

Risk factors associated with the incidence of self-reported COVID-19-like illness: data from a web-based syndromic surveillance system in the Netherlands

Original Paper

Cite this article: McDonald SA, van den Wijngaard CC, Wielders CCH, Friesema IHM, Soetens L, Paolotti D, van den Hof S, van Hoek AJ (2021). Risk factors associated with the incidence of self-reported COVID-19-like illness: data from a web-based syndromic surveillance system in the Netherlands. *Epidemiology and Infection* **149**, e129, 1–7. <https://doi.org/10.1017/S0950268821001187>

Received: 31 December 2020

Revised: 10 May 2021

Accepted: 12 May 2021

Key words:

COVID-19; Netherlands; risk factors; syndromic surveillance

Author for correspondence:

Scott A. McDonald,

E-mail: scott.mcdonald@rivm.nl

Scott A. McDonald¹ , Cees C. van den Wijngaard¹, Cornelia C. H. Wielders¹ , Ingrid H. M. Friesema¹ , Loes Soetens¹, Daniela Paolotti², Susan van den Hof¹ and Albert Jan van Hoek¹

¹Centre for Infectious Disease Control, Netherlands National Institute for Public Health and the Environment, Bilthoven, the Netherlands and ²Institute for Scientific Interchange Foundation, Torino, Italy

Abstract

During the first wave of the severe acute respiratory syndrome-coronavirus-2 epidemic in the Netherlands, notifications consisted mostly of patients with relatively severe disease. To enable real-time monitoring of the incidence of mild coronavirus disease 2019 (COVID-19) – for which medical consultation might not be required – the Infectieradar web-based syndromic surveillance system was launched in mid-March 2020. Our aim was to quantify associations between Infectieradar participant characteristics and the incidence of self-reported COVID-19-like illness. Recruitment for this cohort study was via a web announcement. After registering, participants completed weekly questionnaires, reporting the occurrence of a set of symptoms. The incidence rate of COVID-19-like illness was estimated and multivariable Poisson regression used to estimate the relative risks associated with sociodemographic variables, lifestyle factors and pre-existing medical conditions. Between 17 March and 24 May 2020, 25 663 active participants were identified, who reported 7060 episodes of COVID-19-like illness over 131 404 person-weeks of follow-up. The incidence rate declined over the analysis period, consistent with the decline in notified cases. Male sex, age 65+ years and higher education were associated with a significantly lower COVID-19-like illness incidence rate (adjusted rate ratios (RRs) of 0.80 (95% CI 0.76–0.84), 0.77 (0.70–0.85), 0.84 (0.80–0.88), respectively) and the baseline characteristics ever-smoker, asthma, allergies, diabetes, chronic lung disease, cardiovascular disease and children in the household were associated with a higher incidence (RRs of 1.11 (1.04–1.19) to 1.69 (1.50–1.90)). Web-based syndromic surveillance has proven useful for monitoring the temporal trends in, and risk factors associated with, the incidence of mild disease. Increased relative risks observed for several patient factors could reflect a combination of exposure risk, susceptibility to infection and propensity to report symptoms.

Introduction

For monitoring the incidence and seasonal patterns of infectious diseases – particularly those with a relatively mild course not normally requiring medical consultation – syndromic surveillance is a valuable tool for public health surveillance. In the Netherlands and elsewhere, internet-based syndromic surveillance among the general population has contributed useful data on the timing and magnitude of seasonal epidemics of influenza-like illness (ILI), as well as healthcare-seeking behaviour and other factors associated with self-reported ILI [1–3]. Advantages of such internet-based syndromic over traditional surveillance systems include the potential degree of coverage – very high (voluntary) participation rates can be achieved – and the capability for real-time monitoring of epidemic growth rates and of the impact of public health interventions [4]. In the case of ILI, validation against sentinel surveillance (typically general practitioner (GP) network-based) data indicates a high degree of accuracy with respect to trends and timing [2, 5].

