e1004035.
- 58 Istituto provinciale di statistica (ASTAT). Indagine Multiscopo Sulle Famiglie - 2018. n.d. Available: https://astat.provincia.bz.it/it/news-pubblicazioni-info.asp?news_action=4&news_article_id=624375
- 59 Istituto Provinciale di Statistica (ASTAT). Accesso a Internet E Utilizzo Della Rete - 2018. n.d. Available: https://astat.provincia.bz.it/it/news-pubblicazioni-info.asp?news_action=4&news_article_id=623584
- 60 Ferrari D, Stilloan S, Tonin M. Does COVID-19 mass testing work? the importance of accounting for the epidemic Dynamics. 2021. Available: <http://arxiv.org/abs/2104.14813>