

Assessing SARS-CoV-2 Prevalence in Austria with Sample Surveys in 2020 - a Report

Alexander Kowarik
Statistics Austria

Matea Paškvan
Statistics Austria

Marlene Weinauer
Statistics Austria

Matthias Till
Statistics Austria

Karin Schrittwieser
Statistics Austria

Tobias Göllner
Statistics Austria

Abstract

Since the beginning of the SARS-CoV-2 pandemic a main metric is the amount of infected people at any given time. We present a valid assessment of the population acutely infected by SARS-CoV-2 in Austria at three distinct points in time: April, May and November 2020. The population of these sample surveys includes people aged 16 or older living in private households. Participants were tested with PCR (polymerase chain reaction) tests using nose-throat swabs. Based on these tests, it is assumed that the number of acute SARS-CoV-2 infections was below 11,000 individuals, or 0.15% of the target population, in April, below 6,000, in May and below 265,000 in November (i.e. the upper limit of the 95% confidence interval). In November a comparison with the Austrian Epidemiological Reporting System reveals that more than estimated 50% of acute infections remain undetected by the official reporting obligations.

Keywords: Austria, Covid-19 survey, stratified sample, SARS-CoV-2, R.

1. Introduction

”How many people are currently infected by SARS-CoV-2?” was the main topic in the daily news in 2020. The answer to this question is quite complex as the number of individuals (officially) tested positive has some weaknesses and thus underestimates the actual number of infected people: Data on current SARS-CoV-2 infections, as available from the Austrian Epidemiological Reporting System (EMS), provide insight into the prevalence of infected people and build the basis for determining the effective reproduction number of SARS-CoV-2 in Austria. However, this data might underestimate the actual number of infected people, mainly two reasons: Austria’s testing strategy (restricted by test capacity) and the characteristics of the illness itself: About 50% of infected people are asymptomatic (Nishiura, Kobayashi, Miyama, Suzuki, Jung and Katsuma Hayashi, Kinoshita, Yang, Yuan, Akhmetzhanov, and Linton 2020). Based on these two reasons it is reasonable that official data is unable to accurately estimate the number of unreported cases. To fill this gap, the Federal Ministry of Education, Science and Research (BMBWF) commissioned Statistics Austria in cooperation

with the Medical University of Vienna and the Red Cross Austria to study the prevalence of SARS-CoV-2 in Austria in April, May and November 2020.

For each study a random sample was drawn from Statistics Austria’s sampling frame for households and individuals which is based on the Austrian central register of residents (ZMR). Sampled people were invited to participate. The respondents completed a questionnaire on the presence of any symptoms, on their general health status, their personal well-being and the subjective assessment of the COVID-19 pandemic and its repercussions. From those who consented to being tested a nose-throat swab sample was taken and analyzed for COVID-19 by means of a PCR test. These swab samples were taken by the Austrian Red Cross. PCR tests were carried out and analyzed by the Medical University of Vienna ¹.

Austria was one of the first countries, conducting a nationwide study of this kind. Before Statistics Austria conducted the three surveys, a survey was already conducted by the SORA institute at the beginning of April 2020 (Ogris and Oberhuber 2020): From 1 until 6 April a random sample of 2,880 people living in Austria was sampled to estimate the number of unreported cases of COVID-19 infections. The proportion of people tested positively in the weighted sample was 0.33%. 1,544 PCR test results were included in the final analysis. This proportion represents about 28,500 people among the population. As this pioneering survey helped Austria’s government in decision making, the following three studies were commissioned.

The three aims of this report are:

1. To provide the reader with a compact overview of methodology and main results, of all three studies.
2. To provide the reader with a compact overview of differences in methodology and main results, of all three studies.
3. To provide the reader with a description of how to access the data, as well as how to replicate the results.

The document is structured as follows: Chapter 2 describes the sampling designs of all three studies. Chapter 3 describes the study design: The invitation to participate, the questionnaire - which was to be filled in before receiving a PCR test - and the testing procedure. Furthermore, the response rates are presented within chapter 3. In chapter 4 weighting and inference is discussed: The weighting and calibration procedure are described, as well as methods used for error calculation. The results are presented in chapter 5. Finally a conclusion and a further outlook is given. At the end access to the Austrian Social Science Data Archive (AUSSDA) data sets is introduced. For the sake of clarity, the three different survey waves are indicated with the abbreviations shown in table 1.

Table 1: Survey abbreviations

Survey ID	Time of testing
CV01	21 to 24 April 2020
CV02	26 to 30 May 2020
CV03	12 to 14 November 2020

In parallel to the nationwide studies, three regional studies for SARS-CoV-2 prevalence and seroprevalence were conducted in Austria:

- At the end of April 2020 1,473 people were tested with a PCR and an antibody test in the municipality of Ischgl. The main result was the estimated seroprevalence of 42.4%

¹In the last survey in November 2020 additionally blood was sampled from participants to evaluate the seroprevalence with an antibody test

(Knabl, Mitra, Kimpel, Roessler, Volland, Walser, Ulmer, Pipperger, Binder, Riepler *et al.* 2021).

- In June 2020 835 people from the municipality of Weißenkirchen were tested for antibodies and for the presence of an active infection with a PCR test (Ladage, Höglinger, Ladage, Adler, Yalcin, Harzer, and Braun 2021). The estimated seroprevalence was 9% and the estimated prevalence 1.2%.
- In March and April 2020 1,824 people in Reichenau an der Rax were tested for antibodies (Zifko, Schmiedlechner, Saelens, Zifko, Wagner, Assadian, Grisold, and Stingl 2021) resulting in an estimated seroprevalence of 6.5%.

2. Sampling

In an initial step, a stratified random sample was drawn based on the central population register (ZMR) in each study. The population for sampling consisted of individuals aged 16 or older living in a private household in Austria, corresponding to about 7.4 million people. People in hospitals or institutions such as care homes, institutional households etc. were not included.

Figures 1 and 2 show distributions of the sample for CV01. The illustrations were generated using the R (R Core Team 2020) packages **sp** (Bivand, Pebesma, and Gomez-Rubio 2013) and **leaflet** (Cheng, Karambelkar, and Xie 2019).

