


# BMJ Open Testing behaviour and positivity for SARS-CoV-2 infection: insights from web-based participatory surveillance in the Netherlands

Scott A McDonald ,<sup>1</sup> Lucia C Soetens,<sup>1</sup> C Maarten A Schipper,<sup>1</sup> Ingrid Friesema,<sup>1</sup> Cees C van den Wijngaard,<sup>1</sup> Anne Teirlinck,<sup>1</sup> Nienke Neppelenbroek,<sup>1</sup> Susan van den Hof,<sup>1</sup> Jacco Wallinga,<sup>1,2</sup> Albert Jan van Hoek<sup>1</sup>

**To cite:** McDonald SA, Soetens LC, Schipper CMA, *et al.* Testing behaviour and positivity for SARS-CoV-2 infection: insights from web-based participatory surveillance in the Netherlands. *BMJ Open* 2021;**11**:e056077. doi:10.1136/bmjopen-2021-056077

► 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-2021-056077>).

Received 04 August 2021  
Accepted 22 November 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

<sup>2</sup>Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

## Correspondence to

Dr Scott A McDonald;  
[scott.mcdonald@rivm.nl](mailto:scott.mcdonald@rivm.nl)

## ABSTRACT

**Objectives** We aimed to identify populations at a high risk for SARS-CoV-2 infection but who are less likely to present for testing, by determining which sociodemographic and household factors are associated with a lower propensity to be tested and, if tested, with a higher risk of a positive test result.

**Design and setting** Internet-based participatory surveillance data from the general population of the Netherlands.

**Participants** Weekly survey data collected over a 5-month period (17 November 2020 to 18 April 2021) from a total of 12 026 participants who had contributed at least 2 weekly surveys was analysed.

**Methods** Multivariable analyses using generalised estimating equations for binomial outcomes were conducted to estimate the adjusted ORs of testing and of test positivity associated with participant and household characteristics.

**Results** Male sex (adjusted OR for testing (OR<sub>t</sub>): 0.92; adjusted OR for positivity (OR<sub>p</sub>): 1.30, age groups <20 (OR<sub>t</sub>: 0.89; OR<sub>p</sub>: 1.27), 50–64 years (OR<sub>t</sub>: 0.94; OR<sub>p</sub>: 1.06) and 65+ years (OR<sub>t</sub>: 0.78; OR<sub>p</sub>: 1.24), diabetics (OR<sub>t</sub>: 0.97; OR<sub>p</sub>: 1.06) and sales/administrative employees (OR<sub>t</sub>: 0.93; OR<sub>p</sub>: 1.90) were distinguished as lower test propensity/higher test positivity factors.

**Conclusions** The factors identified using this approach can help identify potential target groups for improving communication and encouraging testing among those with symptoms, and thus increase the effectiveness of testing, which is essential for the response to the COVID-19 pandemic and for public health strategies in the longer term.

## INTRODUCTION

The pillars of effective control of the acute respiratory disease COVID-19 are testing, contact tracing and isolation after a positive test.<sup>1–3</sup> For community-level diagnostic testing to contribute effectively to reducing the spread of infection, convenience and accessibility to testing facilities are only a

## Strengths and limitations of this study

- This study, using web-based weekly survey data collected from voluntary participants, identified sociodemographic and household factors that were associated with a lower chance of getting tested and a higher chance of a positive outcome, if tested.
- This strength overcomes a limitation of most COVID-19 surveillance systems: they do not contain data suitable for inference on the factors associated with willingness—or alternatively, barriers—to present for SARS-CoV-2 testing, as people who are not tested are excluded.
- As with all analyses of participatory surveillance data, one cannot distinguish the effects of the measured variable from reporting behaviour.
- The observed dependence between lower test propensity and higher positivity rate for certain participant groups could also be explained by between-group differences in the symptom severity threshold for getting tested.
- Participants may be more willing to engage with health services compared with the general population ('volunteer bias'), which would limit the generalisability of our estimated SARS-CoV-2 testing rates.

first step. Awareness of governmental guidance regarding testing, recognition of relevant symptoms and personal responsibility or motivation to contact health services to arrange an SARS-CoV-2 test appointment are also required.<sup>3</sup> It, therefore, is vital to identify those characteristics that are associated with willingness (or aversion) to undergo testing, given one or more symptoms compatible with COVID-19, especially for those persons not already identified through contact tracing. Furthermore, knowledge of the characteristics that are most strongly associated with SARS-CoV-2 positivity will optimise interpretation of test data and assist in helping public

health communication in the targeting of advice to individuals that should get tested and/or self-isolate.