In the Netherlands, the mandatory notification system OSIRIS [6] collates patient information for laboratory-confirmed (PCR test positive) coronavirus disease 2019 (COVID-19) cases, of which approximately 25% (prior to 1 June 2020) consisted of relatively severe cases who were tested upon hospital admission. Because patients with a milder, yet symptomatic, disease course did not enter the testing pathway prior to the offer of universal access to testing from June 2020, risk factors for less-severe infection could not easily be assessed. The Infectieradar web-based COVID-19 syndromic surveillance system launched in mid-March 2020 offered an opportunity to address this data gap, as extensive information on patient's characteristics – sociodemographic data, underlying medical conditions and lifestyle and household factors –

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

is collected upon registration. Like the established Influenzanet system with which it is closely affiliated [7, 8], symptom occurrence and other data is elicited from voluntary participants using a survey form submitted on a weekly basis, allowing estimation of the incidence rate of COVID-19-like illness. As with all analyses in which syndromic data from participatory surveillance systems is used as a proxy for confirmed infection, it is important to recognise that discovered associations may be driven by (a combination of) the propensity to report symptoms and the occurrence of symptoms caused by the relevant infectious agent.

We posed the following research question: what are the associations between Infectieradar participant characteristics and other factors, and the incidence of self-reported COVID-19-like illness? Estimation of the relative risks of reporting COVID-19-like illness associated with basic sociodemographic, background health status and household situation variables will help in ascertaining risk factors for susceptibility to severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) infection, exposure to infectious persons, or development of symptoms. Furthermore, we investigated if the information collected by Infectieradar could permit the contribution of other potential causes of the COVID-19 symptom set (e.g. other respiratory infections, allergies) to be estimated.

Methods

Design, study period and timing of public health measures

We conducted a cohort study based on approximately two months of Infectieradar data collected from its launch on 17 March through 24 May 2020; this range covers ISO 8601 week numbers 12–21 of this year. In the Netherlands, national-level public health measures to control the spread of SARS-CoV-2 infection were imposed starting on 12th March, although a subset of measures in one province (North Brabant) were implemented a week earlier. Initial measures consisted of advice to stay at home for those with even mild symptoms, for everyone to stay home as much as possible (with strong encouragement of home-working) and banning of large gatherings. A more stringent lock-down began on the evening of 15th March, with the closure of all catering establishments (other than take-aways), theatres, cinemas, gyms etc, and the advice to maintain 1.5 m distance from persons outside one's household members. Schools and daycares were closed on the following day. The study period end (week 21, 18–24 May; symptom onsets to end week 20 (see below)) was chosen to encompass a relatively homogenous period before public health measures were relaxed. Namely, from the 11th of May primary schools re-opened and 'public contact' occupations (e.g. dentists, hairdressers) were allowed to resume practice, and then on the 1st of June further measures were relaxed (such as the partial opening of cafes and restaurants).

Data source and study population

Infectieradar is a member of the Influenzanet consortium, a European collaboration involving universities and public health institutes that has been focussed on collecting self-reported data on viral infections, such as influenza-like illness, since 2003 [7–9].

Recruitment of volunteers for the Infectieradar surveillance system was via a web announcement, and registration was open to all residents of the Netherlands who were aged 16 years or older (children under 16 years could participate under

supervision of their legal guardian, or if their legal guardian acted on their behalf). Upon registration, participants filled out an intake questionnaire, in which sociodemographic data (age, sex, education level, occupation, partial postal code, number of people in household, etc.) and various aspects of their medical history (e.g. pre-existing health conditions such as allergies/hay fever and chronic diseases) are elicited. These factors were selected to be consistent with previous research by the Influenzanet consortium [2]. Following completion of the intake survey, participants were also requested to fill out a standard questionnaire (and every week thereafter), to provide details of any symptoms experienced during the past week. Specifically, in each weekly survey participants were asked if they had experienced one or more of a set of 19 symptoms (fever, chills, runny or blocked 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, stomachache, loss of sense of smell/taste, other) within the 7-day period prior to reporting, and were also asked to provide a date of symptom onset. Additional weekly survey data collected included information regarding healthcare-seeking behaviour and suspected cause of symptom(s) (if reported). All data were pseudonymised before analysis, with individual participants assigned a unique identifier. Because the initially registered cohort of participants (on 17 March) was supplemented by further registrations during the study period, we used analysis methods suitable for this 'open' cohort.

Data inclusion/exclusion criteria

To minimise the potential impact of selection bias (persons who might have registered to participate and/or only participated once, because they currently or recently experienced symptoms), only those participants who had contributed two or more weekly surveys were retained; these participants were deemed 'active'. In addition, the first weekly survey submitted by each participant was removed. Next, all weekly surveys in which symptom onset was indicated as prior to the date of intake survey, subsequent to the date that the weekly survey was submitted, or >8 days prior to the weekly survey date were excluded. Finally, records for persons with age indicated as <15 years or >100 years (due to low numbers), or with age missing, were removed. The remaining participants were deemed the 'eligible' participants.