An overview of sampling designs is given in table 2: Sample Designs of CV01 and CV02 differed mainly in gross size and small stratification adaptations: In CV01 as well as in CV02, each sample was carried out as a two-stage stratified random sample. To take resources of the Austrian Red Cross into account, the population was divided into two parts (A and B): A - individuals in cities or within a 20-minute radius by car from Red Cross drive-in test sites - and B - all others. The gross sample was drawn in proportion to the population from these two parts.

In CV03 the sampling design was changed: The whole population was assigned to part A. Again, small adaptations in gross size and stratification, were conducted. This change was made, because the requirement of local clustering of individuals disappeared since all tests were now carried out at Red Cross test sites.

The sampling procedure is subsequently described separately for both parts A and B, in which part B is only relevant for CV01 and CV02.

2.1. Subsample A

In CV01 and CV03, the population consisted of individuals in cities or within a 20 minute radius by car from Red Cross drive-in test sites. In CV03, the whole population is included. This part of the sample was treated as a primary sampling unit (PSU) with an inclusion probability of 1 in CV01 and CV02.

For the second stage of part A a stratified sample of individuals was drawn. The probability of being drawn was slightly elevated for the educationally disadvantaged strata (those with completed or non-completed compulsory schooling only). For the sample of the survey conducted in late April, individuals living in areas with higher prevalence of SARS-CoV-2 had an increased sampling probability to increase the chance of getting infected people in the sample, which in turn would have a positive effect on the estimation accuracy.

In CV03 the whole sample was drawn from part A.

2.2. Subsample B

Subsample B was used in CV01 and CV02: The population here consisted of all other individuals not contained in part A; thus, individuals residing in areas that were mostly thinly

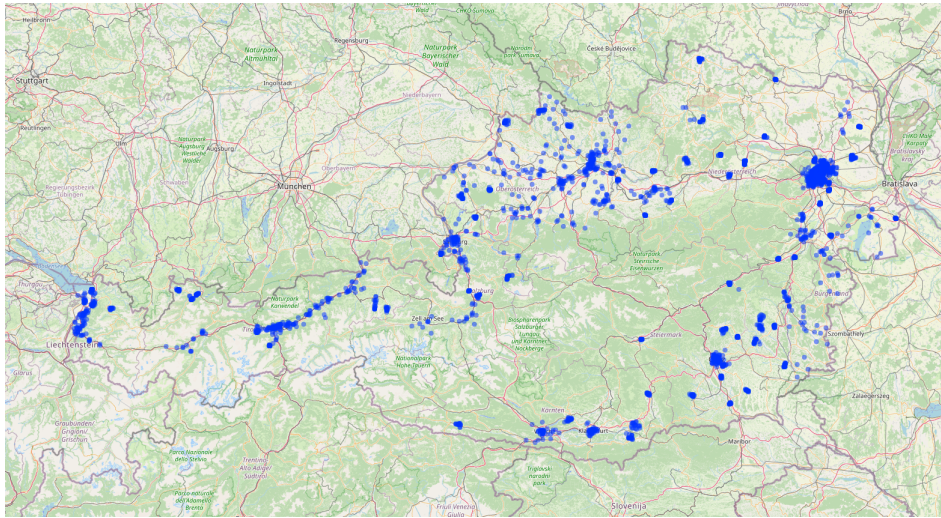


Figure 1: Distribution of the total sample across Austria

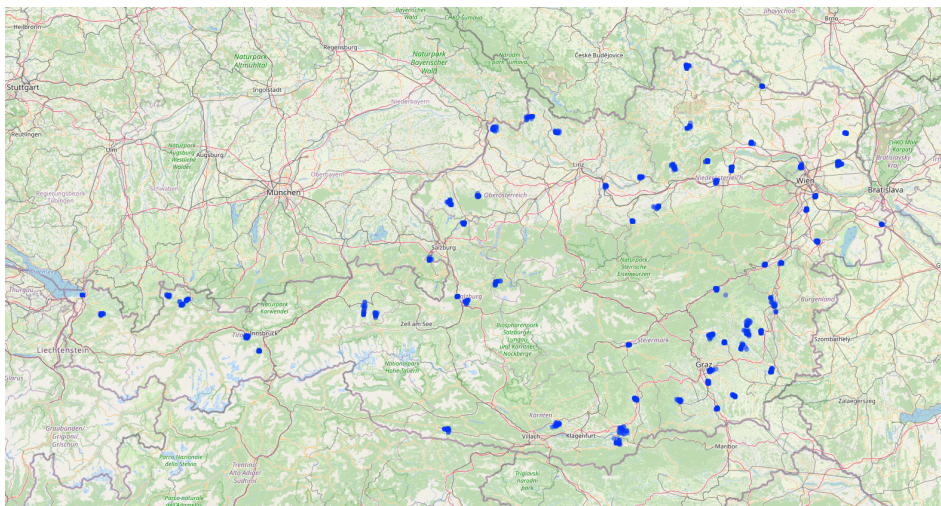


Figure 2: Distribution of the sample by statistical enumeration district (subsample B)

Table 2: Time table CV01, CV02, CV03

Survey ID	Gross size	Sampling design
CV01	2800	division of the population in two parts (A, B); two-stage stratified random sampling; oversampling A: educationally disadvantaged, people living in risk categories oversampling B: educationally disadvantaged stratification A: federal state, risk category, education; stratification B: federal state, risk category, degree of urbanization
CV02	3,470 + 254 *	analogous to CV01; different definition of risk category in part A; no oversampling for risk category in part A; no stratification for risk category and degree of urbanization in part B.
CV03	7,800	no part B; part A: analogous to CV02, with additional stratification of urbanization and citizenship

* redrawn Vienna-sample

populated. This part of the sample entailed drawing statistical enumeration districts (SED) in the first sampling stage to minimize the burden of Red Cross medical personnel (minimize times for travel and for changing personal protective equipment) and, if possible, to reach all individuals for whom tests were planned. The SEDs in part B of the sample were stratified and the probability of their being drawn was approximately proportional to the size of the SED within the stratum. For the second stage of part B a stratified sample of 15 individuals (CV02) and 20 individuals (CV01) was drawn per SED. The probability of being drawn was slightly elevated for people in educationally disadvantaged strata. For the April survey, 63 SEDs were drawn (20 individuals per SED, i.e. a total of 1,260 individuals) and for the May survey, 90 SEDs were drawn (15 individuals per SED, i.e. a total of 1,350 individuals).

