Journal of Clinical Epidemiology 139 (2021) 240–254

Journal of Clinical Epidemiology

ENE-COVID nationwide serosurvey served to characterize asymptomatic infections and to develop a symptom-based risk score to predict COVID-19

Beatriz Pérez-Gómez[a,b,†], Roberto Pastor-Barriuso[a,b,†], Mayte Pérez-Olmeda[c], Miguel A Hernán[d], Jesús Oteo-Iglesias[c,e], Nerea Fernández de Larrea[a,b], Aurora Fernández-García[b,c], Mariano Martín[f], Pablo Fernández-Navarro[a,b], Israel Cruz[g], Jose L Sanmartín[f], Jose León Paniagua[h], Juan F Muñoz-Montalvo[f], Faustino Blanco[f], Raquel Yotti[h,‡], Marina Pollán[a,b,‡,*], on behalf of the ENE-COVID Study Group[§]

[a] *National Centre for Epidemiology, Instituto de Salud Carlos III (ISCIII), Monforte de Lemos 5, 28029 Madrid, Spain*
[b] *Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III (ISCIII), Monforte de Lemos 5, 28029 Madrid, Spain*
[c] *National Centre for Microbiology, Instituto de Salud Carlos III (ISCIII), Ctra de Pozuelo 28, 28222 Madrid, Spain*
[d] *Departments of Epidemiology and Biostatistics, Harvard T H Chan School of Public Health, Harvard-MIT Division of Health Sciences and Technology, 677 Huntington Ave, Boston, MA 02115, USA*
[e] *Spanish Network for Research in Infectious Diseases (REIPI), Instituto de Salud Carlos III (ISCIII), Monforte de Lemos 5, 28029 Madrid, Spain*
[f] *Deputy Directorate of Information Technologies, Ministry of Health, Paseo del Prado 18, 28014 Madrid, Spain*
[g] *National School of Public Health, Instituto de Salud Carlos III (ISCIII), Monforte de Lemos 5, 28029 Madrid, Spain*
[h] *Instituto de Salud Carlos III (ISCIII), Monforte de Lemos 5, 28029 Madrid, Spain*

Received 9 April 2021; Received in revised form 18 May 2021; Accepted 3 June 2021; Available online 11 June 2021

Abstract

Objectives: To characterize asymptomatic SARS-CoV-2 infections and develop a symptom-based risk score useful in primary healthcare.

Study design and setting: Sixty-one thousand ninty-two community-dwelling participants in a nationwide population-based serosurvey completed a questionnaire on COVID-19 symptoms and received an immunoassay for SARS-CoV-2 IgG antibodies between April 27 and June 22, 2020. Standardized prevalence ratios for asymptomatic infection were estimated across participant characteristics. We constructed a symptom-based risk score and evaluated its ability to predict SARS-CoV-2 infection.

Results: Of all, 28.7% of infections were asymptomatic (95% CI 26.1–31.4%). Standardized asymptomatic prevalence ratios were 1.19 (1.02–1.40) for men vs. women, 1.82 (1.33–2.50) and 1.45 (0.96–2.18) for individuals <20 and ≥80 years vs. those aged 40–59, 1.27 (1.03–1.55) for smokers vs. nonsmokers, and 1.91 (1.59–2.29) for individuals without vs. with case contact. In symptomatic population, a symptom-based score (weights: severe tiredness = 1; absence of sore throat = 1; fever = 2; anosmia/ageusia = 5) reached standardized seroprevalence ratio of 8.71 (7.37–10.3), discrimination index of 0.79 (0.77–0.81), and sensitivity and specificity of 71.4% (68.1–74.4%) and 74.2% (73.1–75.2%) for a score ≥3.

Conclusion: The presence of anosmia/ageusia, fever with severe tiredness, or fever without sore throat should serve to suspect COVID-19 in areas with active viral circulation. The proportion of asymptomatics in children and adolescents challenges infection control.

Keywords: COVID-19; SARS-CoV-2 infection; Symptoms; Prediction model; General population; Nationwide serosurvey

Conflicts of Interest: None declared.

† Joint first authors.

‡ Joint senior authors.

§ Collaborators are listed in the Supplementary material.

* Corresponding author: Marina Pollán, National Centre for Epidemiology, Instituto de Salud Carlos III (ISCIII), Monforte de Lemos 5, 28029 Madrid, Spain. Tel.: +34 91 822 2635.

E-mail address: mpollan@isciii.es (M. Pollán).

https://doi.org/10.1016/j.jclinepi.2021.06.005

What is new?

- Little is known about the factors associated with asymptomatic SARS-CoV-2 infection, as most studies are based on clinical COVID-19 cases.
- Reviews highlight the need to explore the predictive ability of symptoms to identify individuals with COVID-19 in the general population, but available prediction models lack enough quality or are not suitable for this purpose.
- This nationwide seroepidemiological study found that nearly 30% of SARS-CoV-2 infections in Spain were asymptomatic, with higher prevalence of asymptomatic infections in regions with lower viral circulation and among men, children and adolescents, old people, and smokers.
- A symptomatic risk score was constructed and validated, providing an easy tool to predict COVID-19 based on three situations: a) presence of anosmia or ageusia, b) fever with severe tiredness, or c) fever without sore throat. It detects over 70% of cases in the general population, with a specificity greater than 70%.
- The proposed symptomatic risk score outperforms other combinations of symptoms frequently used to suspect COVID-19 and can be applied in primary care or community settings.

1. Introduction

Infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) results in a wide range of clinical manifestations [1,2]. While most infected individuals show symptoms of coronavirus disease 2019 (COVID-19), around one third remain asymptomatic [3,4] and, despite their lower secondary attack rate [5], constitute a serious challenge for the control of the pandemic. Little is known about the factors associated with asymptomatic SARS-CoV-2 infection, as most studies are based on clinical COVID-19 cases [1].

Another relevant issue in SARS-CoV-2 research is the evaluation of the predictive ability of symptoms to identify individuals with COVID-19 in primary care settings, to allow a prompt isolation and treatment of the patient, and to identify and quarantine his/her close contacts in order to prevent the spread of infection. In this regard, the information on COVID-19 symptoms relies largely on hospital data [2], not directly applicable to the milder cases attended in primary care [6], or in data from responders to syndromic-surveillance tools, who may not truly represent the general population. A recent review of COVID-19 diagnosis prediction models concludes that those proposed for general population have high risk of bias [7]. Ideally, the evaluation of symptoms of SARS-CoV-2 infection should be ascertained in a population-based study that identifies all individuals infected by the virus.

One such a study is the nationwide seroepidemiological survey in Spain [3]. With more than 61,000 randomly selected individuals, it provides an accurate picture of the complete first pandemic wave in Spain. The study gathered information on COVID-19 related symptoms and blood to measure antibodies against SARS-CoV-2, so it allows to characterize asymptomatic infections and COVID-19 cases in the general population, overcoming the above mentioned limitations. ENE-COVID started 1 month after the peak of the wave, reducing the probability of infra-detection of cases due to insufficient time to seroconvert or due to waning of antibodies in infections occurring several months before testing. Taking advantage of this large and representative study, we explore the characteristics of asymptomatic cases, describe symptoms' patterns, and propose and validate a symptom-based risk score to guide COVID-19 diagnosis in primary healthcare.

2. Methods

2.1. Study design and population

The Seroepidemiological Survey of SARS-CoV-2 Virus Infection in Spain (*Estudio Nacional de Sero-Epidemiología de la Infección por SARS-CoV-2 en España*, ENE-COVID) is a nationwide population-based cohort study to investigate seropositivity for SARS-CoV-2 in the community-dwelling population in Spain. The study design has been described elsewhere [3]. Briefly, 1,500 census tracts, and up to 24 households per tract, were randomly selected through a two-stage sampling stratified by province and municipality size. All residents in the 35,885 selected household were invited to participate in the study. Serial data from epidemiological questionnaires and serology tests were collected for study participants in three rounds between April 27 and June 22, 2020. Each round was completed in 2 weeks, with a 1-week break between rounds. In this report we used the questionnaire and serology obtained in each individual's first round of participation.

2.2. Data collection

Field work was carried out by trained staff from the Spanish regional health services under a common protocol developed and supervised by the Institute of Health Carlos III and the Ministry of Health. Residents in selected households were contacted by phone and invited to go to their primary healthcare centers or allow a home visit. Those who agreed to participate answered a questionnaire on sociodemographic characteristics, risk factors, chronic conditions, contact with suspected or confirmed COVID-19 cases, and presence and date of onset of any of the following nine symptoms compatible with COVID-19:

fever; chills; severe tiredness; sore throat; cough; shortness of breath; headache; nausea and/or vomiting and/or diarrhea; and anosmia or ageusia (anosmia/ageusia). Participants also received a rapid serology test and were asked to donate blood samples, which were centrifuged, refrigerated, and analyzed in one of 29 selected microbiology laboratories [3]. The study protocol and questionnaire are available in Spanish in ENE-COVID study website [8].