Definition of COVID-19-like illness

Following the case definition used by the Influenzanet consortium [10] when Infectieradar was launched (mid-March 2020), COVID-19-like illness was defined as the reporting of: fever (≥ 37.5 C) and/or cough, and at least one other symptom from the set provided in the weekly survey.

Calculation of incidence

We defined a multi-week episode if a participant had reported COVID-19-like illness in two or more contiguous weekly surveys, with no more than eight days between survey dates. The incidence rate for a given week (i.e. the week prior to the week in which the survey was completed) was then defined as the number of episodes of reported COVID-19-like illness divided by the number of eligible participants in that week (i.e. persons at risk: the number of participants returning a weekly survey that was not

indicated as a continuation of the same episode of illness). Thus, the natural denominator for the incidence rate is person-weeks; the incidence rate can be converted to events per person-year by dividing the denominator by 52.

Statistical analyses

Analysis of factors associated with COVID-19-like illness

A multivariable Poisson regression model with log-link was fitted to model the outcome, COVID-19-like illness incidence, as a function of baseline participant characteristics and other factors. This enabled the relative risks (as rate ratios (RRs)) associated with this set of covariates to be estimated. From the intake survey, the following baseline characteristics/factors were investigated: *sex*, *age-group* (15–24, 24–34, 35–44, 45–54, 55–64; 65+ years), *province of residence*, *educational level* (none/primary only, secondary/technical, university), *presence of children aged <5 years in the household* (binary variable), *presence of children aged 5–18 years in the household* (binary variable), *smoker status* (non-smoker, ever-smoker), selected underlying chronic conditions (*lung disease* (e.g. emphysema, COPD), *cardiovascular disease*, *diabetes*; all coded as binary variables) and *hay fever and/or other allergies* at baseline (binary variable). All covariates were included in the multivariate model.

We also investigated the interaction terms *age-group* × *children <5 years in household* and *age-group* × *children 5–18 years in household*, as any effect of age might depend on the presence of children (reflecting possible differences in risk between parents, and persons of the same age without children living with them).

Finally, a natural cubic regression spline with three knots was fitted to week number, to capture the temporal trend in the rate of reported COVID-19-like illness. As more than one episode matching the syndrome definition could be reported by an individual over the study period, we used generalised estimating equations (GEE) Poisson regression approach to account for non-independence (clustering) in the dataset.

In additional analysis, we investigated the distribution of ‘suspected non-COVID cause’, a new variable constructed from the multiple-choice question in the weekly survey ‘Do you have any idea what caused your symptoms?’ This variable was defined as ‘Yes’ if any of the answers ‘flu/flu-like illness’, ‘common cold’ or ‘allergies/hay fever/asthma’, ‘gastrological complaints/stomach flu’ or ‘other’ were selected and as ‘No’ otherwise (thus ‘No’ included the responses ‘coronavirus (COVID-19)’ and ‘don’t know’).

Sensitivity analysis: selected COVID-19-like illness case definition

It is becoming clearer that COVID-19 patients present with a wide range of symptoms, combinations of which correspond to widely varying degrees of severity and prognoses [11]. The case definition we used – aimed at high sensitivity – is rather non-specific for COVID-19, in that it can also match the symptom manifestations for other respiratory infections or conditions such as hay fever. Therefore, in sensitivity analysis we repeated the main analysis using a ‘more specific’ case definition, which was derived using decision tree analysis (for details see Supplementary Materials). With this definition, COVID-19-like illness was defined as: fever and (loss of smell/taste or dyspnoea), or (dyspnoea and chest pain), or (loss of smell/taste and malaise). As usual, the trade-off with improving specificity means that a higher

proportion of true COVID-19 cases will be missed. Although COVID-19-like illness incidence rates are anticipated to be lower using this definition, estimated incidence RRs should be comparable.

Comparison to notified case time series

All laboratory-confirmed COVID-19 cases are reported to the Netherlands national mandatory notification system OSIRIS [6]; these records also contain reported date of symptom onset. We retrieved the total number of notified cases (according to symptom onset date) per week of the study period, excluding the cases for which symptom onset date was missing.

All analyses were conducted using R statistical software, version 3.6.0 [12].

