

**Title:** Seroprevalence of SARS-CoV-2 IgG antibodies in two regions of Estonia (KoroSero-EST-1)

**Authors:** Piia Jõgi<sup>1,2</sup>, Hiie Soeorg<sup>3</sup>, Diana Ingerainen<sup>4</sup>, Mari Soots<sup>5</sup>, Freddy Lättekivi<sup>6,7</sup>, Paul Naaber<sup>3,8</sup>, Karolin Toompere<sup>9</sup>, Pärt Peterson<sup>10</sup>, Liis Haljasmägi<sup>10</sup>, Eva Žusinaite<sup>11</sup>, Hannes Vaas<sup>1</sup>, Merit Pauskar<sup>3</sup>, Arina Shablinskaja<sup>3</sup>, Katrin Kaarna<sup>7,12</sup>, Heli Paluste<sup>13</sup>, Kai Kisand<sup>10</sup>, Marje Oona<sup>14</sup>, Riina Janno<sup>7</sup>, Irja Lutsar<sup>3</sup>

### **Affiliations**

1. Children's Clinic of Tartu University Hospital, N. Lunini 6, 50406, Tartu, Estonia
2. Department of Pediatrics, Institute of Clinical Medicine, University of Tartu, L. Puusepa 8, 50406, Tartu, Estonia
3. Department of Microbiology, Institute of Biomedicine and Translational Medicine, University of Tartu, Ravila 19, 50411, Tartu, Estonia
4. Family doctor center "Järveotsa", Õismäe tee 179, 13517, Tallinn, Estonia
5. Family doctor center „Kuressaare“, Aia 25a, 93815, Kuressaare, Estonia
6. Department of Pathophysiology, Institute of Biomedicine and Translational Medicine, University of Tartu, Ravila 19, 50411, Tartu, Estonia
7. Development Centre, United Chancery Service, Tartu University Hospital, Puusepa 8, 50406, Tartu, Estonia
8. SYNLAB Estonia, Tallinn, Estonia
9. Department of Epidemiology and Biostatistics, Institute of Family Medicine and Public Health, University of Tartu, Ravila 19, 50411, Tartu, Estonia
10. Molecular Pathology, Institute of Biomedicine and Translational Medicine, University of Tartu, Ravila 19, 50411, Tartu, Estonia

**NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

11. Institute of Technology, Faculty of Science and Technology, University of Tartu, Nooruse  
1, 50411, Tartu, Estonia

12. Clinical Research Centre, Institute of Clinical Medicine, University of Tartu, Ravila 19.  
50411, Tartu, Estonia

13. Ministry of Social Affairs, Republic of Estonia, Suur-Ameerika 1, 10122 Tallinn, Estonia

14. Department of Family Medicine, Institute of Family Medicine and Public Health,  
University of Tartu, Ravila 19, 50411, Tartu, Estonia

**Corresponding author:** Piia Jõgi

**Corresponding author email:** piia.jogi@kliinikum.ee

## **Abstract**

**Background:** In Estonia, during the first wave of COVID-19 total number of cases confirmed by PCR was 13.3/10,000, similar in most regions, including capital Tallinn, but in the hotspot of Estonian epidemic, an island Saaremaa, the cumulative incidence was 166.1/10,000.

**Aim:** We aimed to determine the prevalence of SARS-CoV-2 IgG antibodies in these two regions, symptoms associated with infection and factors associated with antibody concentrations.

**Methods:** Participants were selected using stratified (formed by age decades) random sampling and recruited by general practitioners. IgG were determined from sera by four assays. Symptoms of acute respiratory illness associated with seropositivity were analyzed by multiple correspondence analysis, antibody concentrations by multiple linear regression.

**Results:** Total of 3608 individual were invited and 1960 recruited From May 8 to July 31, 2020. Seroprevalence was 1.5% (95% confidence interval (CI) 0.9-2.5) and 6.3% (95% CI 5.0-7.9),

infection fatality rate 0.1% (95% CI 0.0-0.2) and 1.3% (95% CI 0.4-2.1) in Tallinn and Saaremaa, respectively. Of seropositive subjects 19.2% (14/73) had acute respiratory illness. Fever, diarrhea and the absence of cough and runny nose were associated with seropositivity in individuals aged 50 or more years. IgG concentrations were higher if fever, difficulty breathing, shortness of breath, chest pain or diarrhea was present, or hospitalization required.

**Conclusion:** Similarly to other European countries the seroprevalence of SARS-CoV-2 in Estonia was low even in the hotspot region Saaremaa suggesting that majority of population is still susceptible to SARS-CoV-2. Focusing only on respiratory symptoms may delay accurate diagnosis of SARS-CoV-2 infection.

**Keywords:** COVID-19; SARS-CoV-2; Seroprevalence; symptoms; infection fatality rate; IgG

**Conflict of interest:** None declared

**Funding statement:** The study was funded by the Ministry of Social Affairs of the Republic of Estonia.

## Introduction

The first case of COVID-19 was reported in Estonia on February 26, 2020. Two weeks later extensive spread of SARS-CoV-2 virus occurred that led to lockdown, including closure of educational institutions at all levels, recommendations to stay and work at home whenever possible and limitations to public gatherings from March 12, 2020 [1]. By the end of April, the first wave of COVID-19 had been largely contained by the implemented measures that were subsequently lifted or partially relaxed on May 16, 2020. By then, the total cumulative number of cases confirmed by PCR for SARS-CoV-2 in Estonia was 13.3 per 10,000 inhabitants, similar in most of the regions, including the capital Tallinn located in Harju county [2]. In contrast, in the hotspot of Estonian epidemic, an island Saaremaa, the cumulative incidence was 166.1 cases per 10,000 and with its peak incidence of 954.5 cases per 100,000 within last 14 days it was among the European within-country regions experiencing the most extensive spread [1]. However, as the infection-to-case ratio, i.e. the ratio of the seroprevalence to rate of confirmed cases of COVID-19, can vary widely from 6 to 23 times [3], the actual prevalence of infection with SARS-CoV-2 remains unknown in Estonia. In general, the rate of SARS-CoV-2 seropositivity in previously conducted studies have been low, ranging from 0.4 to 10.9% in population-based seroepidemiological studies [4-8]. To accurately estimate the extent of past spread and potential for future spread, seroepidemiological studies are warranted.

In KoroSero-EST-1 study, we aimed to understand the actual prevalence of the infection by determining the seroprevalence of COVID-19 in individuals from two regions in Estonia with very different cumulative incidence of COVID-19. Second, we aimed to describe the symptoms associated with COVID-19 in seropositive individuals and the factors associated with antibody concentrations against SARS-CoV-2.

## **Methods**

### **Study design**

The KoroSero-EST-1 was a cross-sectional seroepidemiological study conducted from May 8 to July 31, 2020 in two general practitioner (GP) practices in capital Tallinn in Harju county and in Saaremaa with total number of patients 13,260 and 7,525, respectively.

Participants were selected using stratified random sampling. Strata were formed by classifying patients from each GP into 10-year age groups, except the age group of 80 years or older due to their small number. From each stratum individuals were randomly identified by Estonian Health Insurance Fund with the aim to include at least 110 participants per age group from both GP practices to achieve desirable precision for the seroprevalence estimates.

Participants filled in a questionnaire based on World Health Organization recommendations [9] that included the presence of acute respiratory illness since March 1, 2020 and if present, its symptoms, known contact with PCR-confirmed COVID-19 cases, previously performed PCR test for SARS-CoV-2 and its result (Supplementary Information).

All participants and/or their legal guardians provided written informed consent. The study was approved by the Research Ethics Committee of the University of Tartu.

### **Sampling and antibody measurements**

Blood samples (3.5 mL) were drawn and stored for 48 hours at +4°C until transported to laboratory, where sera were stored at -30° C until testing in SYNLAB Estonia Central Laboratory in Tallinn or at the research laboratories of the University of Tartu, Estonia.

First, all samples were tested by chemiluminescent microparticle immunoassay for detection of IgG against SARS-CoV-2 nucleoprotein (N) (Abbott Architect SARS-CoV-2 IgG with ARCHITECT i2000SR analyzer; Abbott Laboratories, USA) according to the manufacturer's instructions. Samples close to cut-off value (signal/cut-off ratio 0.3 to 1.39) were additionally

tested by chemiluminescence immunoassay for detection of IgG against S1 and S2 proteins (LIAISON® SARS-CoV-2 S1/S2 IgG test; DiaSorin S.p.A., Italy) to increase sensitivity without influencing specificity [10].