For studies of the factors associated with infectious disease risk in the community, internet-based participatory surveillance is a valuable tool, as demonstrated by studies of influenza-like illness in the Netherlands and elsewhere<sup>4,5</sup> and new app-based monitoring of COVID-19 of symptoms and other outcomes.<sup>6–8</sup> To gain insight into sociodemographic, participant and household factors associated with testing for SARS-CoV-2 infection, we rely on data from the Infectieradar web-based participatory surveillance system<sup>9</sup> for COVID-19 in the Netherlands, which was set up to monitor a number of epidemiological outcomes, including the occurrence of symptom(s), PCR/antigen test behaviour and test results. Other existing surveillance systems in the Netherlands can only collect information on those persons who get tested, as persons who would be eligible, but do not present for testing, are not reached.

Using Infectieradar data, we investigated the associations between participant/household characteristics and the propensity to get tested for SARS-CoV-2 infection, given symptoms. We also quantified the associations between the same factors and a positive test result among those tested, irrespective of symptoms. By combining the results of both analyses, it is possible to identify persons in certain demographic categories or with other risk factors who are less likely to present for testing, but have a higher risk of testing positive.

## METHODS

### Study setting, analysis period and participants

We conducted a cohort study based on approximately 5 months of Infectieradar data collected between 17 November 2020 and 18 April 2021. As of the study period end date, the total number of active users (defined as those who had filled out at least 2 weekly surveys since the date when Infectieradar data collection was restarted in the autumn of 2020 (ie, 2 October 2020) was 17 054. We began analysis on 17 November 2020, as testing eligibility policy prior to this date excluded children younger than 12 years from testing unless severely ill.

### Patient and public involvement

There was no direct patient or public involvement in the design of this study. Infectieradar participants were given opportunities to provide feedback on the survey questionnaires, which were regularly updated to improve clarity.

### Data source

Participants were recruited for the Infectieradar surveillance system (a member of the Influenzanet consortium, a European collaboration involving universities and public health partners) via a web announcement.<sup>9</sup> Registration was open to all residents of the Netherlands. Children younger than 16 years could participate under supervision of their legal guardian, or if their legal guardian

acted on their behalf. After providing informed consent, participants completed registration by filling out an intake questionnaire, which asks for sociodemographic data (age, sex, education level, occupation, partial postal code, number and age make-up of persons in household, etc) and medical history (eg, pre-existing health conditions such as allergies/hay fever and chronic diseases). These elicited data were selected in line with previous research conducted by the Influenzanet consortium.<sup>5</sup> In addition, the intake survey queried if the participant had ever received a positive PCR/rapid test result, and the date the test was conducted.

After completing the intake survey, participants were requested to fill out a standard questionnaire (and every week thereafter), to report any symptoms experienced during the past 7-day period. Namely, in each weekly survey participants were asked to report the occurrence of one or more of a set of 21 symptoms (fever, chills, runny nose, sneezing, sore throat, cough, dyspnoea (shortness of breath), headache, muscle/joint pain, chest pain, malaise, loss of appetite, coloured phlegm, watery or bloodshot eyes, nausea, vomiting, diarrhoea, stomach ache, loss of sense of smell, loss of sense of taste, or nose-bleed). If a participant had been tested for SARS-CoV-2 infection since their previous survey (irrespective of symptoms), they were asked to report the type of test and the test outcome. Additional weekly survey data included information regarding healthcare-seeking behaviour and suspected cause of symptom(s) (if reported). All data were pseudonymised, with individual participants assigned a unique identifier used in subsequent analysis.

### Statistical analysis

#### Data inclusion/exclusion criteria

We attempt to reduce the impact of selection bias caused by persons who registered or only participated once because they recently experienced symptoms, by (1) excluding participants who had contributed fewer than 2 weekly surveys, and (2) removing the first weekly survey submitted by each participant. Prevalent illness at the time of registration was addressed by excluding all weekly surveys in which symptom onset was indicated as before the date of intake survey. Surveys for persons with missing age or sex (n=24) were excluded.

#### Regression analysis

We fitted univariate and multivariable binomial generalised estimating equations to the two distinct outcomes (1) test propensity (among all participants who reported at least one symptom from the set of 21 symptoms queried), and (2) test positivity (among all participants reporting being tested since their last survey). We did not restrict (2) to only those surveys in which symptoms were reported because the participant could have recovered by the time they reported having been tested (the test date may have been more than 1 week previously). For (1) we were interested in the associations between participant/household characteristics and test propensity

(conditioning on symptoms reduces the influence of factors such as requests for testing generated by contact tracing or to be released from quarantine); for (2) we were interested in the associations between participant/household characteristics and a positive result, conditioned on being tested.