In CV02 a small sample was redrawn for Vienna: From a daily monitoring of responses, it was expected early on in CV02 that total response rates would be lower than for CV01 (particularly in urban areas), so an additional sample was drawn exclusively for Vienna ($N = 254$, Vienna only).

3. Study design

A letter was sent to all sampled people, inviting them to participate. People who agreed to participate and answered the questionnaire (online or by phone) were subsequently tested for infection using PCR analysis, the different steps taken are detailed below.

In CV03 participants were additionally tested for SARS-CoV-2 antibodies. These results are not presented in this report, but are available on [Statistic Austria's website](#) (in German).

3.1. Invitation to participate

A similar contact strategy was implemented in CV01, CV02 and CV03 with slight adaptations between surveys.

A letter of invitation and an extensive data privacy sheet were sent to all sampled people inviting them to participate in the study. In CV03 sampled people were invited in two tranches with a time lag of 7 days. In addition to the letter, detailed information on the

different stages of the study was provided online (Statistics Austria 2020).

Those without access to the Internet or unwilling to complete the questionnaire online were requested to send a text message showing their willingness to participate. They were subsequently contacted by Statistics Austria and interviewed on the phone. Additionally, in CV02 and CV03, respondents could call Statistics Austria directly and were surveyed immediately. Whenever the respondent had not yet finalized the questionnaire but a phone number was available, the respondent was contacted by phone and motivated to participate. In some cases, the letter of invitation had not yet arrived (even though it had been sent via priority mail). Moreover, a reminder postcard was sent around six days after the invitation letter to remind all individuals who had not yet participated to take part.

Important dates are summarized in table 3. Due to external circumstances only a very limited amount of time was available for completing the survey. This holds especially true for the first two studies. In April, ten days were provided to complete the entire survey (from the time the letter was sent to the time the last swab was collected), in May fifteen days.

Table 3: Time table CV01, CV02, CV03

Survey ID	Invitation letter	PCR tests
CV01	14 April	21 to 24 April 2020
CV02	15 May / 22 May*	26 to 30 May 2020
CV03	13 Oct. / 20 Oct.**	12 to 14. Nov. 2020

* redrawn Vienna-sample ** second tranche

3.2. Questionnaire

There were three different questionnaires:

- Main Questionnaire: This questionnaire was filled in by respondents via CATI or CAWI days or weeks before the PCR testing.
- Red Cross Questionnaire: This questionnaire about current symptoms was filled in when the swab sample was taken.
- "No-Show"-Questionnaire: This questionnaire was surveyed online in CV03. It was targeted to sampled people who had already set a date for PCR testing within the online questionnaire, but did not keep this appointment.

Main contents and slight changes of the main questionnaire are summarized in Table 4. Changes in question formulations and filtering are not covered in Table 4. In the compilation of the questionnaires, internationally proven scales (WHO questionnaire for COVID-19, WHO-5 Well-Being Index, labour force survey questions on gainful employment and schooling) were used as much as possible. The stability of the questionnaire throughout the three studies was also a major concern, but some improvements were still incorporated based on the experience gained.

3.3. PCR testing

PCR tests were carried out on individuals who had submitted the questionnaire and consented to swab sampling by the Red Cross. The swab samples for the PCR tests were taken nationwide by specifically trained Austrian Red Cross health personnel. Time intervals of PCR tests are available in table 3.

The PCR analyses from the swab specimens taken were conducted on the fully automatic Roche cobas® 6800 Test System using the Roche cobas® SARS-CoV-2 Test (CE/IVD).

Table 4: Questionnaire CV01, CV02, CV03

Survey items	CV01	CV02	CV03
<i>Socio-demographic variables</i>			
Informed consent	X	X	X
Gender	X	X	X
Age	X	X	X
Nationality	X	X	X
Country of birth	X	X	X
Educational attainment	X	X	R
Household composition	X	X	X
<i>Work situation</i>			
Activity status before pandemic	X	X	X
Activity status after pandemic	X	X	X
Workplace	X	X	D
Changes in working hours	X	X	X
Compatibility of work and family	N	X	X
<i>Health factors</i>			
General health status	X	X	X
Well-being	X	X	X
Health risk factors	X	E	X
Chronic illness	X	E	X
Potential SARS-CoV-2 symptoms	X	X	D
SARS-CoV-2 diagnosis	X	X	D
Contact with infected people	X	X	N
SARS-CoV-2 disease process	N	N	X
<i>Subjective assessment</i>			
Social contacts	X	X	N
Subjective risk perception	X	E	X
Subjective assessment of governmental measures	X	D	D

(X) - included

(N) - not included

(D) - included, with different items or in different structure to previous survey

(E) - included, with an extended list of items to previous survey

(R) - included, with a reduced list of items to previous survey

The detection of two target genes (so-called dual target PCR; target regions: ORF1 for SARS-CoV-2; E gene for pan-sarbecoviruses) as well as the simultaneous conduction of an internal control ensured maximum sensitivity (detection limit of 0.009 TCID₅₀* for SARS-CoV-2 and 0.003 TCID₅₀* for pan-sarbecoviruses) and specificity. It is impossible to clearly assess and to quantify potential preanalytical factors with a negative subsequent effect on the PCR test result (e.g. poor swab quality or swabbing at a disadvantageous time in the course of the disease – e.g. swabs done during the incubation period). The results of the analyses were transmitted in fully anonymized form to Statistics Austria. If a test was positive, the procedure set down in law for reporting notifiable diseases was followed by the Medical University of Vienna.

