*Epidemiology and Infection*

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

Scott A McDonald1\*, Cees C van den Wijngaard1, Cornelia CH Wielders1, Ingrid HM Friesema1, Loes Soetens1, Daniela Paolotti2, Susan van den Hof1, Albert Jan van Hoek1

1 Centre for Infectious Disease Control, Netherlands National Institute for Public Health and the Environment, Bilthoven, the Netherlands

2 Institute for Scientific Interchange Foundation, Torino, Italy

**SUPPLEMENTARY MATERIALS**

**Description of how the 'more-specific' COVID-19-like illness case definition
was derived**

The more-specific case defintition was: **“**fever and [loss of smell/taste or dyspnea], or [dyspnea and chest pain], or [loss of smell/taste and malaise]”

This definition was obtained through a decision tree analysis of those participants who were likely to be infected with SARS-CoV-2 and those who were not.

When people reported symptom(s) in the weekly survey, the following question was also asked, which had six possible answers:

*Do you think the symptoms you report today are caused by COVID-19?*

1. No, I don’t think my symptoms are caused by COVID-19
2. I don’t know
3. Yes, I think so, as the symptoms are compatible with COVID-19, however nobody in my surroundings tested positive
4. Yes, I think so, as people I interact with have had contact with a confirmed case
5. Yes, I know so, as I have direct contact with a confirmed case
6. Yes, I know so, as I am a confirmed case

Based on the answer to this question we created two groups: those who indicated they are not likely to be infected with SARS-CoV-2 (i.e., answered a)), and those who indicated they are likely to be infected with SARS-CoV-2 (i.e., answered d), e) or f)). The decision tree method identified combinations of symptoms that could best differentiate these two groups. We then selected three combinations that achieved good specificity and sensitivity, using answer f) as a reference.

**Table S1.** Association of patient characteristics and other factors with the incidence of self-reported COVID-19-like illness using the 'more specific' case definition. The numerator *(n)* for the incidence rate (per 1,000 person-weeks) is the number of self-reported episodes of symptoms matching the case definition.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Factor** | **n** | **Person-weeks** | **Incidence rate** | **Unadjusted****Rate Ratio** | **95% CI** | **Adjusted Rate Ratio\*** | **95% CI** |
| (All) | 1118 | 134032 | 8 |  |  | -- | --  |
| *Sex* |  |  |  |  |  |  |  |
|  Male  | 339 | 53475 | 6 | 0.66 | 0.58-0.75 | 0.71 | 0.62-0.81 |
|  Female | 779 | 80557 | 10 | Ref. |  | Ref. | --  |
| *Age-group* |  |  |  |  |  |  |  |
|  15-24 | 52 | 4961 | 10 | 1.06 | 0.78-1.43 | 0.87 | 0.64-1.18 |
|  25-34 | 243 | 24562 | 10 | 1.00 | 0.85-1.19 | 1.02 | 0.85-1.21 |
|  35-44 | 304 | 31398 | 10 | Ref. |  | Ref. | --  |
|  45-54 | 293 | 30748 | 10 | 0.99 | 0.84-1.16 | 0.93 | 0.78-1.10 |
|  55-64 | 167 | 25218 | 7 | 0.68 | 0.45-0.82 | 0.61 | 0.50-0.74 |
|  65+ | 59 | 17145 | 3 | 0.35 | 0.26-0.47 | 0.32 | 0.23-0.42 |
| *Education level* |  |  |  |  |  |  |  |
| None/lower | 62 | 4802 | 13 | 1.25 | 0.95-1.65 | 1.14 | 0.87-1.51 |
| Middle | 531 | 51005 | 10 | Ref. |  | Ref. | --  |
| Higher | 525 | 78225 | 7 | 0.65 | 0.57-0.73 | 0.68 | 0.60-0.78 |
| Ever smoker | 241 | 19549 | 12 | 1.59 | 1.37-1.84 | 1.29 | 1.11-1.50 |
| Asthma | 219 | 9308 | 24 | 3.31 | 2.85-3.84 | 2.46 | 2.09-2.89 |
| Allergy(s)/hay fever | 531 | 48006 | 11 | 1.62 | 1.44-1.83 | 1.28 | 1.13-1.45 |
| Diabetes | 37 | 3666 | 10 | 1.23 | 0.87-1.73 | 1.36 | 0.96-1.92 |
| Chronic lung disease | 64 | 2922 | 22 | 2.72 | 2.07-3.57 | 2.17 | 1.62-2.90 |
| Cardiovascular disease | 112 | 7759 | 14 | 1.82 | 1.49-2.21 | 2.36 | 1.91-2.93 |
| 1+ children <5 years in household | 168 | 19061 | 9 | 1.07 | 0.91-1.26 | 1.06 | 0.89-1.26 |
| 1+ children 5-18 years in household | 851 | 111747 | 8 | 0.65 | 0.56-0.74 | 1.13 | 0.96-1.33 |

\* Adjusted for all covariates, including province of residence (not shown in the Table). CI = confidence interval

**Table S2.** Distribution of self-reported suspected non-COVID cause among the number of self-reported episodes (*N*) with symptoms matching the COVID-19-like illness case definition.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Case definition** | ***N*** | **Self-reported suspected cause** | ***n*** | **(%)** |
| Reporting of: fever (≥37.5 C) and/or cough, and at least one other symptom | 7060 | *Non-COVID cause* | 3915 | 55.5 |
|  | *COVID-19* | 370 | 5.2 |
|  | *Do not know* | 2575 | 36.5 |
| 'More specific' case definition | 1118 | *Non-COVID cause* | 385 | 34.4 |
|  | *COVID-19* | 194 | 17.4 |
|  | *Do not know* | 515 | 46.0 |

**Table S3.** Distribution of sex, age-group, and education level, comparing the study population (as shown in Table 1, main article) with the general population (data for the year 2020 from Statistics Netherlands [1,2]).

|  |  |  |
| --- | --- | --- |
| ***Variable* Group** | **Study population distribution (%)** | **Netherlands population distribution (%)** |
| *Age-group* 15-24 years 25-34 35-44 45-54 55-64 65+*Sex (15 years and older)* Male Female*Education level (15 years and older)* None/primary Middle Higher | 4.320.124.222.617.411.438.661.43.839.057.3 | 14.615.314.116.826.123.149.450.630.936.732.4 |

1. Statistics Netherlands. Population; sex, age and martial status. <https://opendata.cbs.nl/statline/#/CBS/en/dataset/7461eng/table?ts=1615898674521>. Accessed 15 March 2021.

2. Statistics Netherlands. Bevolking; onderwijsniveau; geslacht, leeftijd en migratieachtergrond. <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82275NED/table?ts=1615899135543>. Accessed 15 March 2021.

**Fig. S1**. Crude incidence of COVID-19-like illness (per 1,000 person-weeks) as function of age-group and the presence of children <5 years and 5-18 years in the household (data aggregated over ISO weeknumbers 12-20 of the year 2020).

**Fig. S2.** Fitted Poisson regression model of the incidence of self-reported COVID-19-like illness using the 'more-specific' case definition over time.