Results

Within our study period (17 March through 24 May (end of week 21) of 2020), a total of 44 914 persons had registered and filled out the intake survey. Of these, 1025 (2.3%) did not return any weekly surveys, 16 561 (36.9%) had returned only one weekly survey, and 27 328 persons (60.8%) had returned two or more weekly surveys by the end of week 21; thus there were a total of 27 328 ‘active’ participants. After applying inclusion/exclusion criteria to the weekly surveys submitted by the active participants, among the remaining (‘eligible’) participants ($n = 25\,663$) there were a total of 7060 episodes of reported COVID-19-like illness among 5196 participants with onset in week 12–20 (representing 131 404 person-weeks, or 2527 person-years of follow-up). Two or more COVID-19-like illness episodes were reported by 1287 participants (24.8% of the 5196 participants who reported at least one COVID-19-like illness episode within this period).

Basic demographic information is presented in Table 1. There was a higher proportion of female (61%) than male participants, and the majority (51.3%) was 45 years or older, with substantial underrepresentation of the 15–24-year age group, who represented 4.3% of all participants. Fifty-seven per cent held a university degree. The distributions for these characteristics differed from the distributions for the Netherlands population (Supplementary Materials, Table S3). Eight per cent of participants had indicated asthma at baseline and 37.5% reported suffering from allergies (including hay fever). The crude incidence of reported COVID-19-like illness decreased over the study period, from 229 per 1000 person-weeks in week 12, to 14 per 1000 person-weeks in week 20 (Fig. 1).

Incidence RRs for participant and other factors

Table 2 shows the estimated relative risks (as RRs) associated with baseline participant characteristics and other factors from the multivariable Poisson regression. Adjusting for all covariates including province of residence and the time trend, the age group 65+ years was associated with a lower incidence rate (RR of 0.77, 95% CI 0.70–0.85), compared with the reference category of 35–44 years. Incidence was also lower for males (RR = 0.80, 95% CI 0.76–0.84). Education level was associated with COVID-19-like illness incidence: persons with higher (university-level) education qualifications had a reduced risk (RR = 0.84, 95% CI 0.80–0.88); those with no/lower education had a raised risk (RR = 1.13, 95% CI 1.01–1.27). Ever smoker, asthma, allergies, diabetes, chronic lung disease and cardiovascular disease as

Table 1. Baseline characteristics of the eligible study participants (the Netherlands, 17 March through 24 May 2020; *n* = 25 663)

Characteristic	<i>N</i>	Percentage
Reported 0 COVID-19-like illness episodes	20 467	79.8
Reported 1 COVID-19-like illness episode	3909	15.2
Reported ≥ 2 COVID-19-like illness episode	1287	5.0
Sex		
Male	9905	38.6
Female	15 758	61.4
Age group (years)		
15–24	1112	4.3
25–34	5159	20.1
35–44	6202	24.2
45–54	5786	22.6
55–64	4468	17.4
65+	2936	11.4
Region ^a		
West	11 917	46.4
South	6388	24.9
East	5052	19.7
North	2306	9.0
(missing)	449	
Education level		
None/primary	961	3.8
Middle	9997	39.0
Higher	14 705	57.3
Ever smoker	4137	16.1
Asthma	1979	7.7
Allergy(s)/hay fever	9618	37.5
Diabetes	681	2.7
Chronic lung disease	567	2.2
Cardiovascular disease	1419	5.5
1+ children <5 years in household	3818	14.9
1+ children 5–18 years in household	21 891	85.3

^aRegion *West* consists of the provinces North Holland, South Holland, Utrecht, Zeeland; *South*: North Brabant, Limburg; *East*: Flevoland, Gelderland, Overijssel; *North*: Drenthe, Friesland, Groningen.

reported at baseline were all associated with a significantly higher incidence rate (RRs of 1.36 (1.28–1.44), 1.65 (1.54–1.77), 1.37 (1.31–1.44), 1.36 (1.20–1.54), 1.69 (1.33–1.60), 1.46 (1.33–1.60), respectively), as was the presence of young children (<5 years of age) (RR = 1.11, 95% CI 1.04–1.19) or older children (5–18 years) in the household (RR = 1.39, 95% CI 1.30–1.48).

Although the crude incidence was higher among participants in the 15–24 and 25–34 years age groups with one or more children <5 years, compared to no children <5 years, in the household (Supplementary Fig. S1), specifying either the interaction term *age-group* \times *children <5 years in household* or the interaction *age-group* \times *children 5–18 years in household* failed to improve regression model fit.