3.4. Response rate

The response rate of the studies varied, see table 5. In CV01, 56% of sampled people participated in the study, in CV02 44% and in CV03 35%.

The limited test capacities were a main driver for high response rates at the beginning of the pandemic and the confinement during the lockdown made contacting people easier. In the last survey CV03, hopes that the included antibody test would help to bring the response rate back up again did not come true.

Table 5: Response rates

Survey ID	Gross size	Net size Q	Net size PCR	Response rate
CV01	2,800	1,577	1,432	56%
CV02	3,724	1,528	1,279	44%
CV03	7,823	2,711	2,263	35%

(Netto size Q) - Netto size of people finishing the questionnaire

(Netto size PCR) - Netto size of people receiving a valid PCR test

The response rate did not only vary between surveys, but also varied strongly between groups: People whose highest level of education was compulsory school had far lower response rates, compared to those of higher education. This effect had already been anticipated and therefore taken into account in the sampling design. Also between the federal states, response rates varied widely, see table 6.

Table 6: Response rates by federal state.

State	CV01	CV02	CV03
Burgenland	63 %	46 %	42 %
Carinthia	65 %	50 %	35 %
Lower Austria	61 %	49 %	38 %
Upper Austria	60 %	43 %	36 %
Salzburg	53 %	43 %	36 %
Styria	63 %	39 %	36 %
Tyrol	53 %	39 %	31 %
Vorarlberg	50 %	35 %	36 %
Vienna	45 %	35 %	30 %

4. Inference

4.1. Weighting

In all three studies, calibration was done using an iterative proportional fitting procedure². As participation varied between filling in the questionnaire and PCR testing, weights were calculated for each of these subsets.

In CV01 and CV02 no detailed modeling of non-response was done: Since many of the variables highly correlated with non-response were later used in the calibration procedure. Response rates were clearly lower in CV03. Therefore, a more sophisticated non response modeling was implemented in CV03.

Table 7 summarizes weighting procedures and slight adaptations of calibration variables used between all three studies.

Table 7: Weighting procedure

Survey ID	Non-response-correction	Calibration variables
CV01	Adaption within strata: Part A: federal state, risk category, education Part B: education	Age class (6) x Gender (2) x Degree of urbanization (3) Household size (4) x Degree of urbanization (3) State (9) x Degree of urbanization (3) Risk category (3) x Degree of urbanization (3) Nationality (2) x Degree of urbanization (3) Education (2) x Degree of urbanization (3)
CV02	analogous to CV01	analogous to CV01; additional: underlying condition (2) x degree of urbanization (3) x gender (2)
CV03	Non-Response-Modeling of Participation in Questionnaire: Age Class + Nationality + Education + Quintil of household income + EMS information Non-Response-Modeling of Participation in PCR-Test: Age Class + Nationality + Education + Quintil of household income + Relative number of Infected in communal + EMS information	analogous to CV02; additional: surveyed educational attainment (4) additional: EMS information - people tested positive until 12 November (2) x Nuts 1 (3)

4.2. Error calculation

In CV01 and CV03 sampling errors and confidence intervals for the prevalence were estimated using a bootstrap procedure (rescaled bootstrap for stratified multistage sampling (Preston 2009), R functions `draw_bootstrap` and `recalib` from R package `surveysd` (Gussenbauer *et al.* 2020; Gussenbauer and de Cillia 2021)). We drew 5,000 bootstrap samples which were calibrated to the same parameters as the original sample.

The 95% confidence intervals are calculated with the percentile method (Efron 1981) as a 2.5% percentile for the lower limit and a 97.5% percentile for the upper limit of the 5,000 bootstrap samples of the estimator (Gray, Haslett, and Kuzmicich 2004). The standard deviation of

²The R function `ipf` from R package `surveysd` (Gussenbauer, Kowarik, and de Cillia 2020) was used. The extreme values of the weights were given upper and lower limits (1,200 and 18,000, respectively) and the relative distance to the design weight (frequently referred to as g weights) was limited by the factor 4

the 5,000 bootstrap estimators can be used for estimating sampling errors. The coverage probability of this bootstrap procedure might be overestimated for very small proportions; thus, given a nominal coverage of 95%, simulation procedures estimate that for proportions less than or equal to 0.27% (about double of the upper confidence limit of CV01), only between 90-95% coverage probability is achieved. Provided that the complex sampling design is taken into consideration, similar estimates are possible with the Clopper-Pearson confidence interval (Clopper and Pearson 1934), which is frequently used for small proportions.

Since bootstrapping does not work for dichotomous characteristics if a category does not occur in a sample, a Bayesian interval was calculated for prevalence in the survey in CV02.

Errors for other results based on the questionnaire were calculated with bootstrap intervals in all three studies.

5. Results

5.1. How many people in Austria were tested positive for COVID-19 in CV01, CV02 and CV03?

Table 8 shows point estimates, and 95% confidence intervals of people tested positive in all three studies.

Statistics Austria decided to publicly present upper confidence interval limits, rather than point estimates of the share of PCR-positive people. This decision was made content-based: In terms of the virus, it was most important to find out the maximum number of infected individuals.

In CV01, PCR tests were successfully carried out on 1,432 sampled people between 21 and 24 April 2020. One person tested positive, i.e. was infected with SARS-CoV-2 at the time of testing. The weights of the 1 432 individuals were calibrated in order to calculate an estimate for the whole of Austria. The estimation for the number of infected cases in April 2020 is 3,420 people, or 0.05% of the population. The confidence interval ranges from 72 to 10,823 people, or 0.001% to 0.148%. This confidence interval was calculated using the bootstrap procedure described above.