Besides sex, age group, education level (higher, middle, none/lower or missing) and underlying conditions (asthma, allergies/hay fever, chronic lung disease, diabetes, cardiovascular disease), smoker status (never; current/ex-smoker), children aged <5 years in household, children 5–18 years in household, and occupation category (see below), the covariate set for outcome (1) only included covariates for test history (defined with respect to the date of each weekly survey): Never, previously reported negative result, previously reported positive result (including a positive result reported at the time of intake), and ‘suspected non-COVID-19 cause’ of reported symptom(s). The latter variable was defined as ‘yes’ if any of the answers ‘influenza/influenza-like illness’, ‘common cold’, or ‘allergies/hay fever/asthma’, ‘gastrological complaints/stomach influenza’, or ‘other’ were selected for the question ‘Do you have any idea what caused your symptoms?’, and as ‘no’ otherwise (thus ‘no’ included the responses ‘COVID-19’ and ‘do not know’). For outcome (2) only, the covariate logarithm-transformed total number of reported symptoms (in that weekly survey) was added. This transformation was selected based on prior assessment of model fit. We defined six categories for the occupation covariate: Not applicable (preschool, pupil, student, household, unemployed, retired); healthcare, education (including day-care staff), knowledge worker (eg, manager, accountant, scientist), sales/administration (eg, shop/supermarket staff, receptionist, administrator, financial assistant) and all other occupations. In regression analysis, knowledge worker—the occupation category with the greatest proportion of participants—was defined as the reference category.

One model for each outcome was fitted, retaining only participant/other factors (excluding symptoms) as covariates. A natural spline was fitted in each model to capture temporal trends in test uptake (five knots) and test positivity (three knots) that is not explained by the other covariates. We used a generalised estimating equations approach with exchangeable correlation structure to correct standard errors for repeated observations per participant.

The predicted absolute risk of getting tested for each variable was estimated from the fitted regression model using marginal standardisation;<sup>10</sup> that is, producing the predicted proportion tested, taking into account the distribution of other variables in the model.

To bring together both sets of association measures for test propensity and positivity, we graphically categorised participant characteristics associated with a low propensity to get tested for SARS-CoV-2, but a relatively high risk of a positive outcome. Finally, we calculated the positive

predictive value (PPV) for each of the 21 possible symptoms that could be reported, to determine if the efficiency of the above-identified participant factors could be improved by adding reported symptom(s).

All analyses were conducted using R statistical software, V3.6.0.<sup>11</sup>

## RESULTS

During the 5-month analysis period between 17 November 2020 and 18 April 2021, 282 277 weekly surveys were submitted by 16 807 unique Infectieradar participants. This represents a participation rate of 0.10% of the total Netherlands population. Overall, 12 026 individuals reported at least one symptom in 50 946 weekly surveys. Among these 50 946 surveys, undergoing a PCR/antigen test and receipt of the test outcome since the previous survey was reported in 8997 surveys (51% of participants) (table 1); 48% of ever-tested individuals were tested two times or more. A previous negative test result (in 42.0% of surveys) or previously testing positive (in 7.6%) was reported. Females (57.2%) and those with a higher education level (60.7%) were over-represented in Infectieradar and the age-distribution was older (median 53.7 years) compared with the Netherlands general population (online supplemental table S1).

The adjusted ORs for the propensity to test associated with participant/household factors from all surveys with at least one reported symptom are provided in table 1 (see also figure 1). The strongest positive associations were observed for: suspected non-COVID-19 cause (OR=1.45 (95% CI: 1.37 to 1.53)), children <5 years in household (1.41 (1.29 to 1.54)), children 5–18 years in household (1.22 (1.14 to 1.31)), occupation: healthcare (1.19 (1.08 to 1.30), with knowledge worker as reference), age 20–29 years (1.19 (1.05 to 1.35)), and age 30–39 years (1.18 (1.08 to 1.29)), compared with the reference category 40–49 years). The lowest odds of being tested were for: previously tested positive (0.35 (0.30 to 0.42)), allergy sufferers (0.77 (0.73 to 0.81)), lower education level (0.78 (0.63 to 0.98), compared with middle education level), and age 65+ years (0.78 (0.69 to 0.89)). To estimate the absolute risk of testing per participant/household category among Infectieradar participants, we calculated the model-predicted test propensity using marginal standardisation. Among all participant/household characteristics, the highest predicted propensity was observed for the age group 20–29 years (22.5%), with the lowest predicted propensity for lower education level and the age group 65+ years (16.2% for both).

Among all those who reported having been tested since their previous weekly survey, irrespective of the presence of symptoms (n=13 219 surveys), 6.5% (n=854) had a positive outcome; among only those tested with reported symptom(s), 8.7% (n=783/9008 surveys) were positive. The associations between participant/household factors and a positive test result among all tested participants (whether or not reporting symptoms) are shown in

**Table 1** Distribution over participant and other factors, and results of univariate and multivariable logistic regressions using generalised estimating equations for the outcome test propensity; study period 17 November 2020 through 18 April 2021 (n=50946 surveys with at least one reported symptom)