Sensitivity analysis

Using the ‘more specific’ COVID-19-like illness case definition, 1118 episodes were reported in the study period, with a total of 134 032 person-weeks, or 2578 person-years of follow-up. More than one episode was reported by only 126 (13.1%) of participants who reported at least one episode. Crude incidence decreased from 46.4 to 2.6 per 1000 person-weeks between weeks 12 and 20 (Supplementary Fig. S2). The estimated incidence RRs were highly similar to those obtained using the COVID-19-like illness case definition (Supplementary Table S1), with the exception of the smaller, and non-significant RRs for the presence of children in the household. As well as, the RRs for several underlying chronic conditions were estimated to be higher (RRs of 2.46 (95% CI 2.09–2.89), 2.17 (95% CI 1.62–2.90), 2.36 (95% CI 1.91–2.93) for asthma, chronic lung disease and cardiovascular disease, respectively) compared with the main analysis.

Fifty-five per cent of all COVID-19-like illness (main case definition) episodes were self-reported as due to a ‘suspected non-COVID cause’ (Supplementary Table S2). Using the more specific COVID-19-like illness case definition, 34% of episodes were indicated as due to a ‘suspected non-COVID cause’.

Notified laboratory-confirmed COVID-19 cases, summed by ISO week number according to first date of symptoms (and excluding cases with missing first symptom date), decreased by 83%: from 34 cases per 100 000 population in week 12, to 5.8 cases per 100 000 with symptom onset in week 20. To compare, the incidence rate of COVID-19-like illness decreased by 94% – from 440 to 28 per 100 000 person-years – over the same period.

Discussion

This web-based syndromic surveillance system had a high participation rate over the analysis time frame, a reasonable regional and demographic coverage of the Netherlands, and yielded a large amount of data that is informative for the incidence of, and risk factors for, symptoms characteristic of SARS-CoV-2 infection.

The significantly reduced incidence RRs associated with being university educated could be related to relatively less exposure to SARS-CoV-2 transmission situations (e.g. by working more often at home), as the study period almost completely overlapped with the lockdown period, or may be related to better health-related outcomes and behaviours in general. The reduced RR for 65+ years could be due to Infectieradar participants in this age group tending to be more active and in better health compared with the general population, and/or this age group have been more stringent in avoiding transmission situations; we have no data to support either possibility. Raised RRs of COVID-19-like illness observed for smokers, none/primary education level, underlying medical conditions, and for the presence of children in the household could reflect a combination of susceptibility to infection and/or development of symptoms (including those caused by other respiratory infections), heightened transmission risks (exposure) and the propensity to report symptoms. Males had a significantly lower COVID-19-like illness incidence rate (RR = 0.80) compared with females, which contrasts with the null association between sex and seroprevalence observed among 3100 participants of a population-level serostudy in the Netherlands (odds ratio = 0.87; 95% CI 0.53–1.41) [13]. Our result notably differed from the reported 1.6-fold increased risk for males in a study cohort of 17 million primary care patients in England [14], but this was for the aggregate outcome