2.3. Detection of SARS-CoV-2 antibodies

We used two serology tests to detect IgG antibodies against SARS-CoV-2: a point-of-care test applied directly to fingerprick blood (Orient Gene Biotech COVID-19 IgG/IgM Rapid Test Cassette, reference GCCOV-402a) and a chemiluminescent microparticle immunoassay (CMIA) using serum samples (SARS-CoV-2 IgG for use with ARCHITECT, Abbott Laboratories, reference 06R8620). Due to its better performance characteristics, this report uses serological results from the CMIA test, which has shown a sensitivity of 90.6% and a specificity of 99.3% in a meta-analysis of 23 diagnostic accuracy studies [9].

2.4. Statistical analysis

The seroprevalence of SARS-CoV-2 was calculated as the proportion of participants with detectable IgG antibodies against SARS-CoV-2 by the CMIA test. The prevalence of asymptomatic SARS-CoV-2 infection was calculated as the proportion of seropositive participants who did not report any symptom compatible with COVID-19. To control for confounding, the prevalence of asymptomatic infection by individual characteristics was standardized to the overall distribution of all other characteristics in the entire seropositive population. To this end, we first fitted a design-based logistic regression model adjusted for the other characteristics, and then computed a weighted average of the predicted probabilities of being asymptomatic, assuming that every seropositive participant was in each category of the individual characteristic [10]. We estimated standardized prevalence ratios and differences for asymptomatic SARS-CoV-2 infection across categories of individual characteristics.

In analyses restricted to symptomatic participants with onset of symptoms at least 21 days before blood draw, we developed a classification tree to construct relevant combinations of symptoms based on their distinct SARS-CoV-2 seroprevalence. The classification tree used the chi-square automatic interaction detection algorithm to recursively split clusters of participants based on symptoms with the lowest Bonferroni-corrected P values obtained from design-based logistic models [11]. The minimum cluster size was set at 1% of symptomatic participants.

We evaluated the diagnostic performance of symptoms in predicting SARS-CoV-2 seropositivity among symptomatic participants. Using the same model-based standardization described above [10], we estimated standardized ratios and differences for SARS-CoV-2 seroprevalence across categories of individual symptoms, total number of symptoms, and a symptomatic risk score. We constructed the symptomatic risk score by assigning to each symptom a weight proportional to its log-transformed standardized seroprevalence ratio. The population discrimination index of the symptomatic risk score for predicting SARS-CoV-2 seropositivity was calculated as the weighted proportion of seropositive-negative pairs in which the seropositive case had a higher symptomatic risk score on 1,000 design-based bootstrap samples, obtaining an overfitting-corrected discrimination index and 95% confidence interval (CI) as the mean and the 2.5th to 97.5th percentiles of the bootstrap replications [12]. We estimated sensitivity, specificity, and predictive values of the symptomatic risk score for the optimal threshold that minimized the overall misclassification rate (sum of false positive and negative rates) [13]. The predictive ability of the symptomatic risk score was compared with that of the classification tree based on symptom interactions.

In all analyses, we assigned sampling weights to study participants to account for the different selection probabilities by province, and to adjust for the distinct response rates to provide blood for the CMIA test by sex, age, and census tract average income. We trimmed extreme weights (upper 0.5%) to prevent highly influential observations. All statistical analyses accounted for the stratification by province and municipality size and the clustering of seropositivity by household and census tract when computing standard errors and CIs [3]. Analyses were performed using survey commands and chaid package in Stata, v16 and survey package in R, v4.

3. Results

Of 88,653 contacted individuals, 61,092 participants (68.9%) provided blood for the CMIA test in any of the three rounds (Supplementary Fig. 1). The proportion of testing was lower in individuals younger than 20 years (45.3%) and older than 80 years (58.9%), and in men aged 20–59 years compared with women (71.9% vs. 78.1%).

The seroprevalence of SARS-CoV-2 (95% CI) was 2.0% (1.8–2.3%) in asymptomatic participants, 10.8% (10.0–11.7%) in symptomatic participants with onset of symptoms at least 21 days before blood draw, 60.0% (48.9–70.1%) in participants who reported past pneumonia, and 36.7% (31.8–41.9%) in those cohabitating with a confirmed COVID-19 case (Table 1). The seroprevalence also varied by province and municipality size and was higher in healthcare workers and nonsmokers, with no differences by body max index, or by any of the selected chronic conditions, frequently associated with higher risk of severe disease (Table 1).

Table 1. Seroprevalence of SARS-CoV-2 by participant characteristics, self-reported symptoms, case contact, and residential features, ENE-COVID study, April 27–June 22, 2020, Spain

Characteristic	No. of participants[I] (%)	No. of positive cases[II]	SARS-CoV-2 seroprevalence[III] (%; 95% CI)
Overall	61,092	2,669	4.6 (4.2–4.9)
Sex			
Men	29,122 (48.9)	1,246	4.4 (4.1–4.8)
Women	31,970 (51.1)	1,423	4.7 (4.3–5.1)
Age (years)			
0–19	7,682 (19.0)	280	3.6 (3.1–4.3)
20–39	13,427 (23.1)	585	5.0 (4.5–5.6)
40–59	22,561 (32.2)	1,071	4.8 (4.4–5.3)
60–79	14,375 (21.2)	621	4.5 (3.9–5.1)
≥80	3,047 (4.5)	112	4.2 (3.2–5.5)
Nationality			
Spain	58,441 (95.2)	2,555	4.5 (4.2–4.8)
Other	2,642 (4.8)	114	5.5 (4.2–7.1)
Occupation[IV]			
Online work	12,676 (21.2)	651	5.8 (5.2–6.4)
Non-healthcare on-site work	12,840 (18.3)	555	4.3 (3.8–4.9)
Healthcare	2,397 (3.4)	211	9.1 (7.7–10.7)
Unemployed	4,764 (7.3)	143	3.3 (2.7–4.2)
Not economically active	28,386 (49.7)	1,107	4.0 (3.6–4.4)
Smoking			
No	45,604 (76.1)	2,229	5.0 (4.7–5.4)
Yes	15,346 (23.9)	433	3.0 (2.6–3.4)
Body mass index[V] (kg/m^2)			
<25	22,064 (42.8)	1,019	5.0 (4.5–5.5)
25–30	20,673 (38.0)	912	4.6 (4.2–5.0)
≥30	10,667 (19.2)	458	4.5 (3.9–5.2)
No. of chronic conditions[VI]			
0	21,182 (52.9)	999	4.8 (4.3–5.2)
1	9,849 (24.6)	412	4.4 (3.9–5.1)
2	5,545 (13.7)	255	4.8 (4.1–5.7)
≥3	3,407 (8.8)	138	4.3 (3.5–5.4)
Chronic condition[VI]			
Diabetes	4,660 (11.5)	190	4.1 (3.3–5.1)
Hypertension	11,742 (29.5)	506	4.7 (4.1–5.3)
Cardiovascular disease	5,828 (14.4)	248	4.5 (3.8–5.4)
Cancer	1,720 (4.5)	81	5.0 (3.8–6.5)
Chronic pulmonary disease	3,264 (8.4)	134	4.3 (3.4–5.3)
Asthma	2,170 (5.5)	102	4.9 (3.9–6.2)
Sleep apnea	1,632 (4.3)	84	5.4 (4.1–7.1)
Chronic kidney disease	859 (2.3)	25	3.7 (2.2–6.0)
Immunosuppressive disease	772 (1.9)	35	5.3 (3.5–7.9)
Self-reported symptoms[VII]			
Asymptomatic	40,090 (64.8)	781	2.0 (1.8–2.3)
Symptomatic <21 days before blood draw	4,565 (7.5)	155	3.2 (2.6–4.0)
Symptomatic ≥21 days before blood draw	16,437 (27.7)	1,733	10.8 (10.0–11.7)

(*continued on next page*)

Table 1 (*continued*)