Next, sera positive for anti-N or anti-S1/S2 IgG or those from participants with previously positive PCR test and two age- and sex-matched negative controls or a random subset of them (Figure 1) were tested for anti-N IgG by a rapid Standard Q COVID-19 IgM/IgG Duo test (SD BIOSENSOR, Republic of Korea) according to the manufacturer's instructions and by in-house luciferase immunoprecipitation system (LIPS) assays as described elsewhere [11].

Finally, an in-house neutralization assay was performed in duplicates for all PCR-positive samples and samples positive for anti-N or anti-S1/S2 IgG. Briefly, patients' sera were two-fold serially diluted in Virus Growth Media (VGM – DMEM/0.2% BSA/Pen-Strep) starting from 1:4 to 1:4096 in a volume of 50 µl per well. The viral stock was diluted in VGM to contain 100 pfu (plaque forming units) per 50 µl. 100 plaque forming units of the virus stock per well (a local Estonian SARS-CoV-2 isolate from a COVID-19-positive patient's nasal swab; from SYNLAB) was added to the wells containing sera dilutions, and the mixtures were incubated for 1 hour at 37°C. After this,  $4 \times 10^4$  Vero E6 cell suspension in 100 µl of VGM per well was added to the wells. Plates were incubated at 37°C and 5% CO<sub>2</sub> for up to 5 days. Neutralization titer was calculated by microscopically evaluating cytopathic effect (CPE) in infected wells (round/floating cells compared to untouched monolayer in non-infected wells). The dilution of sera that did not show CPE was considered as neutralization titer.

### **Statistical analysis**

The software program Stata 14.2 was used for calculating seroprevalence estimates and infection fatality rates and R (version 3.6.1) for other statistical analyses.

Total cumulative number of confirmed COVID-19 cases per 10,000 inhabitants was calculated as of two weeks before the midpoint of sample collection in GP practice. Daily numbers of confirmed COVID-19 cases (positive PCR tests for SARS-CoV-2 from nasopharyngeal swabs) in each age group in Harju county (assumed to reflect the spread of SARS-CoV-2 in Tallinn; 72.3% of Harju county inhabitants live in Tallinn) and Saaremaa were obtained from Health Board [2] and the number of inhabitants as of January 1, 2020 from Statistics Estonia.

### **Sample size calculation**

Sample size of at least 110 participants per stratum was needed to obtain binomial exact 95% CI with width of 5%, assuming seroprevalence was 1% or to estimate seroprevalence with precision  $\pm 10\%$ , assuming seroprevalence was more than 10%. Assuming non-response of about 50% (up to 70% in children), 2341 individuals were sampled from both GP practice to whom invitations were sent until about 110 per stratum agreed.

### **Seroprevalence estimation**

Seroprevalence was estimated as the proportion of participants considered seropositive, i.e. with positive result in at least two different antibody assay out of four assays performed (Figure 1), overall and in each age group, separately for the GP practices. To account for stratification, for each age group in a GP practice sampling weight was used that was further adjusted for response probability by post-stratification. Due to large sampling fractions, finite population correction factor was applied. Binomial exact 95% CI were calculated for seroprevalences in age groups.

Seroprevalence adjusted for test performance of anti-N IgG assay (Abbott Architect SARS-CoV-2 IgG with ARCHITECT i2000SR analyzer) was calculated as

$$\pi = \frac{q + s - 1}{r + s - 1},$$

where  $r$  is sensitivity,  $s$  specificity and  $q$  denotes prevalence [12]. Nonparametric bootstrap confidence intervals were calculated for adjusted seroprevalence estimates by drawing 10,000 bootstrap samples at random from our dataset and binomial distribution of the assay validation dataset. Each bootstrap sample was adjusted for sensitivity of 92.7% and specificity of 99.9% [13] and 2.5th and 97.5th percentiles of the bootstrap distributions were reported. Infection fatality rate was calculated as number of reported death cases related to COVID-19 as of July 31, 2020 divided by the estimated total number of seropositive individuals. Confidence intervals were calculated assuming that number of deaths follow Poisson distribution.

### **Analysis of symptoms**

The association between the presence of symptoms and seropositivity was analyzed by multiple correspondence analysis (MCA) using R package FactoMineR. Only dimensions that explained  $>1/Q$ , where  $Q$  is the number of variables used, of the total inertia were retained for interpretation. Association between seropositivity and symptoms was determined from dimensions with high contribution and quality ( $\cos^2 > 0.1$ ) of seropositivity. Hierarchical clustering using Euclidean distance and Ward's method followed by k-means consolidation was performed on all dimensions of MCA to partition the individuals into groups. The number of clusters was based on elbow of the barplot of the gain in within-inertia.

For comparing categorical variables, Fisher's exact test was used.

### **Analysis of antibody concentrations/titer**

Multiple linear regression was used to analyze the association between IgG and neutralizing (log-transformed and undetectable titers equaled to 0) antibodies concentrations/titer and the presence of acute respiratory illness or its symptoms or hospitalization adjusted for age, sex and time since onset of symptoms, shown to influence antibody concentrations [5, 14].



Multiple imputation was used to replace missing values of the dates of time of onset of symptoms with possible dates of infection in seropositive individuals without acute respiratory illness. Total of 1000 datasets were created where for each individual without acute respiratory illness the date was imputed from the dataset of confirmed COVID-19 cases [2] within the same county and in the inhabitants of the same age group detected until 14 days prior to drawing the individual's blood sample. Coefficients of multiple linear regression models and their standard errors were calculated by Rubin's rules [15].

## **Results**

### **Study population**

In total 3608 subjects were invited to participate of whom 1960 (54.3%) agreed and provided blood samples (Figure 1); all except one filled in the questionnaire. Of the participants 809 (41.3%) were male, 182 (9.3%) had been in contact with confirmed COVID-19 case, 221 (11.3%) had had PCR test for SARS-CoV-2 and 157 (8.0%) had had acute respiratory illness (Table 1). During the study period, the spread of SARS-CoV-2 had largely ended in Saaremaa, while in Tallinn it was ongoing (Figure S1).

### **Laboratory testing**

In total, 73 participants were seropositive. Of those all had detectable concentrations of anti-RBD by LIPS assay, 71 (97.3%) anti-N IgG by Abbott or anti-S1/S2 by DiaSorin assay, 68 (93.2%) anti-N IgG by LIPS assay, 60 (82.2%) neutralizing antibodies, 44 (60.3%) anti-N IgG by Biosensor assay (Figure 2). All 146 samples from seronegative age- and sex-matched controls were negative for anti-N IgG by Biosensor assay; of 22 samples tested by LIPS assays, one was positive for anti-N IgG.

## **Seroprevalence**

Overall seroprevalence was 1.5% (95% CI 0.9-2.5) in Tallinn and 6.3% (95% CI 5.0-7.9) in Saaremaa (Table 1). There were no difference in seroprevalence between males and females, age groups or household sizes. Seroprevalence was higher in the participants with contact with confirmed case or acute respiratory illness compared with those without. The seroprevalence estimates adjusted for test performance were similar to the estimates not adjusted for test performance, both overall (1.5% (95% CI 0.6-2.4) in Tallinn, 6.7% (95% CI 5-8.4) in Saaremaa) and in age groups (Table S1).

Number of deaths due to COVID-19 was 7 in Tallinn and 26 in Saaremaa and the respective infection fatality rates were 0.1% (95% CI 0.0-0.2) and 1.3% (95% CI 0.4-2.1).

### **Seroprevalence vs confirmed cases**

The cumulative number of confirmed COVID-19 cases in Tallinn (as of June 10, 2020) was 11.3 and in Saaremaa (as of June 15, 2020) 167.2 per 10,000 inhabitants. Seroprevalence of COVID-19 in Tallinn was 13.3 (range 8.0-22.1 based on 95% CI of seroprevalence estimate) and in Saaremaa 3.8 (3.0-4.7) times greater than the cumulative incidence of COVID-19 reported in the respective regions.

Of individuals with contact with confirmed case 43.0% (95/221) and of those with acute respiratory illness 33.1% (52/57) had PCR test performed. Considering participants without PCR test performed, in those with contact with confirmed case the proportion of seropositive individuals was 13.5% (17/126) and in those without contact 1.5 % (29/1612) ( $p < 0.001$ ); in those with acute respiratory illness the proportion of seropositive individuals was 6.7% (7/105) and in those without illness 2.4% (39/1633) ( $p = 0.018$ ).