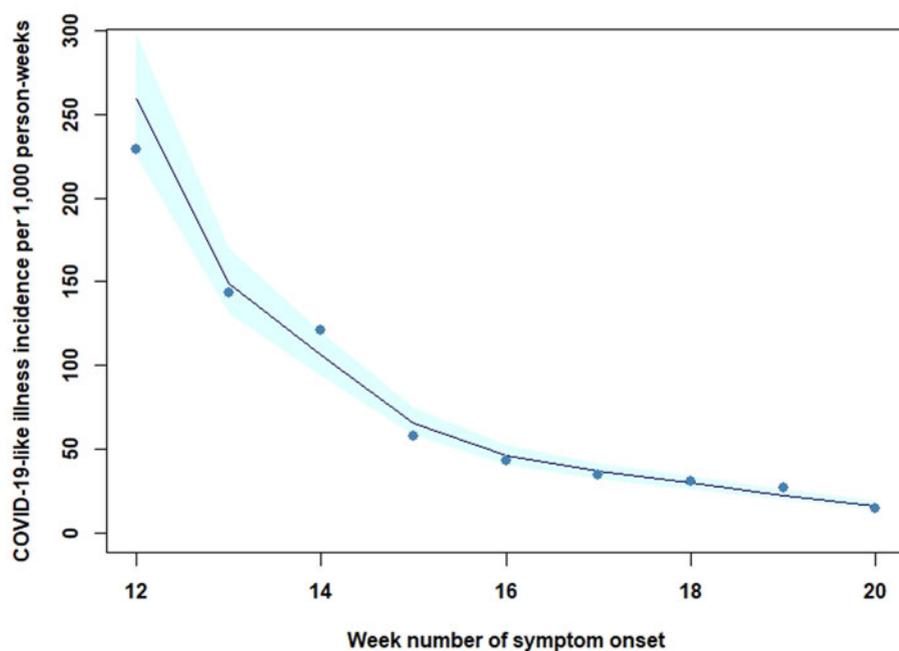


Fig. 1. Fitted Poisson regression of the incidence of self-reported COVID-19-like illness over time, the Netherlands, 17 March through 24 May 2020. Incidence is shown for ISO 8601 week numbers 12 through 20. Points indicate the mean observed weekly incidence rate. Shaded band indicates 95% confidence interval around fitted value.

SARS-CoV-2 infection followed by in-hospital mortality, whereas the cases of COVID-19-like illness in our analysis comprised mild SARS-CoV-2 and other respiratory infections. This comparison is further complicated by the different study bases; only a small percentage of Infectieradar participants with COVID-19-like illness reported consulting their GP.

Interestingly, our estimated RRs for several participant factors were highly similar to findings derived from internet-based syndromic surveillance ILI data from four countries (Netherlands, Belgium, Portugal, Italy) over the 10 seasons from 2003 until 2013 [2]. For instance, the occurrence of at least one ILI episode in a season was found to be significantly associated with (among other factors): male sex (risk ratio = 0.82), smoking (RR = 1.16), asthma or lung disease (RR = 1.58) and the presence of children in the household (RR = 1.31). The concordance between relative risks estimated for ILI in this study with those for COVID-19-like illness in the current study, particularly for sex, chronic conditions and smoker status, may reflect a dependence in susceptibility to respiratory infection in general, to symptom reporting propensity, and the fact that the case definition for COVID-19-like illness is consistent with infection caused by other respiratory agents.

The spline term in the Poisson regression analysis was included to capture variation in the reported COVID-19-like illness incidence rate over time. This could represent a combination of temporal variation due to the actual incidence of SARS-CoV-2 infection, and temporal variation in other contributors to the occurrence of overlapping symptoms (e.g. hay fever). Analytic techniques such as ecological regression might prove useful for attributing symptoms between competing causes [15, 16].

Adopting the more specific case definition for COVID-19-like illness led to a lower incidence rate overall, but with a similar temporal trend as for the standard case definition (Supplementary Fig. S2). This sensitivity analysis also yielded comparable adjusted incidence RR estimates for most participant characteristics (Supplementary Table S1), with the main exception being lower, and non-significant, associations with the presence of children

aged <5 years or 5–18 years in the household. The higher RRs estimated for chronic conditions compared with the main analysis suggest that persons with pre-existing health conditions may be more susceptible to SARS-CoV-2 infection than to a non-specific respiratory syndrome, but this hypothesis would need to be investigated further using an appropriate study design.

Roughly paralleling the time series of notified laboratory-confirmed COVID-19 cases – according to symptom onset date – over the study period, the incidence of COVID-19-like illness as captured by the Infectieradar surveillance system also showed a decreasing trend. The main difference was that the frequency of symptom onset of the notified cases peaked in week 14, but the highest COVID-19-like illness incidence rate was in week 12; this discordance is difficult to explain. Note that because mandatory notifications within our analysis period consisted of the more severe cases (one-quarter of whom were PCR-tested upon or after admission to hospital), the magnitude of the notification rate cannot be directly compared with the COVID-19-like illness incidence rate derived from Infectieradar.