Pneumonia			
No	60,937 (99.7)	2,570	4.4 (4.1–4.7)
Yes	155 (0.3)	99	60.0 (48.9–70.1)
Contact with COVID-19 case[VIII]			
No known contact	48,882 (79.0)	1,189	2.4 (2.2–2.7)
Non-cohabitating suspected case	3,332 (5.9)	225	8.4 (7.0–10.0)
Non-cohabitating confirmed case	4,228 (6.8)	492	11.4 (10.0–12.9)
Cohabitating suspected case	3,504 (6.4)	397	10.4 (8.8–12.3)
Cohabitating confirmed case	1,011 (1.9)	360	36.7 (31.8–41.9)
Household size (residents)			
1	5,306 (7.9)	225	4.4 (3.7–5.3)
2	15,170 (23.5)	667	5.0 (4.4–5.7)
3	15,858 (25.5)	644	4.1 (3.6–4.7)
4	16,458 (29.0)	806	5.0 (4.4–5.7)
≥5	8,300 (14.1)	327	3.7 (3.0–4.5)
Census tract average income[IX]			
<25th percentile	16,143 (24.8)	737	4.5 (3.8–5.3)
25–50th percentile	15,462 (25.7)	561	4.6 (3.8–5.5)
50–75th percentile	13,920 (25.0)	559	4.2 (3.5–5.0)
≥75th percentile	15,567 (24.5)	812	5.0 (4.3–5.8)
Municipality size (inhabitants)			
<5,000	11,159 (11.4)	511	3.8 (3.0–4.7)
5,000–20,000	12,939 (18.2)	507	3.3 (2.8–3.9)
20,000–100,000	18,066 (29.5)	720	3.6 (3.2–4.1)
≥100,000	18,928 (40.9)	931	6.0 (5.4–6.6)
Province seroprevalence[X] (%)			
<3	29,991 (48.6)	505	1.6 (1.4–1.9)
3–5	15,426 (18.2)	586	3.6 (3.1–4.1)
5–10	6,111 (14.1)	432	6.7 (5.6–7.9)
≥10	9,564 (19.2)	1,146	11.3 (10.2–12.6)
Entry round			
First (April 27–May 11)	52,318 (84.2)	2,291	4.5 (4.2–4.9)
Second (May 18–June 1)	7,122 (12.8)	305	4.6 (3.9–5.5)
Third (June 8–June 22)	1,652 (3.0)	73	4.7 (3.3–6.6)

[I] Of the 61,092 participants, 9 (0.0%) had missing data for nationality, 29 (0.0%) for occupation, 142 (0.2%) for smoking, 6 (0.0%) for body mass index, and 135 (0.2%) for contact with COVID-19 case. Data are number of participants (weighted percentage).

[II] Number of seropositive participants with detectable IgG antibodies against SARS-CoV-2 by the chemiluminescent microparticle immunoassay.

[III] Population seroprevalence of SARS-CoV-2 and 95% confidence interval (CI) accounting for sampling weights, nonresponse rates by sex, age, and census tract average income, stratification by province and municipality size, and clustering by household and census tract.

[IV] Online work, non-healthcare on-site work (retail, transport, police/firefighter/public safety, cleaning, or other on-site work), healthcare (hospital, primary care, nursing home, or other health/social work), unemployed, or not economically active (student, retired, permanent/temporal disability, house person, or other unpaid work).

[V] Among participants aged 20 years or older.

[VI] Among participants aged 40 years or older. The number of chronic conditions was computed from those listed in the table.

[VII] Including fever, chills, severe tiredness, sore throat, cough, shortness of breath, headache, nausea/vomiting/diarrhea, and anosmia/ageusia.

[VIII] Contact with non-cohabitating (relative, friend, co-worker, housemaid, caregiver, or client/patient) or cohabitating (household member) suspected case (non-confirmed symptomatic person) or confirmed COVID-19 case. If multiple contacts were reported, we first considered cohabitating cases and then confirmed cases.

[IX] Quartiles from province-specific distributions of census tract average income in 2017.

[X] Provinces with population seroprevalence <3% (Alicante, Almería, Badajoz, Baleares, Cádiz, Castellón, Córdoba, A Coruña, Girona, Gipuzkoa, Huelva, Jaén, Lleida, Lugo, Murcia, Ourense, Asturias, Las Palmas, Pontevedra, Tenerife, Sevilla, Tarragona, Teruel, Valencia, Ceuta, and Melilla), 3–5% (Álava, Burgos, Cáceres, Granada, Huesca, La Rioja, Málaga, Navarra, Palencia, Cantabria, Valladolid, Bizkaia, Zamora, and Zaragoza), 5–10% (Ávila, Barcelona, León, and Salamanca), or ≥10% (Albacete, Ciudad Real, Cuenca, Guadalajara, Madrid, Segovia, Soria, and Toledo) with IgG antibody chemiluminescent microparticle immunoassay.

3.1. Asymptomatic SARS-CoV-2 infection

The prevalence of asymptomatic SARS-CoV-2 infection among 2,669 seropositive participants was 28.7% (95% CI 26.1–31.4%). Asymptomatic infections were more prevalent in men (31.8%), in individuals younger than 20 years (44.9%) and older than 80 years (36.1%), in smokers (33.0%), in those unaware of having had contact with a COVID-19 case (41.4%), in municipalities with less than 20,000 inhabitants (36.2–36.6%), and in provinces with seroprevalence lower than 3% (40.3%; Table 2 and Supplementary Table 1). The standardized prevalence ratio (95% CI) of asymptomatic SARS-CoV-2 infection was 1.19 (1.02–1.40) for men vs. women, 1.82 (1.33–2.50) for individuals younger than 20 years and 1.45 (0.96–2.18) for individuals older than 80 years vs. those aged 40–59, 1.27 (1.03–1.55) for smokers vs. nonsmokers, and 1.91 (1.59–2.29) for individuals without vs. those with known contact with a COVID-19 case (Table 2).

3.2. Seroprevalence-based combination of symptoms

The classification tree (Supplementary Fig. 2) defined 15 combinations of symptoms with marked differences in seroprevalence, which ranged from 63.7% (95% CI 57.8–69.5%) for participants with anosmia/ageusia and fever, but not sore throat, to 3.3% (1.4–5.2%) for those with fever and sore throat, but not anosmia/ageusia, chills, or tiredness. Figure 1 displays the symptoms (presence, absence, or indifference) defining each cluster, together with the cluster distribution among all symptomatic participants and seropositive cases, as well as the SARS-CoV-2 seroprevalence in participants with these clinical presentations. Clusters were mostly defined by four of the nine symptoms considered: anosmia/ageusia, which was the less common but most influential symptom, fever, severe tiredness, and absence of sore throat. None of the clusters included headache as a relevant symptom and cough only appeared in two of them. These two symptoms, in spite of being the most reported complains, did not seem to help differentiate between seropositive and seronegative individuals.

In addition to these data-driven clusters, there are two predefined combinations of symptoms widely used to suspect COVID-19: fever, cough, and shortness of breath, or fever and cough. These combinations were present in 6.5% (95% CI 5.9–7.0%) and 17.7% (16.9–18.7%) of seropositive cases, with a SARS-CoV-2 seroprevalence of 25.4% (21.8–29.4%) and 21.8% (19.3–24.6%), respectively.

3.3. Predictive accuracy of symptoms for SARS-CoV-2 infection

Considering each symptom separately, severe tiredness (32.2% of symptomatic participants with symptom onset $\geq$21 days), fever (30.4%), and anosmia/ageusia (11.3%) were positively associated with SARS-CoV-2 infection, whereas sore throat (42.1%) was negatively related (Table 3 and Supplementary Table 2). Compared with symptom absence, the standardized ratio (95% CI) of SARS-CoV-2 seroprevalence was 1.34 (1.19–1.52) for severe tiredness, 1.66 (1.46–1.89) for fever, 4.10 (3.57–4.72) for anosmia/ageusia, and 0.75 (0.67–0.84) for sore throat.

A symptomatic risk score assigning a weight of 1 to severe tiredness and absence of sore throat, 2 to fever, and 5 to anosmia/ageusia showed standardized seroprevalence ratios (95% CIs) ranging from 2.04 (1.69–2.45) to 8.71 (7.37–10.3), substantially greater than those associated with the number of symptoms (Table 3). The discrimination index (95% CI) of the number of symptoms and the symptomatic risk score for predicting SARS-CoV-2 seropositivity in the symptomatic population were 0.69 (0.67–0.71) and 0.79 (0.77–0.81), respectively (Fig. 2). The discrimination of the symptomatic risk score was higher than 0.75 in all subgroups except individuals younger than 20 years (0.68) and older than 80 years (0.73), those living with a confirmed COVID-19 case (0.74), and those from provinces with seroprevalence below 3% (0.74; Supplementary Table 3).

The optimal diagnostic thresholds that minimized the overall misclassification rate were 4 or more symptoms and 3 or greater symptomatic risk score (Fig. 2), the latter corresponding to the presence of anosmia/ageusia, fever with severe tiredness, or fever without sore throat. The sensitivity and specificity (95% CIs) for the presence of 4 or more symptoms were 57.4% (54.0–60.7%) and 72.1% (71.0–73.1%), respectively, and increased to 71.4% (68.1–74.4%) and 74.2% (73.1–75.2%) for a symptomatic risk score equal to or greater than 3. The sensitivity of the symptomatic risk score remained higher than 65% in all subgroups except individuals younger than 20 years and older than 80 years and those from low-prevalence provinces, whereas the specificity was higher than 65% in all subgroups except those living with a confirmed case (Supplementary Table 4). For an overall SARS-CoV-2 seroprevalence of 10.8% among symptomatic individuals, the positive and negative predictive values (95% CIs) were 20.0% (18.2–22.0%) and 93.3% (92.6–94.0%) for the number of symptoms, respectively, and reached 25.1% (23.1–27.3%) and 95.5% (94.9–96.1%) for the symptomatic risk score (Supplementary Table 4).