In CV02 and therefore in the period from 26 to 30 May 2020, PCR tests were successfully carried out on 1,279 sampled people. None of these individuals tested positive. This preliminary final result essentially confirms the trend reported by the Epidemiological Reporting System (EMS): While around 960 people were infected at the time of CV01 according to the EMS (survey period ± 5 days), only around 380 people were infected when CV02 was conducted in May. Given the underlying sample size, figures of this kind can no longer be reliably estimated. If the results are compared with the previous sample studies, it can be seen that with approximately the same effective sample size, the number of people infected in the sample fell from six (SORA, early April) to one in CV01 and to zero in CV02 (Statistics Austria, late May). It can therefore be assumed that prevalence decreased between each of the surveys, even if no significant evidence of this can be provided by the existing sample data. Table 8 shows the Bayesian-Interval with the a-priori assumption: 0.67 positive and 1,431.33 negative cases corresponding to the weighted proportion of the one positive case in CV01).

Specifically, three different variants were calculated:

1. The non-informative case (a priori assumption: 0.5 positive and 0.5 negative cases), meaning that no prior information is used here.
2. The results of the April survey are used (a priori assumption: 0.67 positive and 1,431.33 negative cases corresponding to the weighted proportion of the one positive case in the sample in the survey conducted in late April 2020).

- The value of the positive case in the April survey is reduced in parallel with the trend in reported new infections in the EMS. The cases counted are the new cases in the observation period ± 5 days (16 to 29 April: around 960 infections were reported; 22 May to 4 June: around 380 infections were reported; a priori assumption: 0.26 positive and 1,431.74 negative cases).

Depending on the scenario, various limits are obtained:

- The non-informative case. No prior information is used here to calculate the confidence interval. The result is an upper limit of around 11,000 people.
- If information based on the survey in late April 2020 is incorporated in a Bayesian confidence interval estimate as an a priori assumption, this results in an upper limit of around 6,000 people for the confidence interval.
- If the decrease in the number of people infected according to the EMS data is also taken into account, this results in an upper limit of around 3,000 people for the Bayesian confidence interval.

In CV03, PCR tests were successfully carried out on 2,263 sampled people between 12 and 14 November 2020. 48 of these people were tested positive. In CV03, Statistics Austria for the first time had access to pseudonymized EMS data of sampled people who had confirmed to participate in the study. This data revealed that 24 people who had already confirmed their participation, but did not show up to the PCR test, were in governmental quarantine because of a positive PCR test at the dates of testing. These official information was mostly in accordance with subjective information of respondents given in a "no-show-survey". Grossed up the number of people tested positive between 12 and 14 November was 233,000 people. This equals a proportion of 3.1% of the population living in Austria aged 16 or older. The confidence interval ranges from 195,000 to 261,000 (2.6 - 3.5%). The new infections reported to the EMS around the survey period (7 - 11 November 2020) were 86,000, so this is again a lower boundary of officially registered cases that should also test positive in our survey.

Table 8: PCR results

Survey	Net size PCR	n PCR positive	Point estimate	2.5%-CI	97.5%-CI
CV01	1,432	1	3,420	71	10,823
CV02	1,279	0	1,814	0	6,276 **
CV03	2,263	48 + 24*	233,000	195,000	261,000

(Netto size PCR) - Netto size of people receiving a valid PCR test

(n PCR positive) - Number of people tested positive in the sample

* - positive according to EMS

** - Bayes scenario 2

The results reflect the trend of official numbers in Austria, but at higher level. Figure 3 shows confidence intervals of estimated prevalence in the three surveys.

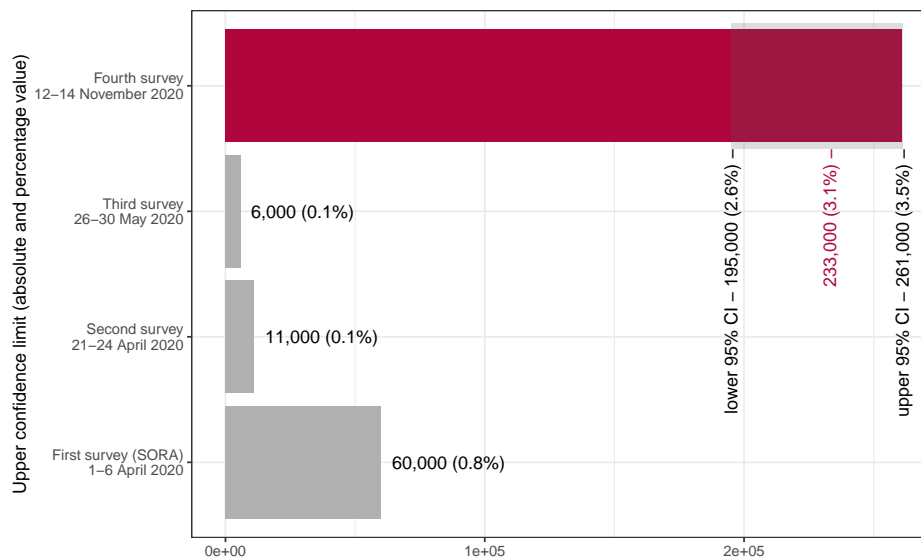


Figure 3: Acute SARS-CoV-2 infections

In CV03 information from the EMS could be used to estimate the number of unknown cases: Less than 50% of the revealed SARS-CoV-2-infections (108 000) were already officially registered. Therefore, estimated 53% of all acute infections, remained undetected by officials. The majority of the group of people with infections not discovered by the government (26 from 37, 70%) declared suffering from none or only one symptom. Also, only a minority (5 of 37, 14%) expected a positive test result. In contrast, only a minority of the group of people with an infection not discovered by the government (6 of 35, 17%) had no symptoms. In general, infections in the group of individuals with undiscovered infections are mostly asymptomatic. However, it is unclear whether these people were presymptomatic and developed symptoms afterwards.

5.2. What is the general opinion about the protective measures taken by the Austrian government?

In this report only one section of the questionnaire is presented: the acceptance of protective measures. More detailed results of the questionnaire are available on [Statistic Austria's website](#) (in German).

The COVID-19 pandemic has had a significant impact on many people's lives. The government introduced legislation on COVID-19 measures at the start of the pandemic, placing some areas under lockdown and universally establishing new social norms such as physical distancing and wearing face masks. We investigated how these measures are perceived by the population and what may affect their acceptance. Respondents were asked about their acceptance of governmental measures in all three studies. As governmental measures were being continuously adjusted to the COVID-19 situation, the measures in place varied from survey to survey. This implies a change in questioning on this content – as already outlined in table 4. In CV01 and CV02 the acceptance of a list of measures was surveyed, in CV03 the participants were just asked for their opinion on the measures in general.