## Symptoms

Of seropositive individuals 19.2% (14/73) and of seronegative individuals 7.6% (143/1886) self-reported having had acute respiratory illness ( $p=0.001$ ) with median (interquartile range) of 6 (5-7) and 5 (3-7) symptoms, respectively. The frequency of different symptoms was similar in seropositive and seronegative individuals with acute respiratory illness, but seropositive individuals required more often hospitalization due to the acute respiratory illness compared with seronegative individuals (Table 2).

In MCA of all individuals with acute respiratory illness, the first two dimensions discriminated poorly seropositive and seronegative subjects (Figure S2). Seropositivity had high quality on dimensions three and four and was associated with the absence of cough and the absence of sore throat (Figure S2). In MCA of 102 participants aged less than 50 years, of whom 5 were seropositive, it was not possible to associate seropositivity with any symptom by the first 6 dimensions due to low quality and contribution of seropositivity on dimensions 1-4 or symptoms on 5-6. In MCA of 55 participants aged 50 years or more, of whom 9 were seropositive, according to the first two dimensions of MCA seropositivity was associated with the absence of cough and runny nose and the presence of diarrhea and fever (Figure 3) and majority (7/9) of seropositive individuals were grouped into a single k-means cluster containing no seronegative individuals (Figure S3).

## Antibody concentrations/titer

According to multiple linear regression, anti-N IgG and anti-RBD IgG concentrations by LIPS assay and anti-N IgG by Abbott assay were significantly higher in seropositive individuals with fever (Table 3). Anti-RBD and anti-N IgG by LIPS assay were significantly higher in hospitalized patients, anti-N IgG by LIPS and/or Abbott assay in patients with difficulty breathing, shortness

of breath or chest pain and anti-N IgG by Abbott assay in patients with diarrhea. There were no statistically significant associations between titer of neutralizing antibodies and characteristics of acute respiratory illness.

## **Discussion**

To the best of our knowledge this is the first seroepidemiological study from Eastern Europe with strengths of including all age groups and using several assays for antibody testing. Our study showed that the seroprevalence of COVID-19 in two regions of Estonia with very different extent of the spread of SARS-CoV-2 by PCR in March and April 2020 were 1.5% and 6.3% with no difference between age groups. Most seropositive individuals had not had acute respiratory illness and only 4% were hospitalized. In individuals aged 50 years or more fever, diarrhea and the absence of cough and runny nose were associated with seropositivity. The concentrations of IgG, but not titer of neutralizing antibodies, were higher if fever, or difficulty breathing, shortness of breath, chest pain or diarrhea was present, or hospitalization was required.

## **Seroprevalence**

The seroprevalence estimates found in our study are within the range of 0.4-10.9% reported in other population-based seroepidemiological studies [4-8] and far from the threshold of 60 to 70% required for herd immunity [16]. Considering the number of cumulative cases per 10,000 inhabitants of 167.2, we would have expected larger seroprevalence in Saaremaa, as seroprevalence estimates were between 4.6-10.9% in Spain, United Kingdom and Geneva [4, 6, 7] with smaller number of cumulative cases per 10,000 (44.7-103.2) and thereby larger infection-to-case ratios 10.6-14.3 [1, 4]. However, these seroepidemiological studies were

conducted shortly after the peak of the spread when the daily incidence curve had flattened at low levels, but the spread was still ongoing [1, 4, 6, 7], suggesting the presence of undocumented cases and thereby underestimation of cumulative cases per 10,000 inhabitants. Similar pattern of ongoing spread during sample collection for seroepidemiological study and infection-to-case ratio 11.3 was in Tallinn. In contrast, in Saaremaa the spread of SARS-CoV-2 had been successfully contained and only few cases were detected during the study period, a situation similar to Reykjavik and Luxembourg, where the cumulative cases per 10,000 inhabitants were 50.4 and 53.3 and seroprevalence estimates 0.4% and 2.0%, respectively [5, 8]. Such discrepancies between infection-to-case ratios underline once again the importance of seroepidemiological studies in estimating the actual prevalence of infection. It is believed that incidence of SARS-CoV-2 infection was severely underreported in spring not only in Estonia but elsewhere [17]. This is further supported by the age distribution of confirmed cases with predominance of older age groups (46.1% and 68.4% aged 50 years or more in Tallinn and Saaremaa, respectively [2]) among reported cases while age distribution in our study was even in our seroepidemiological study.

## **Symptoms**

Of individuals with acute respiratory illness, two thirds had not been tested by PCR, but they were significantly more likely to be seropositive, showing underdetection of cases among them. Detection of cases by individual symptoms has low specificity [18, 19], particularly during the season of respiratory infections [18], but this could be increased by a combination of them [20]. Small number of individuals with acute respiratory illness impeded detection of pattern in our study. Although the proportion of asymptomatic cases among PCR-positive individuals has been up to 87.9% [21], we may have missed some symptomatic cases due to focusing only on acute respiratory illness. However, the proportion of COVID-19 cases

presenting without any respiratory symptoms, such as solely by fever or gastrointestinal symptoms, is low, up to 21% [22]. Furthermore, it is notable that while the proportion of fever, the most specific symptom of COVID-19 [18], in cases detected by PCR is larger than that of cough [18, 23], in seropositive individuals fever is mostly less common or occurs with rate comparable to cough [8, 24]. This may partly corroborate that the proportion of symptoms in seropositive cases attributable causes other than COVID-19 may be notable, up to 31.5% [25]. Additionally, this may suggest underdetection of cases among individuals with respiratory infections presenting without fever, supported by smaller proportion of people with cough being tested compared to those with fever [26], warranting focusing particularly on respiratory infections, as in our study. In adults aged 50 years or more, we found that the symptoms associated with seropositivity were the absence of runny nose and cough and the presence fever and diarrhea, similarly to a previous study [27], but the lack of information on anosmia or ageusia that are most strongly associated with COVID-19 [24, 26] may have compromised our analysis. Nevertheless, our results suggest that focusing only on respiratory symptoms may delay accurate diagnosis of SARS-CoV-2 infection in older people.

### **Antibody concentrations/titer**

In line with several other studies, we found that hospitalized patients or patients with symptoms suggesting more severe disease, such as fever, or shortness of breath had higher anti-N IgG concentrations [5, 14, 28]. However, these factors were not associated with titer of neutralizing antibodies or were less often associated with anti-RBD IgG that contain large proportion of neutralizing antibodies [29], in contrast with other studies [14, 30]. While these studies are mostly conducted on samples collected within two months after development of symptoms, long-term response of neutralizing antibodies is scarcely described [28]. Notably, in contrast to anti-RBD and anti-N IgG, neutralizing antibodies decline to undetectable titer

two months after the onset of symptoms in 20-25% of patients, as in our study, despite development of detectable peak in nearly all of them [28]. Moreover, titers of neutralizing antibodies decline faster compared with anti-RBD and anti-N IgG [14, 28]. Therefore, no associations found for neutralizing antibodies could be attributable to such faster decline and possible attainment of a plateau concentration where differences are less clear, as suggested [28], but further studies on longevity of neutralizing antibodies are warranted.

### **Limitations**

Our study has some limitations. First, as only 54% of individuals asked to participate in the study were willing to do so, participation bias may affect the results of our study. Second, as we did not sample from population register of Saaremaa and Tallinn, but from the list of patients of GP practices, the sampled individuals and thus seroprevalence estimates may not be representative of the regions and not directly comparable to confirmed cases. However, the age distribution of GP practice patients resembled that of the respective counties and testing rates in people near the GP practices was similar to the overall county [2]. Third, our testing algorithm for detection of seropositive individuals has unknown specificity and sensitivity and thus we cannot accurately estimate seroprevalences adjusted for test performance. Furthermore, we cannot rule out exclusion of false-negative individuals by testing only with combination of Abbott and DiaSorin assay, but such approach increases sensitivity by 7% [10].

In conclusion, seroprevalence of COVID-19 in Estonia in spring 2020 was low despite large number of confirmed cases, underlining the importance of seroepidemiological studies. Most people with SARS-CoV-2 infections had mild disease or did not report any respiratory symptoms. Improving detection of cases among individuals with contact with confirmed cases

or with acute respiratory illness or with fever and diarrhea in those aged 50 years or more may contribute to containment of the upsurge of the virus, but further long-term studies are warranted to demonstrate persistence of antibodies.

## Acknowledgements

We thank prof Krista Fischer for her advice in statistical analysis and all the study participants. The study was funded by the Ministry of Social Affairs of the Republic of Estonia. Electronic data collection was supported by the High Performance Computing Center of the University of Tartu.