The classification tree did not improve the predictive ability of the symptomatic risk score, despite its complex symptom interactions, with an overall discrimination index of 0.80 (95% CI 0.78–0.81) (Supplementary Fig. 3).

The date of symptom onset among seropositive cases for SARS-CoV-2 showed a narrow distribution around the peak of the first pandemic wave in Spain, whereas the distribution of symptom onset among seronegative individuals was substantially wider around the same pandemic pick (Supplementary Fig. 4).

Table 2. Prevalence of asymptomatic SARS-CoV-2 infection by participant characteristics, case contact, and residential features, ENE-COVID study, April 27–June 22, 2020, Spain

Characteristic	No. of positive cases[I] (%)	No. of asymptomatic cases[II]	Asymptomatic prevalence[III] (%; 95% CI)	Crude prevalence ratio (95% CI)	Standardized prevalence ratio[IV] (95% CI)
Overall	2,669	781	28.7 (26.1–31.4)		
Sex					
Men	1,246 (47.5)	406	31.8 (28.4–35.5)	1.00 (reference)	1.00 (reference)
Women	1,423 (52.5)	375	25.8 (22.5–29.5)	0.81 (0.69–0.96)	0.84 (0.71–0.98)
Age (years)					
0–19	280 (15.2)	118	44.9 (36.4–53.7)	2.06 (1.61–2.64)	1.82 (1.33–2.50)
20–39	585 (25.6)	143	25.0 (20.6–30.0)	1.15 (0.91–1.44)	1.18 (0.94–1.47)
40–59	1,071 (34.2)	259	21.8 (18.8–25.1)	1.00 (reference)	1.00 (reference)
60–79	621 (20.9)	215	31.1 (26.3–36.4)	1.43 (1.15–1.77)	1.24 (0.94–1.64)
≥80	112 (4.2)	46	36.1 (25.4–48.5)	1.66 (1.16–2.37)	1.45 (0.96–2.18)
Nationality					
Spain	2,555 (94.2)	754	29.3 (26.7–32.2)	1.00 (reference)	1.00 (reference)
Other	114 (5.8)	27	17.7 (10.6–28.2)	0.60 (0.37–0.99)	0.69 (0.44–1.10)
Occupation					
Online work	651 (26.9)	126	19.6 (15.9–23.9)	1.00 (reference)	1.00 (reference)
Non-healthcare on-site work	555 (17.2)	163	27.7 (22.9–33.0)	1.41 (1.09–1.84)	1.22 (0.96–1.55)
Healthcare	211 (6.8)	43	16.9 (11.9–23.6)	0.87 (0.59–1.27)	0.98 (0.71–1.36)
Unemployed	143 (5.3)	44	31.8 (22.3–43.1)	1.62 (1.09–2.42)	1.31 (0.92–1.87)
Not economically active	1,107 (43.8)	405	36.1 (31.6–40.9)	1.84 (1.46–2.33)	1.25 (0.95–1.66)
Smoking					
No	2,229 (84.2)	634	27.8 (25.0–30.8)	1.00 (reference)	1.00 (reference)
Yes	433 (15.8)	143	33.0 (27.2–39.4)	1.19 (0.96–1.46)	1.27 (1.03–1.55)
Body mass index[V] (kg/m^2)					
<25	1,019 (45.0)	260	25.6 (22.0–29.6)	1.00 (reference)	1.00 (reference)
25–30	912 (36.9)	275	26.9 (23.0–31.2)	1.05 (0.85–1.30)	0.93 (0.76–1.14)
≥30	458 (18.1)	128	23.9 (18.9–29.8)	0.94 (0.72–1.21)	0.86 (0.67–1.10)
No. of chronic conditions[VI]					
0	999 (54.1)	269	23.5 (20.3–27.1)	1.00 (reference)	1.00 (reference)
1	412 (23.4)	134	27.8 (22.6–33.6)	1.18 (0.94–1.49)	1.01 (0.80–1.28)
2	255 (14.2)	79	32.3 (25.8–39.7)	1.37 (1.06–1.78)	1.06 (0.81–1.37)
≥3	138 (8.2)	38	27.4 (18.5–38.6)	1.16 (0.78–1.74)	0.93 (0.62–1.41)
Contact with COVID-19 case					
No known contact	1,189 (42.4)	497	41.4 (37.3–45.8)	1.00 (reference)	1.00 (reference)
Non-cohab. suspected case	225 (10.9)	35	15.5 (10.0–23.4)	0.37 (0.24–0.58)	0.43 (0.28–0.66)
Non-cohab. confirmed case	492 (17.1)	93	18.8 (14.3–24.3)	0.45 (0.34–0.59)	0.58 (0.43–0.76)
Cohab. suspected case	397 (14.5)	72	18.8 (13.8–25.3)	0.45 (0.33–0.63)	0.48 (0.36–0.65)
Cohab. confirmed case	360 (15.0)	81	22.9 (17.1–30.0)	0.55 (0.41–0.74)	0.58 (0.43–0.77)
Household size (residents)					
1	225 (7.7)	65	27.0 (20.7–34.3)	0.87 (0.66–1.17)	0.94 (0.71–1.26)
2	667 (25.9)	204	28.7 (24.2–33.7)	0.93 (0.75–1.16)	0.93 (0.75–1.16)
3	644 (23.1)	191	28.4 (23.7–33.7)	0.92 (0.75–1.13)	0.95 (0.76–1.18)
4	806 (31.9)	231	30.8 (26.4–35.6)	1.00 (reference)	1.00 (reference)
≥5	327 (11.3)	90	24.3 (17.6–32.5)	0.79 (0.56–1.12)	0.85 (0.63–1.14)

(*continued on next page*)

Table 2 (*continued*)

Census tract average income					
<25th percentile	737 (24.3)	232	29.0 (24.2–34.4)	1.00 (reference)	1.00 (reference)
25–50th percentile	561 (25.9)	183	30.7 (25.1–36.9)	1.06 (0.82–1.37)	1.11 (0.88–1.41)
50–75th percentile	559 (23.0)	168	27.4 (22.4–33.0)	0.94 (0.73–1.23)	1.02 (0.82–1.29)
≥75th percentile	812 (26.8)	198	27.5 (22.6–33.0)	0.95 (0.73–1.23)	1.02 (0.79–1.31)
Municipality size (inhabitants)					
<5,000	511 (9.4)	203	36.6 (30.8–42.9)	1.00 (reference)	1.00 (reference)
5,000–20,000	507 (13.3)	170	36.2 (31.2–41.5)	0.99 (0.79–1.23)	0.94 (0.75–1.17)
20,000–100,000	720 (23.7)	183	27.5 (22.3–33.3)	0.75 (0.58–0.97)	0.77 (0.61–0.98)
≥100,000	931 (53.7)	225	26.0 (22.1–30.2)	0.71 (0.56–0.89)	0.80 (0.64–1.00)
Province seroprevalence (%)					
<3	505 (17.5)	205	40.3 (35.0–45.9)	1.00 (reference)	1.00 (reference)
3–5	586 (14.2)	177	30.0 (25.2–35.4)	0.74 (0.60–0.93)	0.76 (0.61–0.94)
5–10	432 (20.6)	126	26.9 (21.9–32.7)	0.67 (0.52–0.85)	0.77 (0.61–0.97)
≥10	1,146 (47.7)	273	24.7 (20.6–29.4)	0.61 (0.49–0.77)	0.73 (0.59–0.90)
Entry round					
First (April 27–May 11)	2,291 (83.8)	674	28.5 (25.7–31.4)	1.00 (reference)	1.00 (reference)
Second (May 18–June 1)	305 (13.0)	83	28.0 (21.5–35.7)	0.98 (0.75–1.30)	1.02 (0.79–1.33)
Third (June 8–June 22)	73 (3.2)	24	37.3 (22.1–55.5)	1.31 (0.82–2.09)	1.05 (0.66–1.67)

[I] Analyses restricted to 2,669 seropositive participants with detectable IgG antibodies against SARS-CoV-2 by the chemiluminescent microparticle immunoassay. Data are number of seropositive participants (weighted percentage).

[II] Number of seropositive participants without previous self-reported symptoms, including fever, chills, severe tiredness, sore throat, cough, shortness of breath, headache, nausea/vomiting/diarrhea, and anosmia/ageusia.

[III] Population prevalence of asymptomatic SARS-CoV-2 infection and 95% confidence interval (CI) accounting for sampling weights, nonresponse rates by sex, age, and census tract average income, stratification by province and municipality size, and clustering by household and census tract.

[IV] Prevalence ratio of asymptomatic SARS-CoV-2 infection and 95% confidence interval (CI) standardized to the overall distribution of all other characteristics presented in the table in the entire seropositive population in Spain.