We now consider limitations of this study. First, analyses of self-reported symptom occurrence cannot distinguish association between a given variable and risk of COVID-19-like illness, from the propensity of reporting; for instance from the present data it cannot be concluded that there is an increased risk of syndrome occurrence due to asthma, or whether participants suffering from asthma are more likely to be aware of their symptoms and report them. A second limitation concerns the sensitivity and specificity of the COVID-19 case definition. Persons with SARS-CoV-2 infection who did not report fever or cough – false negatives – will not contribute to the estimated incidence. Symptoms fitting the case definition are also characteristic of other respiratory infections and of conditions such as hay fever; hence, an unavoidable proportion of reports are false positives. From the Netherlands' long-standing GP-based sentinel surveillance system, in which a sample of patients with influenza-like illness or acute respiratory symptoms are routinely swabbed, a decreasing proportion of rhinovirus positivity is apparent during

Table 2. Association of patient characteristics and other factors with the incidence of self-reported COVID-19-like illness in the Netherlands

Factor	<i>n</i>	Person-weeks	Incidence rate	Unadjusted RR (95% CI)	Adjusted RR ^a (95% CI)
(All)	7060	131 404	54	–	–
Sex					
Male	2403	52 556	46	0.78 (0.74–0.82)	0.80 (0.76–0.84)
Female	4657	78 848	59	Ref.	Ref.
Age group (years)					
15–24	280	4878	57	1.01 (0.89–1.15)	0.93 (0.82–1.06)
25–34	1344	24 119	56	0.97 (0.91–1.04)	1.01 (0.94–1.08)
35–44	1753	30 804	57	Ref.	Ref.
45–54	1686	30 094	56	0.99 (0.93–1.06)	1.01 (0.94–1.08)
55–64	1344	30 804	54	0.96 (0.89–1.03)	1.00 (0.93–1.08)
65+	663	16 833	39	0.70 (0.64–0.77)	0.77 (0.70–0.85)
Education level					
None/primary	338	4685	72	1.18 (1.05–1.32)	1.13 (1.01–1.27)
Middle	3047	49 887	61	Ref.	Ref.
Higher	3675	76 832	48	0.78 (0.74–0.82)	0.84 (0.80–0.88)
Ever smoker	1426	19 034	50	1.50 (1.41–1.59)	1.36 (1.28–1.44)
Pre-existing health condition					
Asthma	948	8920	106	2.15 (2.01–2.30)	1.65 (1.54–1.77)
Allergy(s)/hay fever	3232	46 680	69	1.53 (1.46–1.60)	1.37 (1.31–1.44)
Diabetes	250	3590	70	1.33 (1.17–1.50)	1.36 (1.20–1.54)
Chronic lung disease	308	2808	110	2.09 (1.86–2.34)	1.69 (1.50–1.90)
Cardiovascular disease	540	7513	72	1.36 (1.24–1.49)	1.46 (1.33–1.60)
1+ children <5 years in household	1059	18 705	57	1.06 (0.99–1.13)	1.11 (1.04–1.19)
1+ children 5–18 years in household	3915	109 678	52	0.82 (0.77–0.87)	1.39 (1.30–1.48)

CI, confidence interval.

^aAdjusted for all covariates, including province of residence (not shown in the table).

Associations with incidence were estimated as RRs using multivariable Poisson regression analysis and the generalised estimating equation approach. The numerator (*n*) for the incidence rate (per 1000 person-weeks) is the number of self-reported episodes of symptoms matching the COVID-19-like illness case definition; the denominator is the number of person-weeks at risk.

the study period, from about 15% of samples in week 12 to 0% in week 17 [17]. As a further indication of low specificity of the case definition for true COVID-19, 55% of participants reported that they thought their COVID-19-like illness was due to a non-COVID cause (34% for the more specific case definition); however, the reliability of self-reported cause of symptom(s) is not known.

Third, although statistical adjustment could address differences in the distributions over sex, age and education level between the Infectieradar participants and the general population, we could not investigate (nor adjust for) socioeconomic factors that may be associated with SARS-CoV-2 infection [14] (although multivariable adjustment for education level will partially address potential confounding by socioeconomic status). In addition, to reduce selection bias we excluded participants who returned fewer than two weekly surveys; if the risk of COVID-19-like symptoms differed between those excluded and the eligible study population, then estimated risk factor associations might correspond more closely to persons who are more motivated to participate. Finally, access to the internet is a prerequisite for participation, and this may constitute a source of bias particularly for the oldest age group.

Limitations to the implementation of internet-based participatory surveillance platforms at the population level include the need for diverse technical skills among the national teams who manage the systems and the resources to sustain maintenance in the long term. Concerning the recruitment and retention of participants, the strong willingness for engagement observed across existing national systems confirms the feasibility of the approach [18].