## References

- [1] European Centre for Disease Prevention and Control COVID-19 data. Weekly data.; 2020. Available from: <https://www.ecdc.europa.eu/en/covid-19/situation-updates-covid-19/covid-19-data/weekly>. Accessed September 15, 2020.
- [2] Health Board. Data of Coronavirus SARS-CoV-2 test; 2020. Available from: <https://www.terviseamet.ee/et/koroonaviirus/avaandmed>. Accessed September 3, 2020.
- [3] Havers FP, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, et al. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. JAMA Intern Med 2020. Published online July 21, 2020.



- [4] Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet* 2020;396(10247):313-9.
- [5] Gudbjartsson DF, Norddahl GL, Melsted P, Gunnarsdottir K, Holm H, Eythorsson E, et al. Humoral Immune Response to SARS-CoV-2 in Iceland. *N Engl J Med* 2020. Published online September 1, 2020.
- [6] Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* 2020;396(10250):535-44.
- [7] Ward H, Atchison CJ, Whitaker M, Ainslie KEC, Elliott J, Okell LC, et al. Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. *medRxiv* 2020:2020.08.12.20173690. Posted August 21, 2020.
- [8] Snoeck CJ, Vaillant M, Abdelrahman T, Satagopam VP, Turner JD, Beaumont K, et al. Prevalence of SARS-CoV-2 infection in the Luxembourgish population: the CON-VINCE study. *medRxiv* 2020:2020.05.11.20092916. Posted May 18, 2020.
- [9] World Health Organization. Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-19) infection. 2020. Available from: <https://apps.who.int/iris/handle/10665/332188>. Accessed March, 17 2020.
- [10] Naaber P, Hunt K, Pesukova J, Haljasmägi L, Rumm P, Peterson P, et al. Evaluation of SARS-CoV-2 IgG antibody response in PCR positive patients: comparison of nine tests in relation with clinical data. *PLOS ONE* 2020. Accepted for publication.
- [11] Haljasmägi L, Remm A, Rumm AP, Krassohhina E, Sein H, Tamm A, et al. LIPS method for the detection of SARS-CoV-2 antibodies to spike and nucleocapsid proteins. *Eur J Immunol* 2020;50:1234-6.

- [12] Diggle PJ. Estimating Prevalence Using an Imperfect Test. *Epidemiology Research International* 2011;2011:608719.
- [13] Public Health England. Evaluation of sensitivity and specificity of four commercially available SARS-CoV-2 antibody immunoassays; 2020. Available from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/898437/Evaluation\\_of\\_sensitivity\\_and\\_specificity\\_of\\_4\\_commercially\\_available\\_SARS-CoV-2\\_antibody\\_immunoassays.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/898437/Evaluation_of_sensitivity_and_specificity_of_4_commercially_available_SARS-CoV-2_antibody_immunoassays.pdf). Accessed October 1, 2020.
- [14] Klein SL, Pekosz A, Park HS, Ursin RL, Shapiro JR, Benner SE, et al. Sex, age, and hospitalization drive antibody responses in a COVID-19 convalescent plasma donor population. *J Clin Invest* 2020. Published online August 7, 2020.
- [15] Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons; 2004.
- [16] Randolph HE, Barreiro LB. Herd Immunity: Understanding COVID-19. *Immunity* 2020;52(5):737-41.
- [17] Wu SL, Mertens AN, Crider YS, Nguyen A, Pokpongkiat NN, Djajadi S, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nat Commun* 2020;11(1):4507.
- [18] Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MM, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. *Cochrane Database Syst Rev* 2020;7:CD013665.
- [19] Clemency BM, Varughese R, Scheafer DK, Ludwig B, Welch JV, McCormack RF, et al. Symptom Criteria for COVID-19 Testing of Health Care Workers. *Acad Emerg Med* 2020;27(6):469-74.

- [20] Tostmann A, Bradley J, Bousema T, Yiek WK, Holwerda M, Bleeker-Rovers C, et al. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020. *Euro Surveill* 2020;25(16).
- [21] Sutton D, Fuchs K, D'Alton M, Goffman D. Universal Screening for SARS-CoV-2 in Women Admitted for Delivery. *N Engl J Med* 2020;382(22):2163-4.
- [22] Götzinger F, Santiago-García B, Noguera-Julián A, Lanaspá M, Lancella L, Calò Carducci FI, et al. COVID-19 in children and adolescents in Europe: a multinational, multicentre cohort study. *Lancet Child Adolesc Health* 2020;4(9):653-61.
- [23] Borges do Nascimento IJ, von Groote TC, O'Mathúna DP, Abdulazeem HM, Henderson C, Jayarajah U, et al. Clinical, laboratory and radiological characteristics and outcomes of novel coronavirus (SARS-CoV-2) infection in humans: A systematic review and series of meta-analyses. *PLoS One* 2020;15(9):e0239235.
- [24] Steensels D, Oris E, Coninx L, Nuyens D, Delforge ML, Vermeersch P, et al. Hospital-Wide SARS-CoV-2 Antibody Screening in 3056 Staff in a Tertiary Center in Belgium. *JAMA* 2020;324(2):195-7.
- [25] Iversen K, Bundgaard H, Hasselbalch RB, Kristensen JH, Nielsen PB, Pries-Heje M, et al. Risk of COVID-19 in health-care workers in Denmark: an observational cohort study. *Lancet Infect Dis* 2020. Published online August 3, 2020.
- [26] Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;26(7):1037-40.

- [27] Lechien JR, Chiesa-Estomba CM, Place S, Van Laethem Y, Cabaraux P, Mat Q, et al. Clinical and epidemiological characteristics of 1420 European patients with mild-to-moderate coronavirus disease 2019. *J Intern Med* 2020;288(3):335-44.
- [28] Seow J, Graham C, Merrick B, Acors S, Steel KJA, Hemmings O, et al. Longitudinal evaluation and decline of antibody responses in SARS-CoV-2 infection. *medRxiv* 2020:2020.07.09.20148429. Posted July 11, 2020.
- [29] Rogers TF, Zhao F, Huang D, Beutler N, Burns A, He WT, et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* 2020;369(6506):956-63.
- [30] Wang X, Guo X, Xin Q, Pan Y, Hu Y, Li J, et al. Neutralizing Antibody Responses to Severe Acute Respiratory Syndrome Coronavirus 2 in Coronavirus Disease 2019 Inpatients and Convalescent Patients. *Clin Infect Dis* 2020. Published online June 4, 2020.

**Table 1.** Seroprevalence and confidence interval (CI) in the two regions by general characteristics

	Saaremaa			Tallinn		
	Number of participants	Seroprevalence, % (95% CI)		Number of participants	Seroprevalence, % (95% CI)	
<b>Overall</b>	954	6.3 (5.0-7.9)		1006	1.5 (0.9-2.5)	
<b>Sex</b>						
Male	365	6.5 (4.4-9.4)		444	0.5 (0.1-1.8)	
Female	589	6.1 (4.5-8.3)		562	2.3 (1.3-4.1)	
<b>Age group (years)</b>						
0-9	101	5.9 (2.2-12.5)		106	3.8 (1-9.4)	
10-19	110	5.5 (2-11.5)		107	0 (0-3.4)	
20-29	104	5.8 (2.1-12.1)		111	0.9 (0-4.9)	
30-39	121	7.4 (3.5-13.7)		117	0.9 (0-4.7)	
40-49	114	7 (3.1-13.4)		120	1.7 (0.2-5.9)	
50-59	113	7.1 (3.1-13.5)		117	1.7 (0.2-6)	
60-69	112	5.4 (2-11.3)		116	2.6 (0.5-7.4)	
70-79	112	3.6 (1-8.9)		114	0 (0-3.2)	
80-100	67	9 (3.4-18.5)		98	1 (0-5.6)	
<b>Acute respiratory illness</b>						
Yes	85	15.6 (9.6-24.4)		72	1.7 (0.3-10.8)	
No	869	5.3 (4.1-6.9)		933	1.4 (0.8-2.5)	
<b>PCR</b>						
Tested	135	15.1 (10.2-21.8)		86	10.7 (5.3-20.5)	
PCR-negative	117	1.3 (0.3-5.1)		81	5.1 (1.4-16.3)	
Not tested	819	4.9 (3.6-6.5)		919	0.6 (0.3-1.3)	
<b>Household size</b>						
1	14	0.0 (0-23.1)*		15	6.3 (0.9-32.1)	
2	157	6.4 (3.6-11.0)		189	1.5 (0.4-6.0)	
3	268	6.9 (4.5-10.4)		285	0.4 (0.1-1.6)	
4	190	6.6 (4.0-10.8)		214	2.2 (0.9-5.1)	
5	185	5.6 (3.1-9.8)		219	2.0 (0.8-5.2)	
>5	140	6.2 (3.3-11.5)		83	0.0 (0-4.3)*	
<b>Contact with confirmed case</b>						
Yes	163	16.9 (12.2-23.0)		19	24.2 (9.7-48.6)	
No	791	4.0 (2.9-5.5)		986	0.9 (0.5-1.8)	

\*95% CI calculated for raw seroprevalence estimates.