[V] Among seropositive participants aged 20 years or older.

[VI] Among seropositive participants aged 40 years or older.

4. Discussion

We estimated the proportion of SARS-CoV-2 infections and of asymptomatic cases in the general population according to age, sex, presence of chronic conditions and COVID-19 risk factors in a large nationwide representative population-based seroprevalence survey. Additionally, symptomatic participants served to characterize the usual clinical presentation of COVID-19 in the population and to propose an easy-to-use symptomatic risk score, based on anosmia/ageusia, fever, severe tiredness and absence of sore throat, with acceptable sensitivity and specificity, intended to be useful as a first tool to suspect COVID-19 in primary health-care centers.

To our knowledge, this is the first attempt in a large population-based study to characterize asymptomatic cases, which, according to our results, represent nearly 30% of SARS-CoV-2 infections. This estimation, based on people with IgG antibodies, minimizes the possibility of including pre-symptomatic cases [14]. Identification of asymptomatic infections through serology has also its drawbacks. Apart from the exclusion of PCR+ cases who do not seroconvert, very recent infections and those occurring many months ago may not be detected due to insufficient time to develop a humoral response or to the waning of antibodies. While the exclusion of the low proportion of never seroconverters is unavoidable, the timing of our study allows minimizing the other two situations. As shown by the epidemiological information available, most ENE-COVID participants would have been infected one month before their first participation, and it is known that IgG antibodies are detected 2–3 weeks after symptom onset in more than 90% of COVID-19 cases [15] and decrease 2–3 months after infection [16].

The relevance of asymptomatic infections is undeniable. Available data suggest that they have a similar viral load [17,18], with conflicting reports about comparative duration of viral shedding [16,18,19]. However, even though their infectivity might be lower [5], asymptomatics probably account for a substantial fraction of new infections, still not well quantified [20], constituting a challenge in the prevention of transmission. In addition, the effect of SARS-CoV-2 vaccines on asymptomatic infections is still under study [21].

Table 3. Seroprevalence of SARS-CoV-2 among participants with self-reported symptoms compatible with COVID-19, ENE-COVID study, April 27–June 22, 2020, Spain

Symptom	No. of symptomatic participants[I] (%)	No. of positive cases[II]	SARS-CoV-2 seroprevalence[III] (%; 95% CI)	Crude seroprevalence ratio (95% CI)	Standardized seroprevalence ratio[IV] (95% CI)
Overall	16,437	1,733	10.8 (10.0–11.7)		
Fever					
No	11,890 (69.6)	797	6.8 (6.1–7.5)	1.00 (reference)	1.00 (reference)
Yes	4,547 (30.4)	936	20.1 (18.2–22.2)	2.97 (2.61–3.37)	1.66 (1.46–1.89)
Chills					
No	12,232 (74.2)	1,006	8.5 (7.8–9.4)	1.00 (reference)	1.00 (reference)
Yes	4,205 (25.8)	727	17.5 (15.6–19.5)	2.04 (1.81–2.31)	1.02 (0.89–1.16)
Severe tiredness					
No	11,127 (67.8)	774	6.9 (6.2–7.7)	1.00 (reference)	1.00 (reference)
Yes	5,310 (32.2)	959	19.1 (17.3–20.9)	2.75 (2.43–3.12)	1.34 (1.19–1.52)
Sore throat					
No	9,726 (57.9)	1,079	11.2 (10.2–12.3)	1.00 (reference)	1.00 (reference)
Yes	6,711 (42.1)	654	10.3 (9.2–11.6)	0.92 (0.81–1.05)	0.75 (0.67–0.84)
Cough					
No	8,320 (49.8)	799	9.6 (8.6–10.6)	1.00 (reference)	1.00 (reference)
Yes	8,117 (50.2)	934	12.1 (11.0–13.4)	1.27 (1.12–1.44)	1.01 (0.91–1.13)
Shortness of breath					
No	13,213 (80.3)	1,266	9.7 (8.9–10.6)	1.00 (reference)	1.00 (reference)
Yes	3,224 (19.7)	467	15.5 (13.7–17.6)	1.60 (1.40–1.83)	0.99 (0.87–1.12)
Headache					
No	8,008 (48.3)	768	9.6 (8.7–10.6)	1.00 (reference)	1.00 (reference)
Yes	8,429 (51.7)	965	12.0 (10.8–13.2)	1.25 (1.10–1.41)	0.92 (0.83–1.02)
Nausea/vomiting/diarrhea					
No	12,533 (75.6)	1,136	9.4 (8.6–10.3)	1.00 (reference)	1.00 (reference)
Yes	3,904 (24.4)	597	15.3 (13.6–17.1)	1.63 (1.44–1.84)	1.06 (0.95–1.17)
Anosmia/ageusia					
No	14,589 (88.7)	884	6.3 (5.7–7.0)	1.00 (reference)	1.00 (reference)
Yes	1,848 (11.3)	849	46.2 (42.8–49.7)	7.29 (6.49–8.19)	4.10 (3.57–4.72)
No. of symptoms					
1–2	9,157 (54.2)	547	5.8 (5.2–6.5)	1.00 (reference)	1.00 (reference)
3–4	4,115 (25.5)	446	11.7 (10.2–13.3)	2.00 (1.71–2.35)	1.72 (1.48–1.99)
5–6	2,124 (13.5)	432	19.4 (17.0–21.9)	3.32 (2.83–3.89)	2.53 (2.18–2.95)
7–9	1,041 (6.7)	308	31.0 (26.9–35.4)	5.30 (4.49–6.27)	3.58 (3.05–4.21)
Symptomatic risk score[V]					
0–1	9,574 (56.7)	352	3.9 (3.4–4.5)	1.00 (reference)	1.00 (reference)
2–3	4,211 (26.7)	380	9.2 (8.0–10.7)	2.38 (1.95–2.90)	2.04 (1.69–2.45)
4–5	964 (6.2)	195	19.4 (16.2–23.2)	5.00 (3.95–6.33)	3.76 (3.02–4.68)
6–7	845 (4.9)	344	40.8 (36.2–45.5)	10.5 (8.75–12.6)	6.61 (5.46–8.00)
8–9	843 (5.5)	462	54.4 (49.5–59.2)	14.0 (11.9–16.5)	8.71 (7.37–10.3)

[I] Analyses restricted to 16,437 symptomatic participants with onset of any of the nine symptoms at least 21 days before blood draw. Data are number of symptomatic participants (weighted percentage).

[II] Number of symptomatic participants with detectable IgG antibodies against SARS-CoV-2 by the immunoassay.

[III] Population seroprevalence of SARS-CoV-2 and 95% confidence interval (CI) accounting for sampling weights, nonresponse rates, stratification by province and municipality size, and clustering by household and census tract.

[IV] Ratio of SARS-CoV-2 seroprevalence and 95% confidence interval (CI) standardized to the overall distribution of sex, age group, nationality, occupation, smoking, body mass index, number of chronic conditions, contact with COVID-19 case, household size, census tract average income, municipality size, province seroprevalence, and entry round in the entire symptomatic population in Spain. Seroprevalence ratios for individual symptoms were further standardized to the overall distribution of all other symptoms.

[V] Symptomatic risk score assigning a weight of 1 to severe tiredness, 2 to fever, and 5 to anosmia/ageusia, together with a weight of 1 to absence of sore throat, which were proportional to their individual log-transformed standardized ratios and rounded to the nearest integer.