While around three quarters of the measures were accepted by more than 85% of the population in CV01 (figure 4), only a quarter of the partly new measures compared to those in CV01 found broad acceptance of this kind in May 2020 (figure 5): the protection of identified vulnerable groups at work (97%), physical distancing (94%) and wearing a face mask (86%). Measures less likely to be considered appropriate by those surveyed in May 2020 included, most notably, the ban on mass gatherings/events with 10 or more people (62%), home learn-

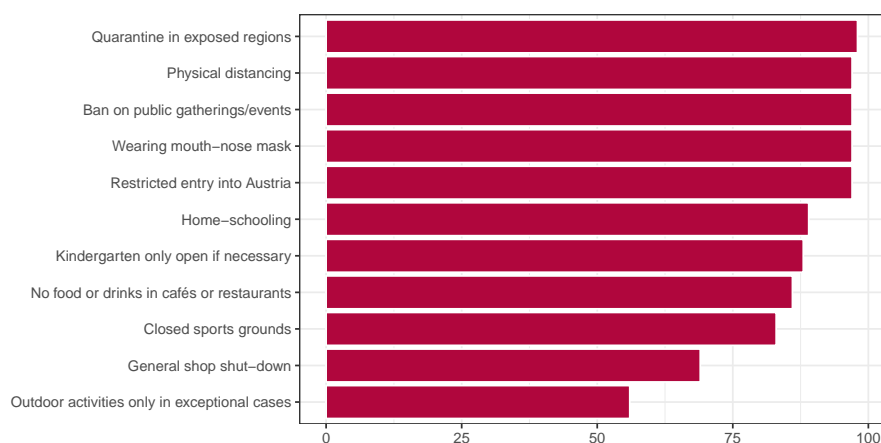


Figure 4: CV01: What do people think in general about the measures – percentage considering measure as appropriate

ing for upper secondary school pupils (60%) and the closure of swimming pools and indoor sports facilities, which had reversed since (59%).

In addition to the general acceptance for the measures in CV02, figure 5 exemplarily presents the acceptance by age group and the 95% confidence interval for these estimators. The confidence intervals were computed using the bootstrap procedure described in 4.2. The measures restricted childcare, alternating home/school learning and home learning have a significant lower acceptance in the group of 35 to 44 year old people compared to other age groups. Furthermore, a difference between young and old is visible in the acceptance of the restrictions in restaurants, older groups showed a higher acceptance than the youngest two age groups. In CV01 the closure of sports grounds had a significant higher acceptance in the older age groups compared to people below 35 years. As in CV02 there are more differences between age groups, but due to the sampling error most of them are not significant. In both CV01 and CV02 the protection of vulnerable groups, physical distancing and wearing a face mask was accepted by all age groups up to a similar extent.

In CV03 respondents were only asked to judge their overall acceptance with measures in five categories: clearly overstated, overstated, appropriate, understated and clearly understated. 81% of all respondents judged measures as appropriate over even (clearly) understated. In this period (Oct. 13 to Oct. 28) Austria had a very high number of official SARS-CoV-2-cases – and one of the worst incidences worldwide –, but there was no lockdown. The situation however was assessed differently by the youngest age group surveyed – people aged from 16 to 24. More than every fourth of this group (26%) agreed that measures were (clearly) overstated.

6. Conclusion

This study provides a valid assessment of SARS-CoV-2 infections in Austria for three time periods in 2020: April 2020, May 2020 and November 2020.

Samples of individuals were tested for acute SARS-CoV-2 infections. Grossed up figures of acute infections reflected the trend of official figures, but at a higher level. This study proves that more than 50% of all infections remained undiscovered in Austria in November 2020. Undiscovered infections showed to be mostly asymptomatic. The high rate of undiscovered cases is severe, as the virus may be spread by asymptomatic individuals as well.

This study has several strengths, as e.g. sampling from the ZMR and timeliness. Due to the timely repetitions of the study, the study design could be slightly improved from survey to