**Table 2.** Characteristics of acute respiratory illness since March 1, 2020 in seropositive or seronegative individuals

	Seropositive (n=14)	Seronegative (n=143)
Fever $\geq 38^{\circ}\text{C}$ – n (%)	7 (50)	45 (31.5)
Chills – n (%)	7 (50)	57 (39.9)
Fatigue – n (%)	11 (78.6)	96 (67.1)
Muscle aches – n (%)	7 (50)	45 (31.5)
Sore throat – n (%)	5 (35.7)	88 (61.5)
Cough – n (%)	6 (42.9)	100 (69.9)
Runny nose – n (%)	8 (57.1)	99 (69.2)
Difficulty breathing – n (%)	5 (35.6)	34 (23.8)
Shortness of breath – n (%)	4 (28.6)	22 (15.4)
Chest pain – n (%)	4 (28.6)	29 (20.3)
Other respiratory symptoms – n (%)	2 (14.3)	21 (14.7)
Headache – n (%)	10 (71.4)	75 (52.4)
Nausea/vomiting – n (%)	0	17 (11.9)
Stomach ache – n (%)	2 (14.3)	19 (13.3)
Diarrhea – n (%)	5 (35.7)	32 (22.4)
Sought medical attention – n (%)	8 (57.1)	90 (62.9)
Missed work or school – n (%)	8 (57.1)	61 (42.7)
Hospitalized* – n (%)	3 (21.4)	7 (4.9)

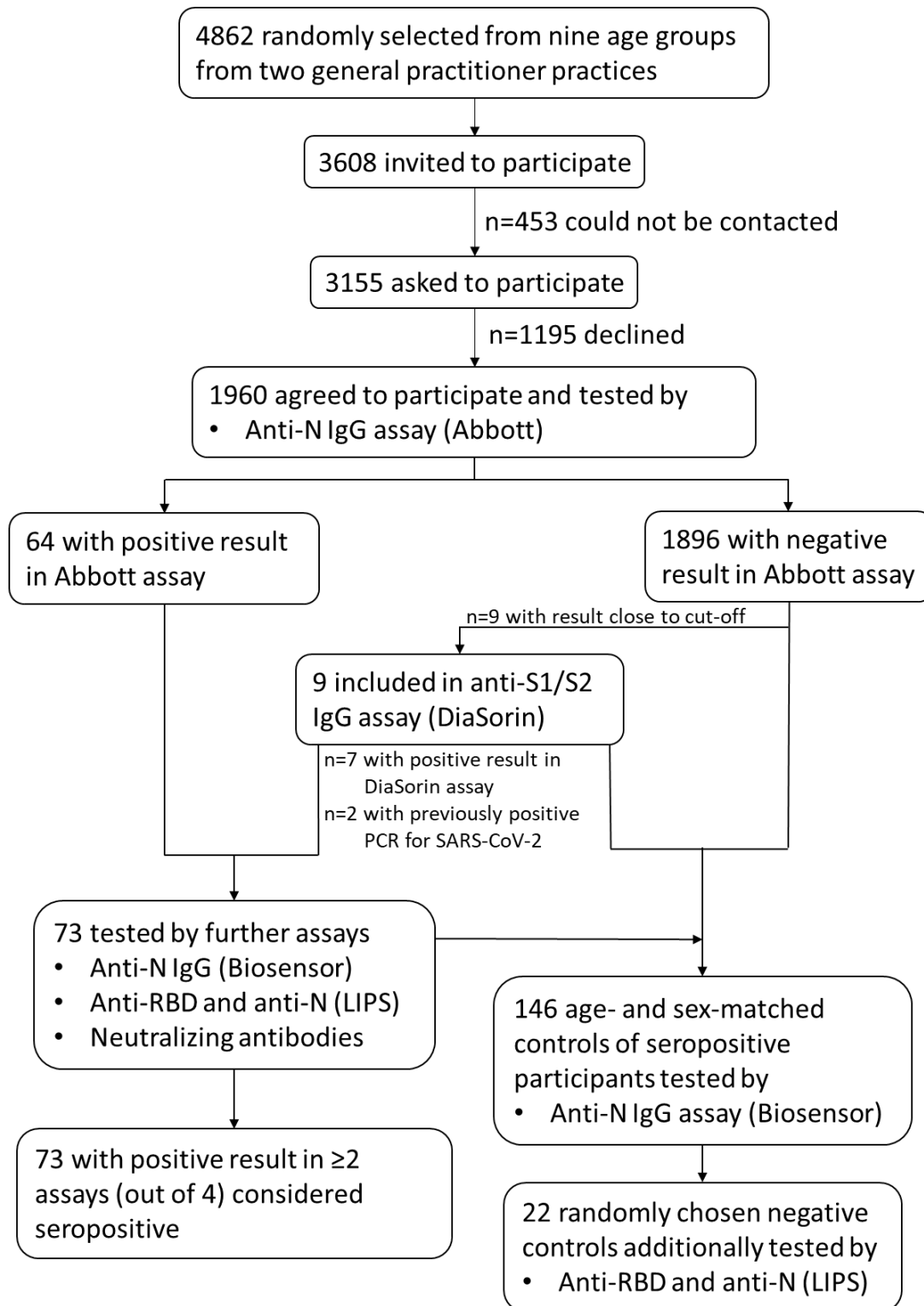
\*Only difference between the proportion of hospitalized patients between seropositive and seronegative individuals was statistically significant ( $p=0.047$ ).

**Table 3.** Regression coefficients and 95% confidence intervals of multiple linear regression models of association between characteristics of acute respiratory illness and IgG concentrations or neutralizing antibody titer in seropositive individuals, adjusted for age, sex or time since onset of symptoms or multiply imputed positive PCR test

	Anti-RBD IgG by LIPS	Anti-N IgG by LIPS	Anti-N IgG by Abbott	Neutralizing antibodies
Acute respiratory illness	29.0 (-19.0...77.0)	5.0 (-0.1...10.1)	1.5 (-0.1...3.0)	0.6 (-0.7...1.9)
Fever $\geq 38^{\circ}\text{C}$	<b>62.4</b> <b>(0.6...124.2)*</b>	<b>9.8</b> <b>(3.4...16.3)**</b>	<b>3.0</b> <b>(1.1...4.9)**</b>	1.2 (-0.5...2.9)
Chills	-10.5 (-73.8...52.8)	1.6 (-5.2...8.4)	0.4 (-1.6...2.5)	0.01 (-1.7...1.7)
Fatigue	14.2 (-37.0...65.5)	4.5 (-0.9...10.0)	1.1 (-0.6...2.7)	0.3 (-1.0...1.7)
Muscle aches	23.3 (-37.0...83.5)	4.2 (-2.3...10.7)	1.3 (-0.6...3.2)	0.5 (-1.1...2.2)
Sore throat	-50.1 (118.9...18.7)	-2.9 (-10.4...4.6)	0.3 (-2.0...2.5)	-0.5 (-2.4...1.3)
Cough	20.9 (-45.3...87.1)	5.6 (-1.5...12.6)	1.7 (-0.4...3.8)	0.6 (-1.2...4.1)
Runny nose	12.3 (-46.4...71.0)	3.8 (-2.5...10.0)	1.5 (-0.3...3.4)	1.0 (-0.5...2.6)
Difficulty breathing	69.7 (-4.1...143.5)	<b>9.3</b> <b>(1.4...17.1)*</b>	2.3 (-0.1...4.6)	1.0 (-1.0...3.1)
Shortness of breath	53.6 (-25.8...132.9)	<b>9.4</b> <b>(1.1...17.8)*</b>	<b>2.6 (0.1...5.1)*</b>	1.7 (-0.4...3.9)
Chest pain	46.0 (-31.2...123.3)	8.1 (-0.1...16.4)	<b>3.2</b> <b>(0.9...5.6)**</b>	1.0 (-1.1...3.1)
Other respiratory symptoms	-8.9 (-116.2...98.4)	2.7 (-8.8...14.3)	1.5 (-1.9...5.0)	-0.1 (-3.0...2.8)
Headache	4.1 (-49.2...57.4)	1.9 (-3.8...7.6)	0.5 (-1.2...2.2)	0.1 (-1.3...1.6)
Stomach ache	-60.2 (-165.4...45.0)	-7.6 (-18.9...3.7)	-1.0 (-4.4...2.4)	-2.0 (-4.8...0.9)
Diarrhea	31.1 (-38.2...100.4)	6.4 (-1.0...13.7)	<b>2.5 (0.4...4.7)*</b>	0.3 (-1.6...2.2)
Hospitalized	<b>102.8</b> <b>(13.6...192.0)*</b>	<b>10.5</b> <b>(0.9...20.2)*</b>	2.7 (-0.2...5.6)	1.3 (-1.2...3.8)