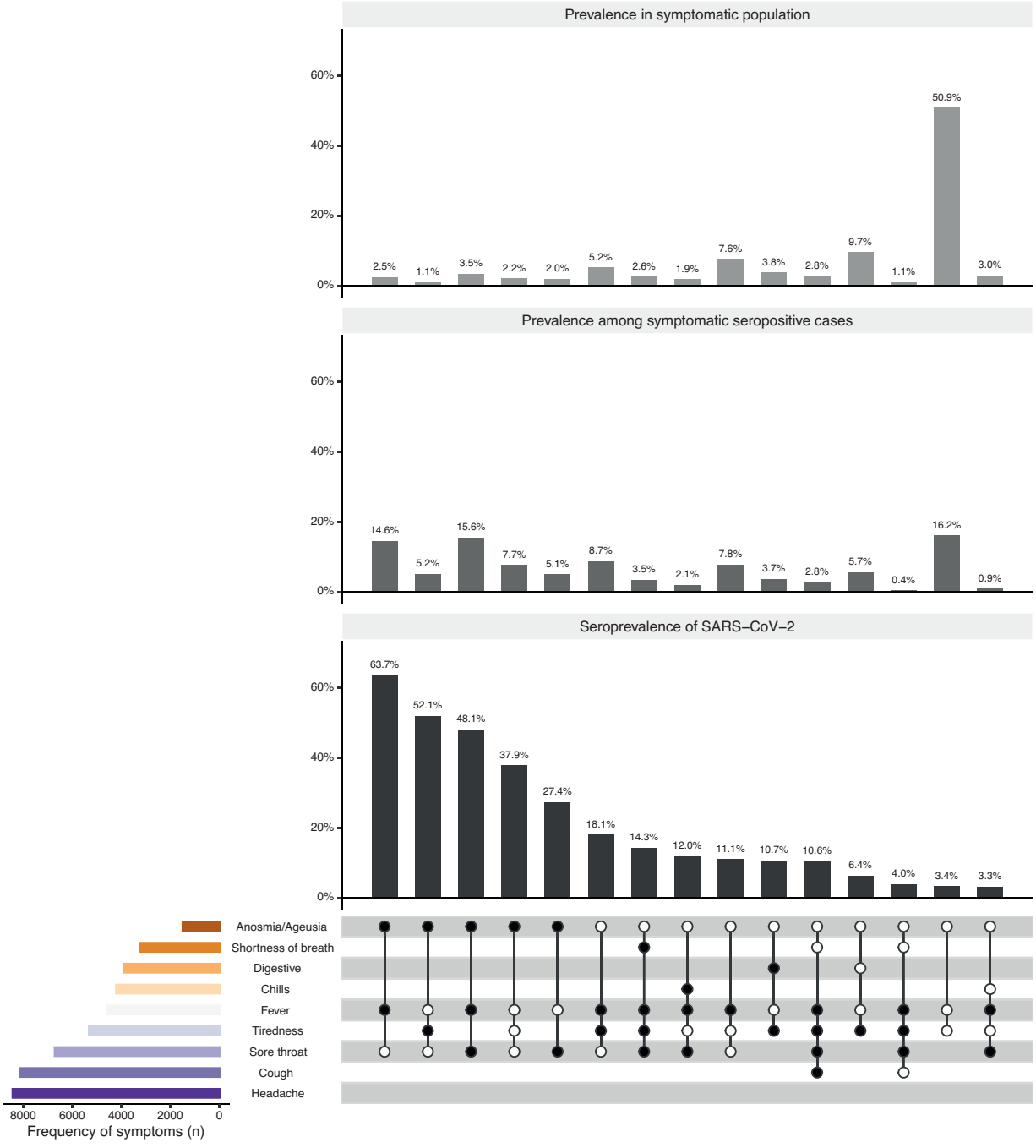

Fig. 1. Seroprevalence of SARS-CoV-2 by clusters of symptoms among participants with self-reported symptoms compatible with COVID-19, ENE-COVID study, April 27–June 22, 2020, Spain. Analyses restricted to 16,437 symptomatic participants with onset of symptoms at least 21 days before blood draw. Clusters were obtained from a classification tree algorithm and are defined by the presence (black circle) or absence (white circle) of specific symptoms, irrespective of the other symptoms without a circle.

The prevalence of asymptomatic cases was higher in men, in people younger than 20 years and in the elders. These differences, also observed in large hospital-based series [22] and in other studies [23], may reflect either a distinct clinical presentation or different sensitivity to notice and report common unspecific symptoms. Interestingly, in general population studies, the number of symptoms reported is higher in women [24,25] and tends to increase with age [25], although with a slight decline in elder groups [24]. The less severe infection in children and teenagers [26] and their high proportion of asymptomatic infections (44.9%) pose a problem to control strategies, given their higher mobility, as they are also involved in SARS-CoV-2 transmission [27,28], with similar viral loads

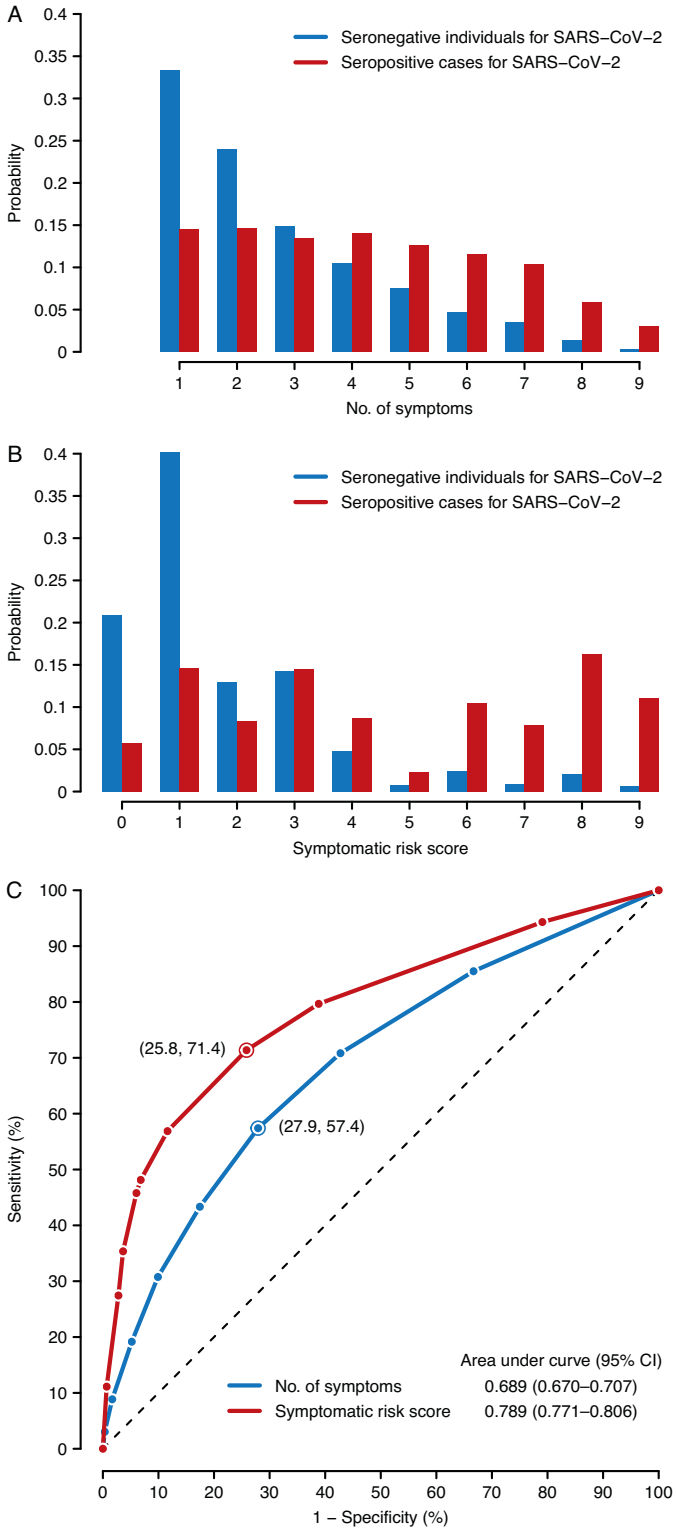

Fig. 2. Distribution of number of symptoms (A), distribution of symptomatic risk score (B), and ROC curves for predicting SARS-CoV-2 seropositivity (C) among participants with self-reported symptoms compatible with COVID-19, ENE-COVID study, April 27–June 22, 2020, Spain. Analyses restricted to 16,437 symptomatic participants with onset of any of the nine symptoms at least 21 days before blood draw. The symptomatic risk score assigned a weight of 1 to severe tiredness and absence of sore throat, 2 to fever, and 5 to anosmia/ageusia. Probability mass functions and sensitivities and specificities were estimated accounting for sampling weights and nonresponse rates. The outlined point on the ROC curves corresponded to the optimal threshold that minimized the overall misclassification rate (number of symptoms $\geq$4 and symptomatic risk score $\geq$3). The area under the ROC curves and its 95% confidence interval (CI) were calculated as the mean and the 2.5th to 97.5th percentiles of 1,000 design-based bootstrap replications.

than adults [29]. Finally, the higher proportion of asymptomatics in areas with lower seroprevalence, particularly among people without close contact with a COVID-19 case, may be explained by the reported association between viral load and symptomatic COVID-19 [30], since lower viral circulation at population level implies lower probability of contact with a heavily loaded viral carrier.

Smokers had also higher presence of asymptomatic infections and lower seroprevalence. A certain underdetection of cases among smokers could not be ruled out, since they may develop lower antibody levels [31]. However, a recent review concluded that smokers have lower risk of infection [32], although the mechanisms are not well understood [33]. A lower probability of participation of infected smokers in our study due to their higher risk of hospitalization is unlikely, as we contacted each person up to 3 times –one per round-, so they had several chances to participate. Furthermore, in a later contact, we asked about previous hospitalization and the seroprevalence among hospitalized participants was again lower among smokers.