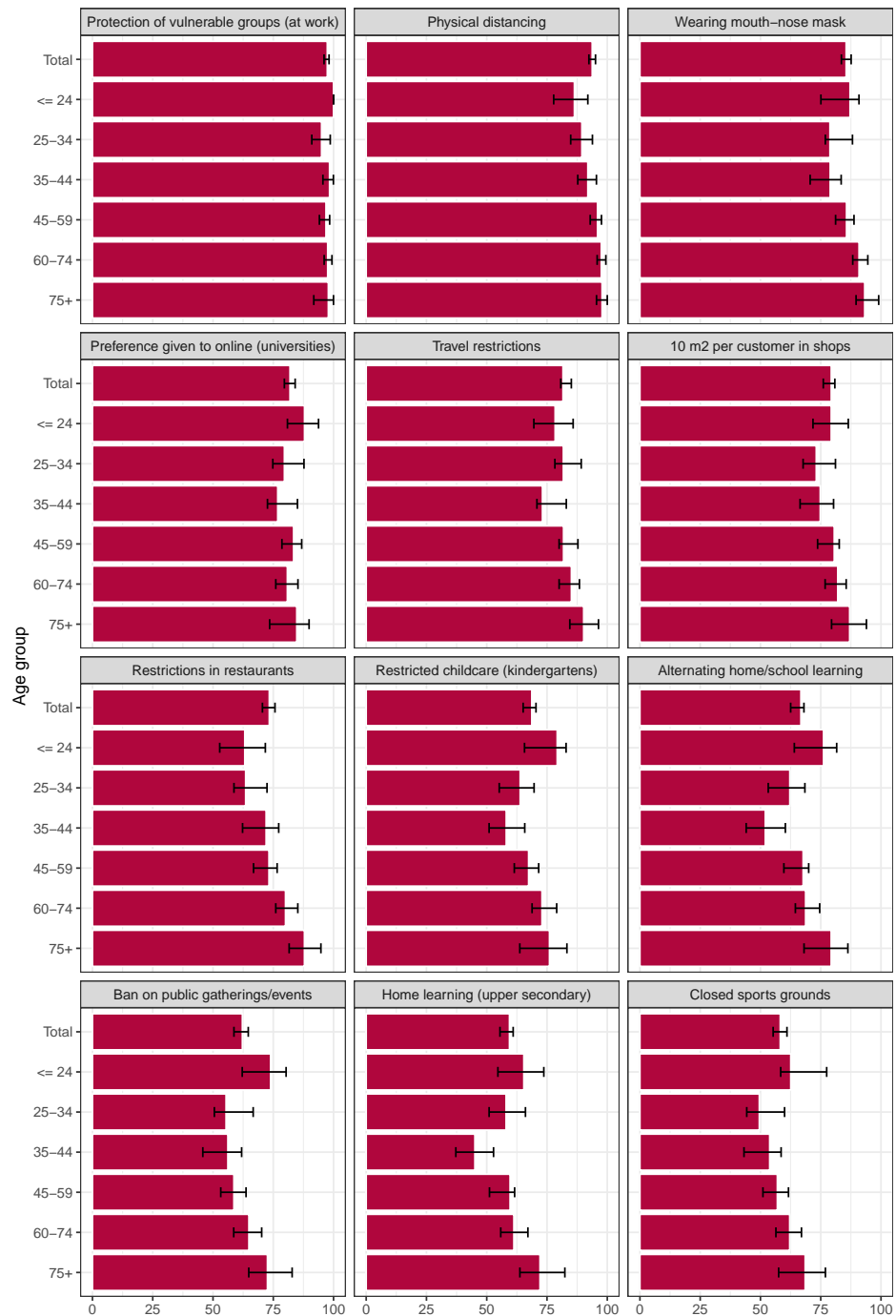


Figure 5: CV02: What do people think in general about the measures – percentage considering measure as appropriate by age group

survey. Especially the access to the EMS in CV03 tremendously improved the validity of the figures. However, several weaknesses have to be pointed out: The response rates declined from survey to survey. While high response rates in CV01 may be explained by the novelty of the pandemic, low response rates in later studies may be due to a certain restricted affect. Even though, in the CV03 design weights were corrected with non-response modeling, results may be biased by non-response. Bias could occur in both directions, e.g. people at high risk may be more likely to attend to be informed about their health status or in contrast may be less likely to attend for safety reasons.

Firstly, the SARS-CoV-2 virus is characterized by regional clustering – which might be missed in a grossed up sample. Secondly, specificity and sensitivity of tests were not taken into account in the calculations. Still, as outlined in section 3.3, maximum sensitivity was ensured. Nevertheless, this study underlines that broad testing is one way of detecting asymptomatic cases early and thus containing the spread of the virus. In Austria, shortly after the publication of CV03 regular mass testing was implemented: In these mass tests every individual has the opportunity to have a free antigen test done at a specific date. Because of the high rate of false positives in these antigen tests, a positive test result has to be followed up by a PCR test which is more accurate. Austrian schools follow this strategy too: On February 8 2021, after a long period of distance learning, schools reopened and pupils had to test themselves at the schools once a week. Also from February 8 2021 a negative antigen test was necessary for routines of daily live, as e.g. to visit the hairdresser or a museum. Therefore, this study design has contributed to governmental judgment and adaption of SARS-CoV-2 measures and is recommended for further monitoring of the SARS-CoV-2 situation.

Dataset notes

Data on the Prevalence Study of all three surveys are available free of charge from AUSSDA, The Austrian Social Science Data Archive. For further information see AUSSDA Dataverse (Paškvan, Kowarik, Schrittwieser, Till, Weinauer, Göllner, Hartleib, Klimont, Plate, Baumgartner, Edelhofer-Lielacher, Grasser, and Kytir 2020a,b).

The data sets include all questionnaire variables and the symptoms specified on the occasion of swab sample taking for the PCR test. In addition, another data set is available for the bootstrap weights. For CV01, CV02 only the information if a PCR test was performed is available, but for CV01 the single person with a positive PCR result is not disclosed. In the CV03 data set both test results (PCR and antibody) are included. For the first survey only Statistics Austria's Safe Center is available for analysis of PCR results.

Data can be analyzed with the **surveysd** package (Gussenbauer *et al.* 2020; Gussenbauer and de Cillia 2021) to compute confidence intervals. The **surveysd** package is used to compute bootstrap intervals and standard errors for complex survey designs. Coding of variables are saved as SPSS-labels, and can also be looked up in the subjoining codebook.

A short example for CV01-AUSSDA to compute the share of people being in favor of the measurement "physical distancing", is given in the following:

Read the Data:

```
> library(data.table)
> library(surveysd)
> library(haven)
> # path <- "" # define the directory of files
>
> cv01_survey_data <- data.table(read_sav(
+ file.path(path,
+ "Daten_COVID-19_Prävalenz_April2020.sav"))
+ )
```

```

> cv01_bootstrap_data <- data.table(read_sav(
+ file.path(path,
+ "Schichtung-Bootstrap_COVID-19_Prävalenz_April2020.sav"))
+ )
> cv01_dat <- merge(cv01_survey_data ,
+                 cv01_bootstrap_data,
+                 by = "lfnr"
+ )

```

Explore the variable of interest (variable name taken from the codebook, available on AUSSDA):

```

> table(cv01_dat$subjpolitMass1)

```

```

  1    2
1531  46

```

Equip data with attributes for the usage of `surveysd`:

```

> n_b <- 4989 #number of bootstrap weights
> setattr( cv01_dat, "weights", "gew1")
> setattr( cv01_dat, "b.rep", paste0("w", 1:n_b))

```

Compute Estimates, first for the whole population, then by sex:

```

> # whole population
> (Estimates <- calc.stError(cv01_dat,
+                           var = "subjpolitMass1",
+                           fun = weightedRatio,
+                           p = c(.025,.975)
+ )$Estimates[,-1])