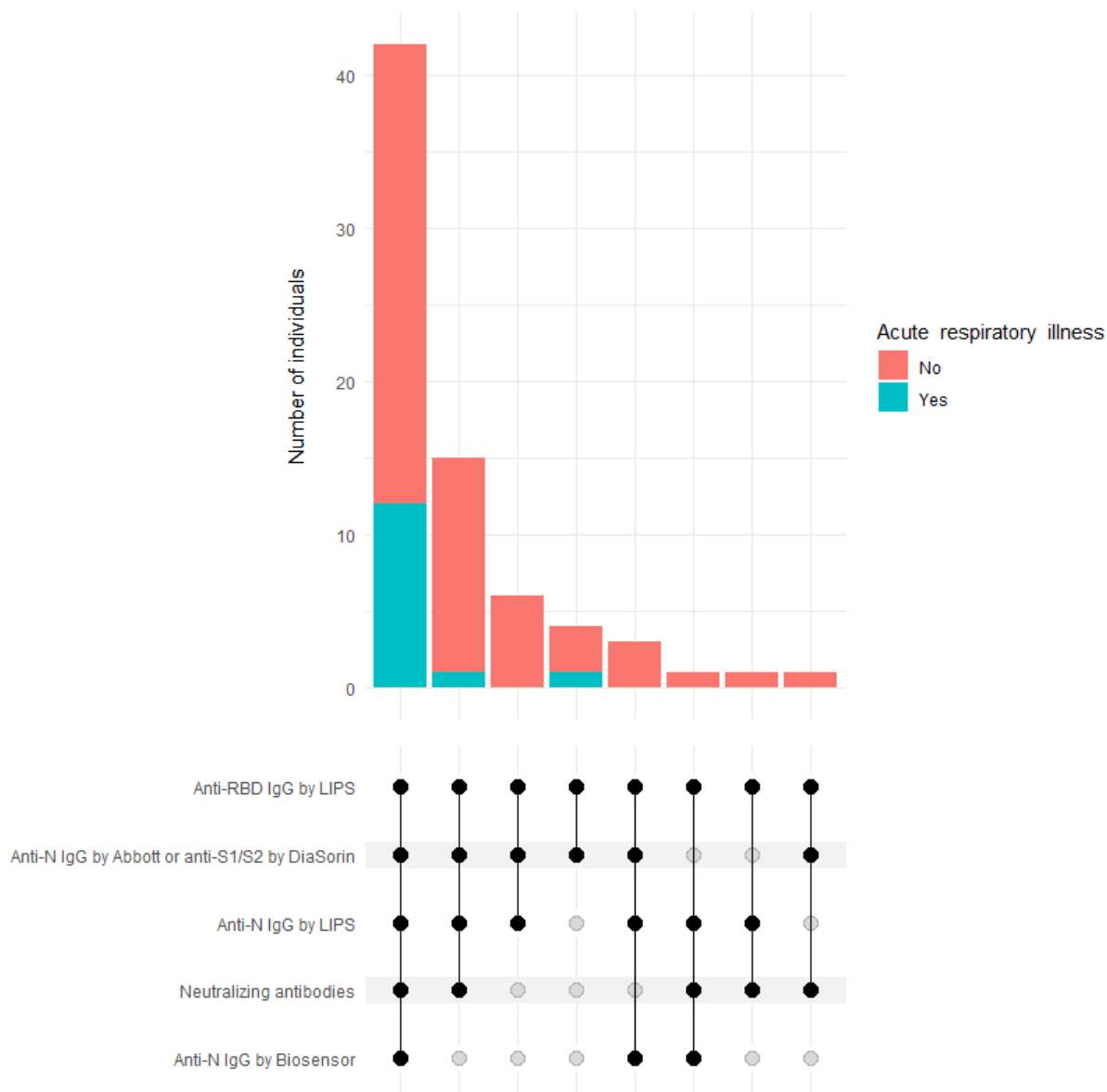
Statistically significant regression coefficients are shown in bold.

\* p-value between 0.01 to 0.05, \*\* p-value <0.01.

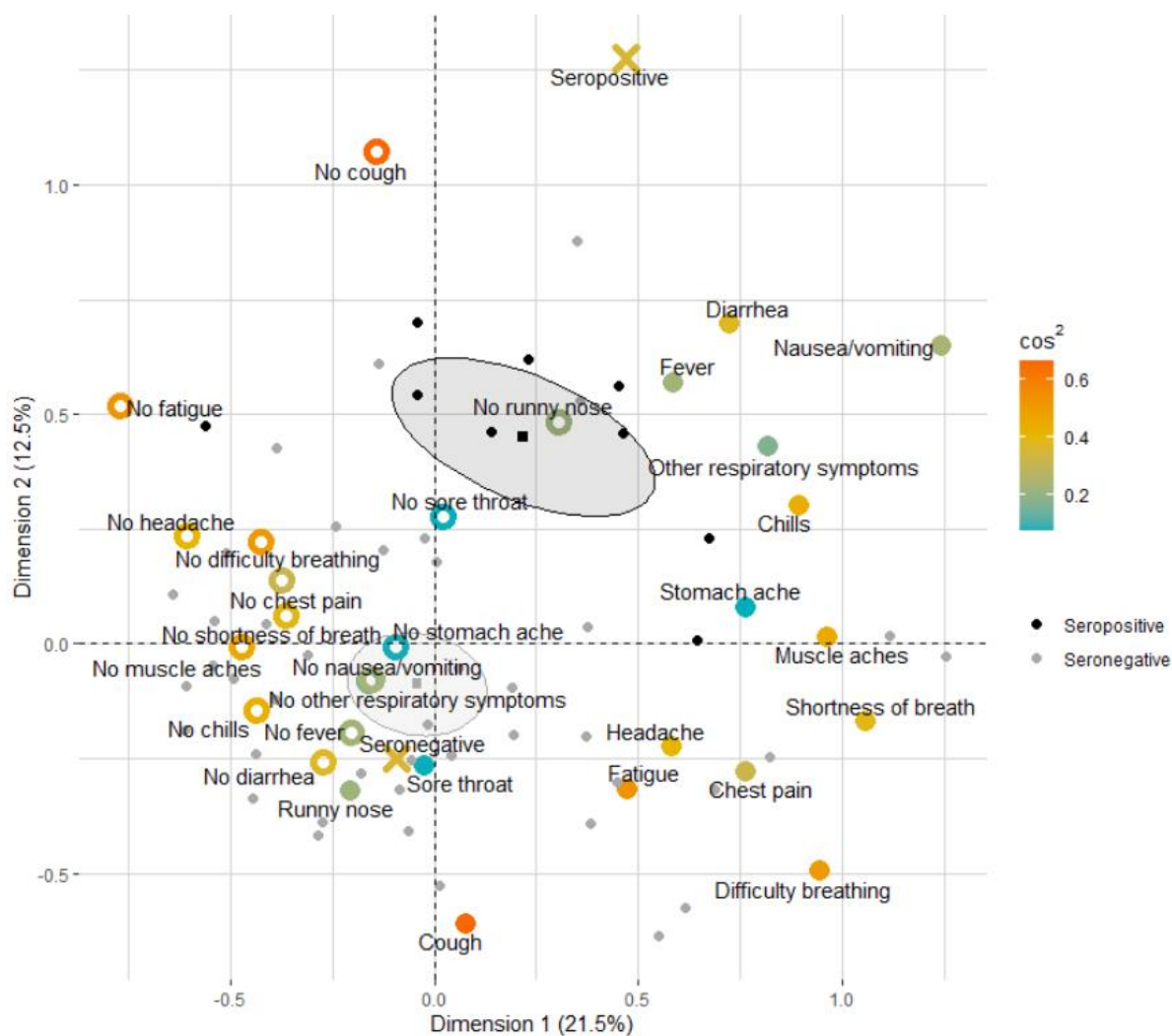


**Figure 1.** Flowchart of recruitment of study participants, testing and definition for being considered seropositive





**Figure 2.** Upset plot of number of seropositive individuals with positive test results in different assays. Each bar shows number of individuals with positive results from the assays indicated by black circles beside the plot. Bars are colored according to number of individuals who had had acute respiratory illness (red) or not (green). Circles in the bottom show for each set of individuals negative (transparent) or positive (black) test result by the respective assay. Solid lines connect black circles.