COVID-19 is a multifaceted illness. Clinical presentation is highly variable, with most cases reporting common unspecific symptoms. A recent review highlights the urgent need to explore the syndromic presentation at population level [6], and a review of predictive models to suspect COVID-19 in the population concludes that those available do not have enough quality [7]. Our study contributes to fill this gap [6,7,34]. Since the beginning of the pandemic, the paradigmatic triad of symptoms observed in hospitalized patients [22], fever, cough, and shortness of breath, was used to suspect COVID-19 [1,2,34]. However, in our study only 6.5% of symptomatic cases experienced this triad, while 18% reported fever and cough. The updated WHO definition for suspected COVID-19 [35] includes fever and cough, or three or more symptoms from a list similar to ours, while the presence of anosmia/ageusia classifies the patient as possible COVID-19.

From a population-based study, we propose an easy-to-use symptoms score that might help to suspect COVID-19 and guide subsequent epidemiological and clinical decisions. Three situations serve to suspect COVID-19 among symptomatic patients: sudden loss of smell/taste, the combination of fever and severe tiredness and the presence of fever without sore throat. In our study, 71% of symptomatic infected cases had scores equal to or greater than 3, compared to 24% of non-infected symptomatic people, which supports the utility of this score. We corroborate that loss of smell and/or taste, present in 49% of symptomatic infections, is the best predictor of COVID-19 [36–38], while fever and severe tiredness were present in more than half of them. However, we do not have information on the day of onset of each symptom or the order in which they appear, so the validity of our tool for early detection is unknown. Our score has poorer performance in the oldest age groups. COVID-19 might be harder to suspect in the elders, whose symptoms can be masked, so the threshold for testing should be lowered in their case, given their greater risk of serious complications and higher lethality [1,9].

A recent study in the UK general population identified four symptoms from a list of 26 –new persistent cough, anosmia, ageusia and fever- as the most discriminant between PCR positive and negative participants [39]. Using the information provided by UK and US individuals through an smartphone app, Menni et al built up a regression model to predict COVID-19, adjusting for age, sex and body mass index [36]. Their model included "persistent cough," "loss of appetite" and two of the symptoms considered in ours: anosmia/ageusia and tiredness. This model had lower sensitivity (65% vs. 71%) and slightly higher specificity (78% vs. 74%) than ours [36]. We collected digestive symptoms other than loss of appetite and cough was not specifically required to be persistent, which may explain in part why they did not add any discrimination. However, cough did not help to predict COVID-19 in a smaller study from Singapore [40]. Interestingly, in our data sore throat, a symptom that usually accompanies many respiratory infections, was less prevalent among the infected symptomatic participants, compared to their uninfected counterparts (37% vs 59%), and the score included the absence of this symptom. Most studies define infections based on PCR tests, while ours is based on IgG antibodies, so results are not fully comparable. PCR is the gold-standard to detect early infections, but it has a narrow window of positivity and may misclassify COVID-19 patients tested outside it. In contrast, antibodies serve to explore infections in a retrospective way.

Among the strengths of this study is the use of a CMIA test with good characteristics of sensitivity and specificity [9]. Specificity is a crucial requirement to avoid false-positives in a context of low prevalence, such as ours. In fact, in a recent review by the UK National SARS-CoV-2 Serology Assay Evaluation Group, our test was the one with the highest specificity (99.9%) [41]. IgG antibodies are detected in more than 90% of COVID-19 cases 2 weeks after symptom onset [15], something we confirmed in participants reporting a positive PCR [3]. Antibodies seem to decrease after 3 months from infection, particularly those against the N protein, that often become undetectable by 5–7 months [42], more notably among mild and asymptomatic cases [16,42]. However, according to the first epidemic wave in Spain [43], most of our seropositive participants would have been infected around 1 or 2 months before being tested, being unlikely that antibodies had already waned to undetectable levels. One unavoidable limitation is the reliance on self-reported information -or in that provided by proxies in children and mentally disabled people-, heavily depending on personal and contextual factors [44]. Symptoms are, by definition, subjective and prone to recall bias, and symptom awareness may be different in areas of high and low viral circulation. How-

ever, in our case the date of interview was close (weeks) to symptoms onset, and probably memory was good, while most participants were interviewed during a very strict lockdown, with the pandemic being the first concern of a society that was constantly informed and reminded by the media. Finally, even though most frequent symptoms were included in the questionnaire, it did not cover the whole range communicated by COVID-19 patients [1,34].

5. Conclusion

The presence of sudden anosmia/ageusia, or a combination of fever with severe tiredness or fever without sore throat can be useful markers of SARS-CoV-2 infection in areas with active viral circulation that may help guide epidemiological and clinical actions. However, any symptom associated with COVID-19 should be an indication for testing in the elderly, given the high lethality of SARS-CoV-2 infection among them [9]. The high prevalence of asymptomatic infections in children and teenagers poses a challenge to stop SARS-CoV-2 dissemination.

Author contributions

BP-G and RP-B are joint first authors, and RY and MP are joint senior authors. BP-G, RP-B, and MP are responsible for the conception and design of the study; FB and RY are the executive coordinators of the project and led the relationship with regional health services; MP-O, JO-I, and AF-G are responsible for validation studies to select the serological tests, the coordination of participant microbiological labs, and the acquisition of laboratory data. MM, JLS, JLP, and JFM-M are responsible for the study operative, including the coordination of data acquisition and logistics; NFL and IC developed the operational protocols for field work and conducted the training of the involved administrative and health personnel; BP-G, RP-B, MAH, NFL, PF-N, and MP were in charge of statistical analyses and tables and figures design; other authors included in the ENE-COVID group contributed to data acquisition, laboratory analyses, and quality control at their respective regions or at national level. The first draft was written by BP-G, RP-B, MAH, RY, and MP. All authors contributed to data interpretation, critically reviewed the first draft, approved the final version, and agreed to be accountable for the work. BP-G, RP-B, and MP act as guarantors, accept full responsibility for the work, had access to the data, and controlled the decision to publish. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Ethical approval

The ENE-COVID study was approved by the Institutional Review Board of the Institute of Health Carlos III (register number PI 39_2020). Written informed consent was obtained from all participants, with specific forms for adolescents, parents of participant children, and guardians of mentally disabled participants, and assistance of witnesses for those not able to read.

Funding

The ENE-COVID study was supported by the Spanish Ministry of Health, the Institute of Health Carlos III, and the Spanish National Health System. The funders were involved in the study logistics, but they had no role in study design or in the collection, analysis, interpretation of data, or the decision to submit the article for publication.

Data availability

ENE-COVID has stablished a procedure for data request, with a Scientific Board that evaluates these petitions and guarantees the safeguard of participants' rights, under the limits imposed by the Ethical Committee. Requests should be addressed to direccion-ENECOVID@isciii.es.

Acknowledgements

This work was supported by the Spanish Ministry of Health, the Institute of Health Carlos III (Ministry of Science and Innovation), and the National Health System, including the Health Services of all autonomous communities and autonomous cities: Servicio Andaluz de Salud, Servicio Aragonés de Salud, Servicio de Salud Principado de Asturias, Servei de Salut Illes Balears, Servicio Canario de la Salud, Servicio Cántabro de Salud, Servicio de Salud Castilla-La Mancha, Servicio de Salud de Castilla y León, Servei Català de Salut, Conselleria de Sanitat Universal i Salut Pública Generalitat Valenciana, Servicio Extremeño de Salud, Servizo Galego de Saúde, Servicio Riojano de Salud, Servicio Madrileño de Salud, Servicio Murciano de Salud, Servicio Navarro de Salud-Osasunbidea and Instituto de Salud Pública y Laboral de Navarra, Servicio Vasco de Salud-Osakidetza, and Instituto Gestión Sanitaria. The Spanish Institute of Statistics provided the random selection of households and the information required for participants' contact. We thank the nurses, general practitioners, administrative staff, and other healthcare workers who collaborated in this study and all participants. This study is the result of the efforts of many professionals, and the trust and generosity of more than 61,000 participants who have understood the interest of providing time, information, and samples to learn about the COVID-19 pandemic in Spain.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2021.06.005.

References

[1] Vetter P, Vu DL, L'Huillier AG, et al. Clinical features of covid-19. BMJ 2020;369:m1470. doi:10.1136/bmj.m1470.

[2] Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of Coronavirus Disease 2019 (COVID-19): a review. JAMA 2020;324:782–93. [Internet][cited 2020 Jul 15]Available from https://jamanetwork.com/journals/jama/fullarticle/2768391 . doi:10.1001/jama.2020.12839.

[3] Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. Lancet 2020;396:535–44. [Internet][cited 2020 Jul 15];0Available from https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31483-5/abstract . doi:10.1016/S0140-6736(20)31483-5.

[4] Oran DP, Topol EJ. The proportion of SARS-CoV-2 infections that are asymptomatic : a systematic review. Ann Intern Med 2021. *[Internet]*[cited 2021 Mar 26]Available from: https://www.acpjournals.org/doi/10.7326/M20-6976 . doi:10.7326/M20-6976.