Variable	n reporting tested	N weekly surveys	Predicted proportion tested	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
(All)	8997	50946	0.195	–	–
Sex					
Male	3118	18993	0.185	0.89 (0.84 to 0.94)	0.92 (0.86 to 0.98)
Female	5879	31953	0.199	Ref.	Ref.
Age group					
<20 years	161	1052	0.179	0.75 (0.61 to 0.91)	0.89 (0.69 to 1.14)
20–29 years	661	3338	0.225	1.01 (0.90 to 1.13)	1.19 (1.05 to 1.35)
30–39 years	1962	7974	0.224	1.33 (1.22 to 1.45)	1.18 (1.08 to 1.29)
40–49 years	2007	10312	0.197	Ref.	Ref.
50–64 years	3115	19393	0.187	0.80 (0.75 to 0.86)	0.94 (0.86 to 1.02)
65+ years	1091	8877	0.162	0.60 (0.55 to 0.66)	0.78 (0.69 to 0.89)
Education level					
None/lower or missing	203	1556	0.162	0.78 (0.65 to 0.93)	0.78 (0.63 to 0.98)
Middle	2840	17279	0.197	Ref.	Ref.
Higher	5954	32111	0.194	1.12 (1.06 to 1.19)	0.98 (0.92 to 1.04)
Smoking status					
Never smoker	8192	46129	0.195	Ref.	Ref.
Current/ex-smoker	805	4817	0.185	0.94 (0.86 to 1.04)	0.94 (0.85 to 1.03)
Asthma					
No asthma	8212	46281	0.194	Ref.	Ref.
Asthma	785	4665	0.197	0.95 (0.86 to 1.05)	1.02 (0.92 to 1.14)
Allergy(s)/hay fever					
No allergy(s)/hay fever	5198	27474	0.213	Ref.	Ref.
Allergy(s)/hay fever	3799	23472	0.173	0.84 (0.80 to 0.89)	0.77 (0.73 to 0.81)
Diabetes					
No diabetes	8379	49116	0.194	Ref.	Ref.
Diabetes	258	1830	0.190	0.77 (0.66 to 0.90)	0.97 (0.83 to 1.14)
Chronic lung disease					
No chronic lung disease	8775	49489	0.194	Ref.	Ref.
Chronic lung disease	222	1457	0.201	0.84 (0.70 to 1.01)	1.05 (0.86 to 1.27)
Cardiovascular disease					
No cardiovascular disease	8324	46264	0.194	Ref.	Ref.
Cardiovascular disease	673	4682	0.194	0.77 (0.70 to 0.86)	1.00 (0.89 to 1.11)
Household composition					
1+ children <5 years in household					
No	7498	45457	0.187	Ref.	Ref.
Yes	1499	5489	0.243	1.83 (1.70 to 1.97)	1.41 (1.29 to 1.54)
1+ children 5–18 years in household					
No	6195	37835	0.186	Ref.	Ref.
Yes	2802	13111	0.216	1.36 (1.28 to 1.44)	1.22 (1.14 to 1.31)
Occupation category					
Education	821	3916	0.213	1.09 (0.98 to 1.22)	1.12 (1.00 to 1.25)
Healthcare	1502	7077	0.222	1.10 (1.00 to 1.20)	1.19 (1.08 to 1.30)
Knowledge	1965	9964	0.195	Ref.	Ref.
Sales/admin	774	4587	0.183	0.86 (0.77 to 0.96)	0.93 (0.82 to 1.04)
Other	1528	8279	0.200	0.94 (0.86 to 1.03)	1.03 (0.94 to 1.13)
Not applicable	2407	17123	0.176	0.69 (0.64 to 0.74)	0.88 (0.80 to 0.97)

Continued

Table 1 Continued

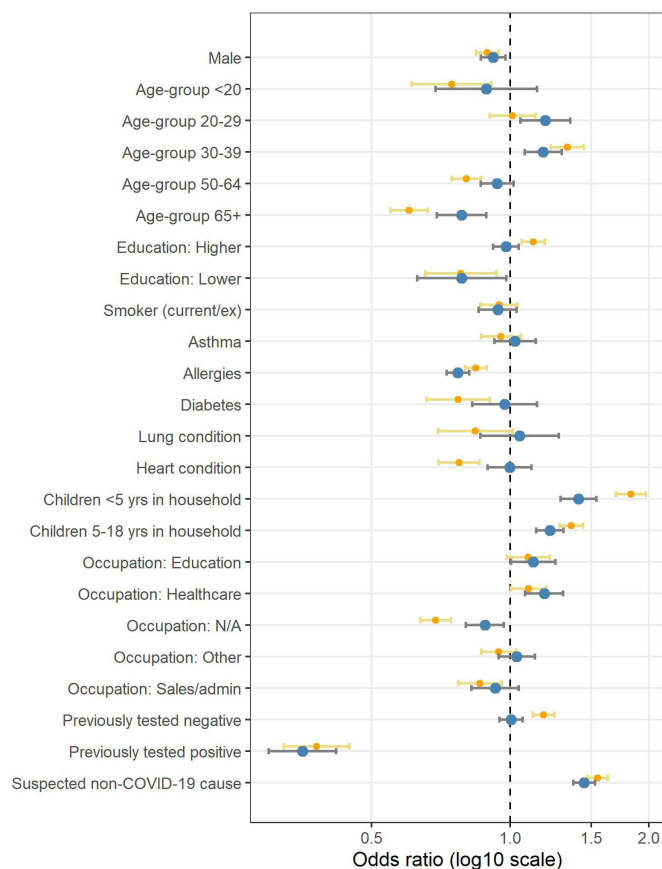
Variable	n reporting tested	N weekly surveys	Predicted proportion tested	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Previous swab test					
No	3917	25 699	0.202	Ref.	Ref.
Yes, negative	4708	21 399	0.203	1.18 (1.12 to 1.25)	1.00 (0.95 to 1.06)
Yes, positive	372	3848	0.084	0.38 (0.32 to 0.45)	0.35 (0.30 to 0.42)
Suspected non-COVID-19 cause					
No	2855	20 128	0.160	Ref.	Ref.
Yes	6142	30 818	0.215	1.55 (1.47 to 1.63)	1.45 (1.37 to 1.53)