**Figure 3.** Biplot of the first two dimensions of multiple correspondence analysis on association between the presence (filled large dots) or absence (empty large dots) of symptoms and seropositivity (crosses) in all individuals aged 50 years or more with acute respiratory illness. The variables included in the analysis are colored according to the sum of  $\cos^2$  of the variable on dimensions one and two. Grey and black small dots represent seronegative and seropositive individuals, respectively. Practices of gravity of seropositive and seronegative individuals are shown by squares surrounded by 95% confidence ellipses.

## Supplementary Content

### Supplementary Information 1. Study questionnaire

#### 1. Personal data

1.1.	First name	
1.2.	Family name	
1.3.	Patient ID	
1.4.	Age (in years)	
1.5.	Phone number	
1.6.	E-mail	
1.7.	Place of residence (county)	
1.8.	Number of family members living together	
1.9.	Form completion date (dd/mm/yyyy)	

#### 2. History of COVID-19

	Yes	No
2.1. Have you had acute respiratory illness since March 1, 2020?		

2.1.1. If Yes, then date of becoming ill? ...../...../..... (dd/mm/yyyy)

	Yes	No
2.2. Have you had a COVID-19 PCR test (nasal swab)?		
2.2.1. If Yes, then did the PCR test confirm COVID-19 infection?		
2.3. Have you been in contact with someone who has been diagnosed with COVID-19 infection?		

2.3.1. If Yes, then date of the last contact? ...../...../..... (dd/mm/yyyy)

	Yes	No	Not sure
2.4. Have you been vaccinated against tuberculosis in childhood?			

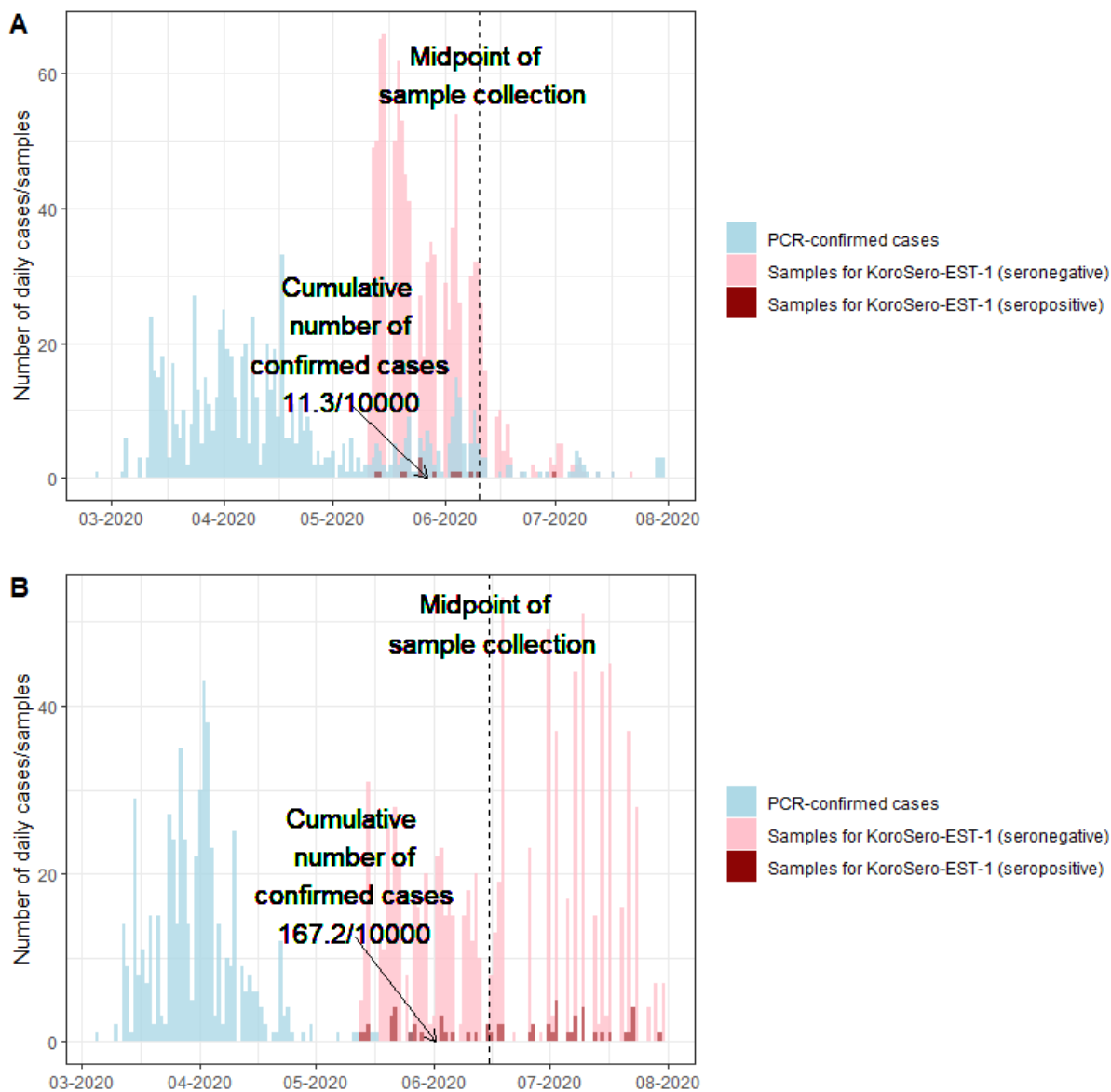
#### 3. Symptom history (only for those who answered Yes in 2.1)

	Yes	No
Fever $\geq 38^{\circ}\text{C}$		
Chills		
Fatigue		
Muscle aches		
Sore throat		
Cough		
Runny nose		
Difficulty breathing		
Shortness of breath		

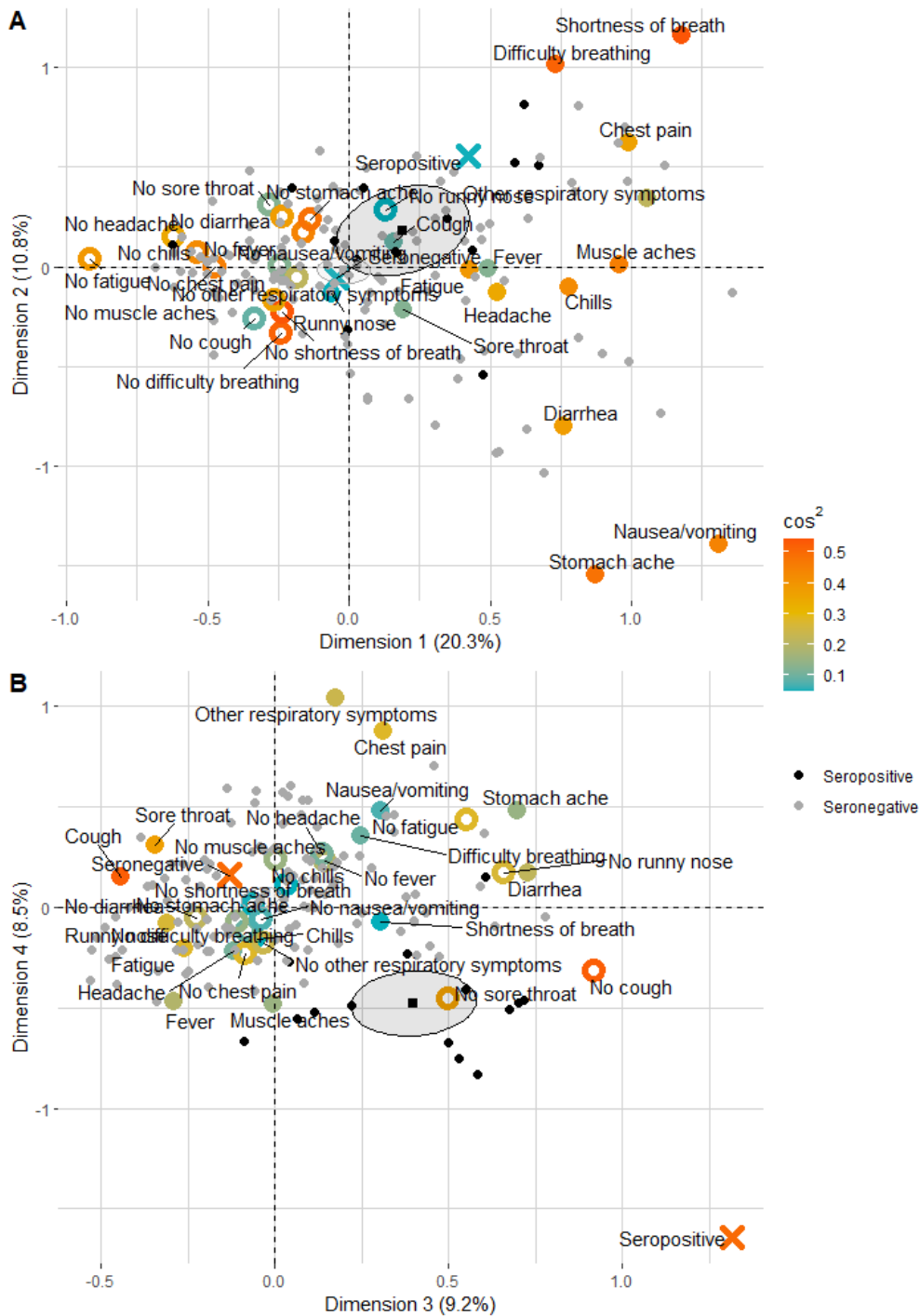
Chest pain		
Other respiratory symptoms		
Headache		
Nausea/vomiting		
Stomach ache		
Diarrhea		
Did any of these symptoms require you to seek medical attention?		
Did any of these symptoms require you to miss work?		
Did any of these symptoms require you to be hospitalized?		