[5] Qiu X, Nergiz AI, Maraolo AE, et al. Defining the role of asymptomatic and pre-symptomatic SARS-CoV-2 transmission - a living systematic review. Clin Microbiol Infect 2021. [Internet][cited 2021 Mar 26];Available from https://www.sciencedirect.com/science/article/pii/S1198743X21000380?via%3Dihub . doi:10.1016/j.cmi.2021.01.011.

[6] Struyf T, Deeks JJ, Dinnes J, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. Cochrane Database Syst Rev 2021;2:CD013665. doi:10.1002/14651858.CD013665.pub2.

[7] Wynants L, Calster BV, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369:m1328. doi:10.1136/bmj.m1328.

[8] Instituto de Salud Carlos III. National Study of SARS-CoV-2 Sero-Epidemiology in Spain (ENE-COVID) [Internet]. 2020 [cited 2020 Jul 24]. Available from: https://portalcne.isciii.es/enecovid19/

[9] Pastor-Barriuso R, Pérez-Gómez B, Hernán MA, et al. Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: nationwide seroepidemiological study. BMJ 2020;371:m4509. Supplementary material.. doi:10.1136/bmj.m4509.

[10] Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. Am J Epidemiol 2004;160:301–5. doi:10.1093/aje/kwh221.

[11] Kass GV. An exploratory technique for investigating large quantities of categorical data. Appl Stat 1980;29:119. doi:10.2307/2986296.

[12] Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

[13] Perkins NJ, Schisterman EF. The inconsistency of 'optimal' cut-points obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163:670–5. doi:10.1093/aje/kwj063.

[14] Meyerowitz EA, Richterman A, Bogoch II, Low N, Cevik M. Towards an accurate and systematic characterisation of persistently asymptomatic infection with SARS-CoV-2. Lancet Infect Dis 2020. [Internet][cited 2021 Feb 26];Available from: https://www.sciencedirect.com/science/article/pii/S1473309920308379?via%3Dihub . doi:10.1016/S1473-3099(20)30837-9.

[15] Health Information and Quality AuthorityHIQA Evidence summary of the immune response following infection with SARS-CoV-2 or other human coronaviruses [Internet], Ireland: HIQA; 2020. Aug. Report No.: V3.0. Available from https://www.hiqa.ie/sites/default/files/2020-08/Evidence-summary_SARS-CoV-2-immune-response.pdf .

[16] Long Q-X, Tang X-J, Shi Q-L, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat Med 2020;26:1200–4. doi:10.1038/s41591-020-0965-6.

[17] Lavezzo E, Franchin E, Ciavarella C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. Nature 2020;584:425–9. doi:10.1038/s41586-020-2488-1.

[18] Walsh KA, Jordan K, Clyne B, et al. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. J Infect 2020;81:357–71. doi:10.1016/j.jinf.2020.06.067.

[19] Lee S, Kim T, Lee E, et al. Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. JAMA Intern Med 2020;180:1447. doi:10.1001/jamainternmed.2020.3862.

[20] Johansson MA, Quandelacy TM, Kada S, et al. SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. JAMA Netw Open 2021;4:e2035057. doi:10.1001/jamanetworkopen.2020.35057.

[21] Voysey M, Costa Clemens SA, Madhi SA, et al. Single Dose Administration, And The Influence Of The Timing Of The Booster Dose On Immunogenicity and Efficacy Of ChAdOx1 nCoV-19 (AZD1222) Vaccine [Internet], NY: Rochester; 2021. *Social Science Research Network*Feb [cited 2021 Feb 4]. Report No.: ID 3777268. Available from https://papers.ssrn.com/abstract=3777268 .

[22] Pritchard MG, Group ICC. COVID-19 symptoms at hospital admission vary with age and sex: ISARIC multinational study. medRxiv; 2020. [Internet][cited 2020 Dec 21]Available from https://www.medrxiv.org/content/10.1101/2020.10.26.20219519v2 . doi:10.1101/2020.10.26.20219519.

[23] Teherán AA, Ramos GC, Guardia RP, et al. Epidemiological characterisation of asymptomatic carriers of COVID-19 in Colombia: a cross-sectional study. BMJ Open 2020;10:e042122. doi:10.1136/bmjopen-2020-042122.

[24] Ladwig KH, Marten-Mittag B, Formanek B, Dammann G. Gender differences of symptom reporting and medical health care utilization in the German population. Eur J Epidemiol 2000;16:511–18. doi:10.1023/a:1007629920752.

[25] Bardel A, Wallander M-A, Wallman T, et al. Age and sex related self-reported symptoms in a general population across 30 years: Patterns of reporting and secular trend. PLoS One 2019;14:e0211532. doi:10.1371/journal.pone.0211532.

[26] Götzinger F, Santiago-García B, Noguera-Julián A, et al. COVID-19 in children and adolescents in Europe: a multinational, multicentre cohort study. Lancet Child Adolescent Health 2020;4:653–61. doi:10.1016/S2352-4642(20)30177-2.

[27] Arons MM, Hatfield KM, Reddy SC, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N Engl J Med 2020;382:2081–90. doi:10.1056/NEJMoa2008457.

[28] Sayampanathan AA, Heng CS, Pin PH, et al. Infectivity of asymptomatic versus symptomatic COVID-19. Lancet 2020;397:93–4. doi:10.1016/S0140-6736(20)32651-9.

[29] Jones TC, Biele G, Mühlemann B, et al. Estimating infectiousness throughout SARS-CoV-2 infection course; 2021 Science:eabi5273. doi:10.1126/science.abi5273.

[30] Marks M, Millat-Martinez P, Ouchi D, et al. Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study. Lancet Infectious Dis 2021. [Internet][cited 2021 Feb 4]Available from https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30985-3/abstract . doi:10.1016/S1473-3099(20)30985-3.

[31] Watanabe M, Balena A, Tuccinardi D, et al. Central obesity, smoking habit, and hypertension are associated with lower antibody titres in response to COVID-19 mRNA vaccine. Diabetes Metab Res Rev 2021. doi:10.1002/dmrr.3465.

[32] Simons D, Shahab L, Brown J, Perski O. The association of smoking status with SARS-CoV-2 infection, hospitalisation and mortality from COVID-19: A living rapid evidence review with Bayesian meta-analyses (version 11). Qeios; 2021. [Internet][cited

2021 May 15]Available from https://www.qeios.com/read/UJR2AW.13 . doi:10.32388/UJR2AW.13.
[33] Wenzl T. Smoking and COVID-19: Did we overlook representativeness? Tob Induc Dis 2020;18:1–2. doi:10.18332/tid/129584.
[34] Burke RM, Killerby Marie E, Newton Suzanne, et al. Symptom Profiles of a Convenience Sample of Patients with COVID-19 — United States, January–April 2020. MMWR Morb Mortal Wkly Rep 2020;69:904–8. doi:10.15585/mmwr.mm6928a2.
[35] WHO. COVID-19 Case definition [Internet]. 2020 [cited 2021 Feb 1]. Available from: https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-Surveillance_Case_Definition-2020.2
[36] 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. doi:10.1038/s41591-020-0916-2.
[37] Lechien JR, Chiesa-Estomba CM, De Siati DR, et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. Eur Arch Otorhinolaryngol 2020;277:2251–61. doi:10.1007/s00405-020-05965-1.
[38] Lao WP, Imam SA, Nguyen SA. Anosmia, hyposmia, and dysgeusia as indicators for positive SARS-CoV-2 infection. World J Otorhinolaryngol Head Neck Surg 2020;6:S22–5. doi:10.1016/j.wjorl.2020.04.001.
[39] Elliott J, Whitaker M, Bodinier B, et al. Symptom reporting in over 1 million people: community detection of COVID-19. medRxiv; 2021. 2021.02.10.21251480. doi:10.1101/2021.02.10.21251480.
[40] Sun Y, Koh V, Marimuthu K, et al. Epidemiological and Clinical Predictors of COVID-19. Clin Infect Dis 2020;71:786–92. doi:10.1093/cid/ciaa322.
[41] The Nacional SARS-CoV-2 Serology Assay Evaluation group. Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. Lancet Infectious Dis 2020;20:1390–400. doi:10.1016/S1473-3099(20)30634-4.
[42] Ripperger TJ, Uhrlaub JL, Watanabe M, et al. Orthogonal SARS-CoV-2 serological assays enable surveillance of low-prevalence communities and reveal durable humoral immunity. Immunity 2020;53:1–9. doi:10.1016/j.immuni.2020.10.004.
[43] Working group for the surveillance and control of COVID-19 in Spain. The first wave of the COVID-19 pandemic in Spain: characterisation of cases and risk factors for severe outcomes, as at 27 April 2020. Eurosurveillance 2020;25:2001431. doi:10.2807/1560-7917.ES.2020.25.50.2001431.
[44] Schwarz N. Self-reports: How the questions shape the answers. Am Psychol 1999;54:93–105. doi:10.1037/0003-066X.54.2.93.