The model-predicted proportion of surveys in which testing was reported is also shown.

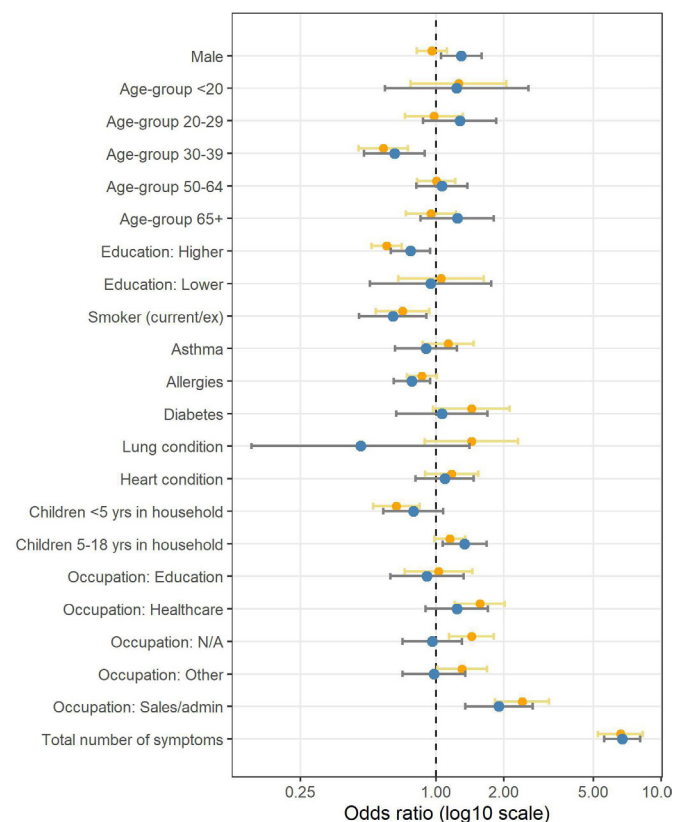
**figure 2.** The highest odds of testing positive were for: (log) total number of symptoms (6.71 (5.58 to 8.07)), occupation: sales/admin (1.90 (1.35 to 2.68)), compared with knowledge worker), children 5–18 years in household (1.34 (1.07 to 1.68)), males (1.30 (1.05 to 1.59)), and age group <20 years (1.27 (0.88 to 1.85)), though OR

was not significant. The lowest odds of testing positive were for: ever smoker (0.64 (0.46 to 0.91)), age group 30–39 (0.65 (0.48 to 0.89), compared with 40–49 years) and higher education level (0.77 (0.63 to 0.94)).

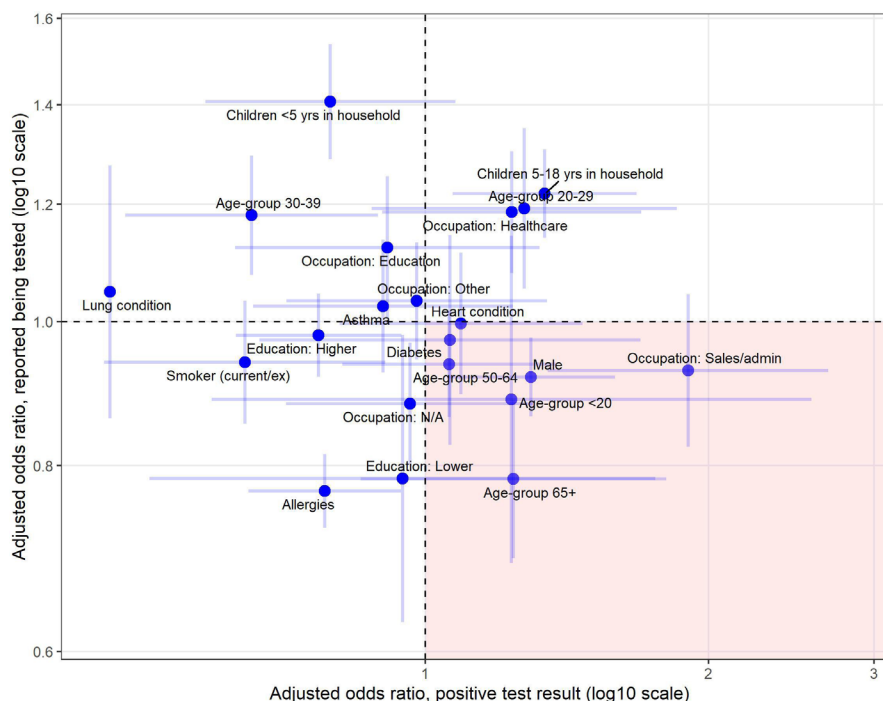
To bring together both sets of association measures for test propensity and positivity, we categorised participant characteristics associated with a low propensity to get tested for SARS-CoV-2, but a relatively high risk of