**Table S1.** Seroprevalence and seroprevalence adjusted for test performance according to age groups and confidence interval (CI)

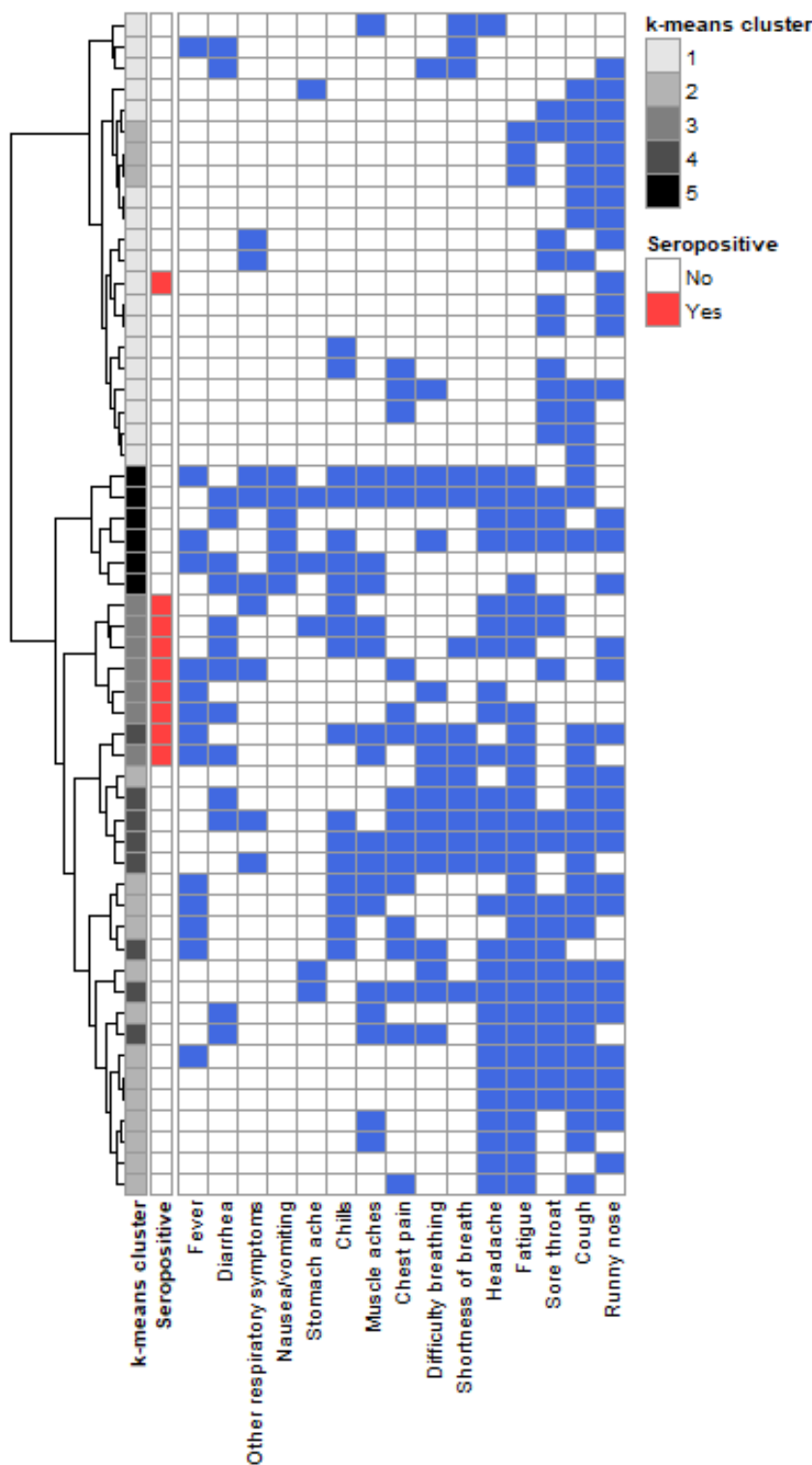
<b>Age group (years)</b>	<b>Seroprevalence, % (95% CI)</b>	<b>Seroprevalence adjusted for test performance, % (95% CI)</b>
<b>Tallinn</b>		
0-9	3.8 (1-9.4)	4.0 (0.8-8.1)
10-19	0 (0-3.4)	0.0 (0-0)
20-29	0.9 (0-4.9)	0.9 (0-2.9)
30-39	0.9 (0-4.7)	0.8 (0-2.8)
40-49	1.7 (0.2-5.9)	1.7 (0-4.5)
50-59	1.7 (0.2-6)	1.7 (0-4.6)
60-69	2.6 (0.5-7.4)	2.7 (0-6.3)
70-79	0 (0-3.2)	0.0 (0-0)
80-100	1 (0-5.6)	1.0 (0-3.3)
<b>Saaremaa</b>		
0-9	5.9 (2.2-12.5)	6.3 (2-11.7)
10-19	5.5 (2-11.5)	5.8 (1.8-10.7)
20-29	5.8 (2.1-12.1)	6.1 (1.9-11.4)
30-39	7.4 (3.5-13.7)	7.9 (3.4-13.4)
40-49	7 (3.1-13.4)	7.5 (2.7-13)
50-59	7.1 (3.1-13.5)	7.5 (2.8-13.1)
60-69	5.4 (2-11.3)	5.7 (1.7-10.6)
70-79	3.6 (1-8.9)	3.7 (0.7-7.8)
80-100	9 (3.4-18.5)	9.6 (3.1-17.6)



**Figure S1.** PCR-confirmed cases (blue) and number of seropositive (dark red) or seronegative (pink) serum samples drawn for the seroepidemiological study KoroSero-EST-1 (A) in Tallinn and (B) in Saaremaa



**Figure S2.** Biplot of the dimensions 1 and 2 (panel A) and 3 and 4 (panel B) of multiple correspondence analysis on association between the presence (filled large dots) or absence (empty large dots) of symptoms and seropositivity (crosses) in all individuals with acute respiratory illness. The variables included in the analysis are colored according to the sum of  $\cos^2$  of the variable on dimensions 1 and 2 (panel A) or 3 and 4 (panel B). Grey and black small dots represent seronegative and seropositive individuals, respectively. Practices of gravity of seropositive and seronegative patients are shown by squares surrounded by 95% confidence ellipses. According to dimensions three and four seropositivity was associated with the absence of cough and the absence of sore throat



**Figure S3.** Clustering of individuals aged 50 years or more with acute respiratory illness according to hierarchical clustering on components of multiple correspondence analysis and subsequent k-means consolidation into 5 groups (leftmost column, designated with different shades of grey). Seropositivity is shown in the second leftmost column in red. The presence of the respective symptom is indicated in blue. The majority (7/9) of seropositive individuals are grouped into a single k-means cluster containing no seronegative individuals.