Conclusions

The Infectieradar syndromic surveillance system has proven to be useful for monitoring the national-level incidence of symptoms associated with SARS-CoV-2 infection during the period of lockdown in the spring of 2020 in the Netherlands, and for determining the associations between self-reported COVID-19-like illness occurrence and sociodemographic variables, pre-existing health conditions and other factors. Future work will continue efforts to improve the specificity of the symptom reports, and so complete other surveillance data sources in providing a more accurate picture of patient risk factors and the time course and intensity of the epidemic in the Netherlands.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268821001187>

Acknowledgements. We are grateful to Clément Turbelin for his patient assistance in setting up the ifnBase analysis software, and to Maurice van Heuveln for extensive software support for the delivery and processing of online surveys.

Financial support. This work was supported by the EU Horizon 2020 Research and Innovations Programme project EpiPose (Epidemic Intelligence to Minimize COVID-19's Public Health, Societal and Economical Impact, No 101003688).

Conflict of interest. None declared.

Ethical standards. The research protocol was shared with the Medical Ethics Review Committee Utrecht, and an official waiver for ethical approval (reference number: WAG/avd/20/008757) was obtained given the non-invasive nature of data collection.

Data availability statement. 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.

References

1. Friesema IHM *et al.* (2009) Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands. *Vaccine* **27**, 6353–6357.
2. Van Noort SP *et al.* (2015) Ten-year performance of Influenzanet: ILI time series, risks, vaccine effects, and care-seeking behaviour. *Epidemics* **13**, 28–36.
3. Peppas M, Edmunds WJ and Funk S (2017) Disease severity determines health-seeking behaviour amongst individuals with influenza-like illness in an internet-based cohort. *BMC Infectious Diseases* **17**, 238.
4. Guerrisi C *et al.* (2018) The potential value of crowdsourced surveillance systems in supplementing sentinel influenza networks: the case of France. *Eurosurveillance* **23**, 1700337.
5. Vandendijck Y, Faes C and Hens N (2013) Eight years of the Great Influenza Survey to monitor influenza-like illness in Flanders. *PLoS One* **8**, e64156.
6. Ward M *et al.* (2005) Electronic reporting improves timeliness and completeness of infectious disease notification, The Netherlands, 2003. *Eurosurveillance* **10**, 7–8.
7. Paolotti D *et al.* (2014) Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. *Clinical Microbiology and Infection* **20**, 17–21.
8. Koppeschaar CE *et al.* (2017) Influenzanet: citizens among 10 countries collaborating to monitor influenza in Europe. *JMIR Public Health Surveillance* **3**, e66.
9. Van Noort SP *et al.* (2007) Gripenet: an internet-based system to monitor influenza-like illness uniformly across Europe. *Eurosurveillance* **12**, 5–6.
10. Borges do Nascimento IJ *et al.* (2020) Novel coronavirus infection (COVID-19) in humans: a scoping review and meta-analysis. *Journal of Clinical Medicine* **9**, 941.
11. Sudre CH *et al.* (2021) Symptom clusters in COVID-19: a potential clinical prediction tool from the COVID symptom study app. *Science Advances* **7**, eabd4177.
12. R Core Team (2019) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at <https://www.R-project.org/>.
13. Vos RA *et al.* (2020) Nationwide seroprevalence of SARS-CoV-2 and identification of risk factors in the general population of the Netherlands during the first epidemic. *Journal of Epidemiology and Community Health* **75**, 489–495.
14. Williamson EJ *et al.* (2020) OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature* **84**, 430–436.
15. Van Asten L *et al.* (2012) Mortality attributable to 9 common infections: significant effect of influenza A, respiratory syncytial virus, influenza B, norovirus, and parainfluenza in elderly persons. *Journal of Infectious Diseases* **206**, 628–639.
16. Hardelid P, Pebody R and Andrews N (2013) Mortality caused by influenza and respiratory syncytial virus by age group in England and Wales 1999–2010. *Influenza and Other Respiratory Viruses* **7**, 35–45.
17. Netherlands Institute for Health Research (Nivel) (2020) Nivel Surveillance Bulletin 2020 week 29. 22 July 2020. Utrecht: Nivel. https://www.nivel.nl/sites/default/files/algemene-content/NivelSurveillanceSamenvatting_2020_wk29b_0.pdf.
18. Bajardi P *et al.* (2014) Determinants of follow-up participation in the Internet-based European influenza surveillance platform Influenzanet. *Journal of Medical Internet Research* **16**, e78.