**Figure 1** Coefficients from generalised estimating equations regression analysis of test propensity, among all participants reporting at least one symptom (analysis period 17 November 2020 to 18 April 2021); n=50 946 surveys. Orange (lighter) and blue (darker) symbols represent unadjusted and adjusted coefficients, respectively. Reference categories for age group was 40–49 years; for education level, middle; for occupation, knowledge worker; for previous swab test, no.



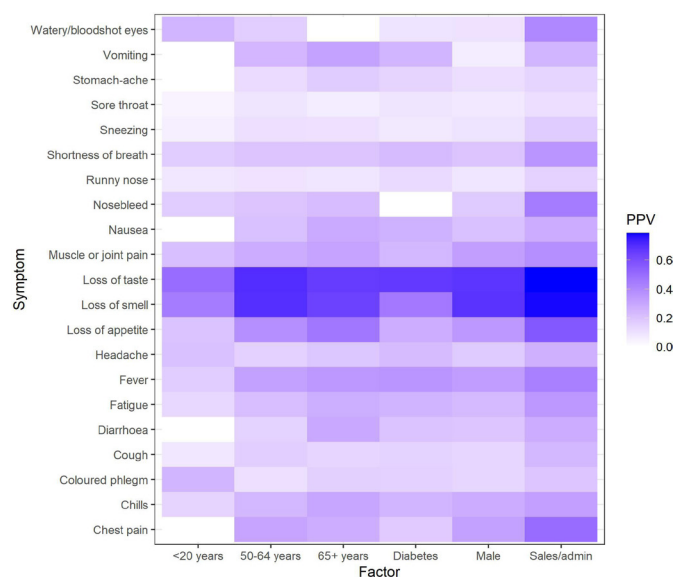
**Figure 2** Coefficients from generalised estimating equations regression analysis of positive test result among all participants who reported undergoing a swab test since their previous survey (analysis period 17 November 2020 to 18 April 2021; n=12 315). Orange (lighter) and blue (darker) symbols represent unadjusted and adjusted coefficients, respectively.



**Figure 3** Crossplot comparing adjusted ORs for test propensity and for positivity, for the same participant factors (analysis period 17 November 2020 to 18 April 2021). Light blue crosses indicate the 95% CIs in each dimension. The shaded quadrant indicates the combination of interest: lower test propensity and higher positivity. Occupation category N/A (not applicable) consists of children, students, household, unemployed and retired persons.

a positive outcome (figure 3). These factors were identified as: males, age groups 65+, 50–64 and <20 years (compared with 40–49), diabetes and occupation: sales/admin (compared with knowledge worker).

These identified risk factors are by definition associated with a higher probability of a positive outcome. We next calculated the PPV for each symptom, conditional on each of the identified participant factors (figure 4).



**Figure 4** Heatmap showing positive predictive value (PPV) per symptom, among subpopulations corresponding to each of the six identified lower test propensity/higher positivity participant factors.

For all factors, the highest PPVs were estimated for loss of taste (range of 50.0%–78.6% over the six factors) and for loss of smell (44.4%–77.8%). Other symptoms with higher PPVs were loss of appetite for males and the 50–64 and 65+ years age groups (34.5%, 38.2% and 46.2%, respectively) and fever for occupation: sales/admin and diabetics (42.2%, 35.9%). Across the subpopulations defined by our identified risk factors, very low PPVs were calculated for the mild respiratory symptoms runny nose (8.0%–14.9%), sneezing (5.4%–17.2%) and sore throat (4.0%–11.1%).

Finally, in sensitivity analysis we restricted the dataset to participants who fell within the approximate age-range of the working population, defined as 20–64 years ( $n=41\,017$  weekly surveys). An overlapping set (excluding the two extreme age groups) of low propensity/high positivity risk factors was identified (online supplemental figure S1); the main difference was that diabetes fell outside of the area of interest.