```

```

      n      N val_subjpolitMass1 stE_subjpolitMass1
1: 1577 7323486          97.21483      0.4033959
  p0.025_subjpolitMass1 p0.975_subjpolitMass1
1:          96.40382          97.97164

```

```

> # by group
> (Estimates <- calc.stError(cv01_dat,
+                           var = "subjpolitMass1",
+                           fun = weightedRatio,
+                           group = list("bsex"),
+                           p = c(.025,.975)
+ )$Estimates[,-1])

```

```

      n      N bsex val_subjpolitMass1 stE_subjpolitMass1
1:  762 3573281    1          96.69265      0.6645776
2:  815 3750205    2          97.71237      0.4871973
3: 1577 7323486   NA          97.21483      0.4033959
  p0.025_subjpolitMass1 p0.975_subjpolitMass1
1:          95.35183          97.96036
2:          96.72025          98.63816
3:          96.40382          97.97164

```

```

>

```


n indicates the number of observation, N the weighted number of observations, val_variablename the weighted share, stE_variablename the weighted standard error, $\text{p0.025_variablename}$ the lower confidence bound, $\text{p0.975_variablename}$ the upper confidence bound.

7. Acknowledgement

We gratefully acknowledge the great cooperation with the Red Cross and the Medical University of Vienna. Therefore, we thank especially Lukas Weseslindtner and Robert Strassl from the Medical University of Vienna. Additionally, several colleagues from Statistics Austria played a crucial role in this project and provided insights used in this work: Sarah Hartleib, Jeannette Klimont, Marc Plate, Irene Baumgartner, Edith Edelhofer-Lielacher, Alfred Grasser, Josef Kytir and Franziska Mayrhofer.

References

- Bivand RS, Pebesma E, Gomez-Rubio V (2013). *Applied Spatial Data Analysis with R, Second edition*. Springer, NY. URL <https://asdar-book.org/>.
- Cheng J, Karambelkar B, Xie Y (2019). *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 2.0.3, URL <https://CRAN.R-project.org/package=leaflet>.
- Clopper CJ, Pearson ES (1934). “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial.” *Biometrika*, pp. 404–413.
- Efron B (1981). “Nonparametric Standard Errors and Confidence Intervals.” *canadian Journal of Statistics*, **9**(2), 139–158.
- Gray A, Haslett S, Kuzmich G (2004). “Confidence Intervals for Proportions Estimated from Complex Sample Designs.” *Journal of Official Statistics*, **20**(4), 705.
- Gussenbauer J, de Cillia G (2021). “The R-package *surveysd*: Estimating Standard Errors for Complex Surveys with a Rotating Panel Design.” *Statistical Journal of the IAOS*, **37**(1), 115–121. doi:10.3233/SJI-200709.
- Gussenbauer J, Kowarik A, de Cillia G (2020). *surveysd: Survey Standard Error Estimation for Cumulated Estimates and their Differences in Complex Panel Designs*. R package version 1.3.0, URL <https://CRAN.R-project.org/package=surveysd>.
- Knabl L, Mitra T, Kimpel J, Roessler A, Volland A, Walser A, Ulmer H, Pipperger L, Binder SC, Riepler L, et al. (2021). “High SARS-CoV-2 Seroprevalence in Children and Adults in the Austrian Ski Resort of Ischgl.” *Communications Medicine*, **1**(1), 1–9.
- Ladage D, Höglinger Y, Ladage D, Adler C, Yalcin I, Harzer O, Braun RJ (2021). “SARS-CoV-2-specific Antibody Prevalence and Symptoms in a Local Austrian Population.” *Frontiers in Medicine*, **8**.
- Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung JK, Katsuma Hayashi S, Kinoshita R, Yang Y, Yuan B, Akhmetzhanov AR, Linton NM (2020). “Estimation of the Asymptomatic Ratio of Novel Coronavirus Infections (COVID-19).” *International Journal of Infectious Diseases*, **94**, 154.
- Ogris G, Oberhuber F (2020). *Spread of SARS-CoV-2 in Austria*. URL https://www.sora.at/fileadmin/downloads/projekte/Austria_Spread_of_SARS-CoV-2_Study_Report.pdf.

- Paškvan M, Kowarik A, Schrittwieser K, Till M, Weinauer M, Göllner T, Hartleib S, Klimont J, Plate M, Baumgartner I, Edelhofer-Lielacher E, Grasser A, Kytir J (2020a). “COVID-19 Prevalence April 2020 (SUF edition).” doi:10.11587/FTDGA4. URL <https://doi.org/10.11587/FTDGA4>.
- Paškvan M, Kowarik A, Schrittwieser K, Till M, Weinauer M, Göllner T, Hartleib S, Klimont J, Plate M, Baumgartner I, Edelhofer-Lielacher E, Grasser A, Kytir J (2020b). “COVID-19 Prevalence May 2020 (SUF edition).” doi:10.11587/8GP09W. URL <https://doi.org/10.11587/8GP09W>.
- Preston J (2009). “Rescaled Bootstrap for Stratified Multistage Sampling.” *Survey Methodology*, **35**(2), 227–234.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Statistics Austria (2020). *Online Information COVID-19 Prevalence*. URL http://www.statistik.at/web_de/frageboegen/private_haushalte/covid19/index.html.
- Zifko U, Schmiedlechner T, Saelens J, Zifko K, Wagner M, Assadian O, Grisold W, Stingl H (2021). “Covid-19: Involvement of the Nervous System. Identifying Neurological Predictors Defining the Course of the Disease.” *Journal of the Neurological Sciences*, **425**, 117438. ISSN 0022-510X. doi:<https://doi.org/10.1016/j.jns.2021.117438>. URL <https://www.sciencedirect.com/science/article/pii/S0022510X21001325>.

Affiliation:

Alexander Kowarik
Methods Unit
Statistics Austria
Guglgasse 13
A-1110 Vienna, Austria
E-mail: alexander.kowarik@statistik.gv.at
URL: <https://www.statistik.gv.at>