## DISCUSSION

We identified the participant characteristics male sex, younger and older age, diabetes, and the occupation category sales/administration (relative to knowledge worker) to be associated with a lower propensity to test given symptoms, but with a higher risk of a positive outcome. Among all subpopulations characterised by these factors, the symptoms loss of taste and loss of smell had the highest predictive value of a positive outcome (figure 4), confirming previous results for the general population,<sup>6,7</sup>

with loss of appetite (from age 50 years) also improving predictions.

The high OR for test positivity for the sales/admin category (relative to knowledge workers) suggests a higher exposure risk, potentially due to higher contact rates with the public, but our study cannot elucidate the cause. PPVs for individual symptoms were higher overall compared with the other identified participant factors (figure 4) due to the relatively high positivity rate for sales/admin employees.

Ever smokers and allergies/hay fever were participant characteristics that identified a low propensity to test with a relatively low risk of a positive outcome, suggesting that these persons correctly self-identified their symptoms as unlikely to be attributable to COVID-19. Interestingly, a higher propensity to test (OR of 1.45) was associated with the variable suspected non-COVID cause, the definition of which included allergy. If allergy is removed from the definition of this variable, the OR increases to 1.89, indicating that causes other than allergy drive this result. A higher willingness to be tested (or alternatively, a greater degree of cautiousness) accompanied by a low risk of positivity was observed among those with children aged <5 years in the household, age group 30–39 years, asthma sufferers and those with an underlying lung condition. Such information may not necessarily help guide testing policy, but nevertheless is valuable in understanding the participant and household factors associated with other aspects of the data.

The low odds of getting tested given symptoms if had previously tested positive (OR of 0.35) has implications for the detection of reinfections within test-based COVID-19 surveillance data. This lower test propensity could lead to severe underascertainment of reinfection, if there is a reduced willingness to get tested—even given symptoms—when one had previously tested positive.

Previous research on factors related to SARS-CoV-2 test uptake was conducted using ecological analysis to compare state-level testing rates in the USA with prevalence of underlying conditions,<sup>12</sup> which suggested associations between testing rates and certain COVID-19 risk factors. For states with higher prevalences of hypertension, diabetes or lung cancer, overall testing rates were lower and positivity rates were higher compared with states in which the prevalence of these risk factors was lower. Ecological associations between testing rate and ethnic composition, and between positivity rate and socioeconomic status were also observed at neighbourhood level in New York City.<sup>13</sup> A large study using individual-level data collected from a mobile/web app during April/May 2020 in the USA (when screening was not universally available) reported a higher odds of being tested associated with healthcare and other essential occupations, and a higher odds for ages 55–64 and 65+ years and a lower odds for females;<sup>7</sup> these age and sex findings are inconsistent with our results. Our study builds on this previous research and will be useful in contexts beyond the Netherlands.

Our approach has a number of limitations. First, position in the testing/positivity space (figure 3) for categorical variables (eg, occupation) is affected by the reference category selected. Second, as with all participatory surveillance systems, observed associations with participant or other factors and survey responses cannot distinguish the effects of the measured variable from reporting behaviour. For instance, if participants reporting underlying allergies are more prone to be aware of their symptoms (and thus report them) compared with non-allergy sufferers, associations between this risk factor and test propensity may be distorted. Third, the dependence between lower test propensity and higher positivity rate for certain groups is consistent with risk behaviours associated with both unwillingness to test and with infection, or with between-group differences in symptom severity thresholds for getting tested, or with differences in the ability to accurately assess infection risk and so avoid unnecessary testing. Our data cannot distinguish between these accounts. Fourth, volunteer bias may have influenced the absolute test propensity estimates if Infectieradar participants are more willing to engage with healthcare services compared with the general population. We did not have sufficient data to investigate whether ORs of a positive result for certain groups (such as age) changed over our study period, in line with group-specific infection risks varying over time. The 'suspected non-COVID-19 cause' variable may be affected by misclassification bias, if receipt of a positive test result influenced the response to the relevant survey question, with the consequence that the obtained OR for this variable would be overestimated.

Although those individuals who had previously tested positive appear much less likely to get tested again—even with eligible symptoms—behaviour may depend on the severity of these symptoms; we cannot easily assess this. Also, having previously tested positive might increase the chance of reporting (continued) symptoms in following surveys (ie, either because of experienced prolonged symptoms ('long-COVID-19') or because the participant's threshold for symptom reporting may be lowered due to confirmation bias), which would strengthen the inverse relation with positive test history. We note that within our data collection period, there were no restrictions regarding access to testing for symptomatic persons.

Infectieradar participants are not representative of the Netherlands population in terms of the demographic variables sex, age and education level (online supplemental table S1); we cannot determine if our findings generalise beyond the study participants. Finally, certain levels of the occupation covariate are not represented among all age categories; hence adjustment for both of these variables may lead to biased ORs, or the occupation category may be too restrictive or too heterogeneous with respect to age (eg, very few persons younger than 20 years are in the healthcare or education occupation categories; as well, occupation: N/A includes ages from infants through retirees). However, the sensitivity analysis

restricting to the working age population largely corroborated the principal results.

The level of willingness to undergo testing for SARS-CoV-2 infection in the Netherlands is undesirably low. Cross-sectional behavioural surveys demonstrated that despite 85%–89% of respondents with COVID-19-compatible symptoms being supportive of getting tested, 44%–64% nevertheless did not do so (period 10 November 2020 to 6 April 2021).<sup>14</sup> This behaviour, also observed in other countries,<sup>15</sup> occurs within the context of public health communications urging voluntary, cost-free testing, and thus constitutes a challenge for the pandemic response.

Estimation of the statistical relationships between test propensity and sociodemographic, background health status and household situation variables will help in ascertaining risk factors and thus identify subpopulations for whom testing should be facilitated or otherwise encouraged, and/or are likely under-represented in the case notification and testing surveillance systems. This analysis using participatory surveillance data has assisted in determining these potential barriers to testing. Moreover, this study has enabled identification of symptoms that are strong predictors of a positive test outcome among the identified subpopulations with higher probability of a positive result. Improving communication to the public (and thus their motivation) to present for SARS-CoV-2 testing, especially to those individuals less willing or able to do so, poses challenges for public health planning, but will be important for effective and sustainable control and management of what is expected to become an endemic infection.<sup>16</sup>

**Acknowledgements** We greatly appreciate the voluntary contributions to this research by all Infectieradar participants.

**Contributors** SAM, LCS and AJvH designed the study. CMAS advised on statistical analysis. SAM produced tables and figures and drafted the manuscript. SAM, LCS, CMAS, IF, CCvdW, AT, NN, SvdH, JW and AJvH contributed to interpretation of the data and revision of the manuscript. SAM acts as guarantor for the manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants. The research protocol was shared with the Medical Ethics Review Committee Utrecht, and an official waiver for ethical approval was obtained, which stated that ethical assessment was unnecessary given the non-invasive nature of data collection (reference number: WAG/avd/20/008757). Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. Individual-level data collected by Infectieradar cannot be publicly released for reasons of privacy protection; aggregated data that fulfill privacy requirements may be obtained by contacting the authors.

**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 iD

Scott A McDonald <http://orcid.org/0000-0003-0788-6011>

#### REFERENCES

- 1 Kucharski AJ, Klepac P, Conlan AJK, *et al*. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect Dis* 2020;20:1151–60.
- 2 Aleta A, Martín-Corral D, Pastore Y Piontti A, *et al*. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav* 2020;4:964–71.
- 3 Rubin GJ, Smith LE, Melendez-Torres GJ, *et al*. Improving adherence to ‘test, trace and isolate’. *J R Soc Med* 2020;113:335–8.
- 4 Guerrisi C, Turbelin C, Souty C, *et al*. The potential value of crowdsourced surveillance systems in supplementing sentinel influenza networks: the case of France. *Eurosurv* 2018;23:1700337.
- 5 van Noort SP, Codeço CT, Koppeschaar CE, *et al*. Ten-year performance of Influenzanet: ILI time series, risks, vaccine effects, and care-seeking behaviour. *Epidemics* 2015;13:28–36.
- 6 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.
- 7 Allen WE, Altae-Tran H, Briggs J, *et al*. Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nat Hum Behav* 2020;4:972–82.
- 8 Zens M, Brammertz A, Herpich J, *et al*. App-based tracking of self-reported COVID-19 symptoms: analysis of questionnaire data. *J Med Internet Res* 2020;22:e21956.
- 9 McDonald SA, van den Wijngaard CC, Wielders CCH, *et al*. Risk factors associated with the incidence of self-reported COVID-19-like illness: data from a web-based syndromic surveillance system in the Netherlands. *Epidemiol Infect* 2021;149:e129.
- 10 Muller CJ, MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *Int J Epidemiol* 2014;43:962–70.
- 11 R Core Team. R: a language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria, 2019. Available: <https://www.R-project.org/>
- 12 Souch JM, Cossman JS. A commentary on rural-urban disparities in COVID-19 testing rates per 100,000 and risk factors. *J Rural Health* 2021;37:188–90.
- 13 Lieberman-Cribbin W, Tuminello S, Flores RM, *et al*. Disparities in COVID-19 testing and positivity in New York City. *Am J Prev Med* 2020;59:326–32.
- 14 RIVM Behaviour Unit. Compliance and support: data until 9 March 2021. Available: <https://coronadashboard.government.nl/landelijk/gedrag> [Accessed 16 May 2021].
- 15 Smith LE, Potts HWW, Amlöt R, *et al*. Adherence to the test, trace, and isolate system in the UK: results from 37 nationally representative surveys. *BMJ* 2021;372:n608.
- 16 Lavine JS, Bjornstad ON, Antia R. Immunological characteristics govern the transition of COVID-19 to endemicity. *Science* 2021;371:741–5.