EPICOVID19: Psychometric assessment and validation of a short diagnostic scale for a rapid Covid-19 screening based on reported symptoms

Luca Bastiani, Loredana Fortunato, Stefania Pieroni, Fabrizio Bianchi, Fulvio Adorni, Federica Prinelli, Andrea Giacomelli, Gabriele Pagani, Stefania Maggi, Caterina Trevisan, Marianna Noale, Nithiya Jesuthasan, Aleksandra Sojic, Carla Pettenati, Massimo Andreoni, Raffaele Antonelli Incalzi, Massimo Galli, Sabrina Molinaro

Submitted to: Journal of Medical Internet Research
on: August 27, 2020

Table of Contents

EPICOVID19: Psychometric assessment and validation of a short diagnostic scale for a rapid Covid-19 screening based on reported symptoms

Luca Bastiani[1] PhD; Loredana Fortunato[1] MSc; Stefania Pieroni[1] MSc; Fabrizio Bianchi[1] PhD; Fulvio Adorni[2] MPH; Federica Prinelli[2] PhD; Andrea Giacomelli[3] MD; Gabriele Pagani[3] MD; Stefania Maggi[4] PhD; Caterina Trevisan[5] MD; Marianna Noale[4] MSc; Nithiya Jesuthasan[2]; Aleksandra Sojic[2] PhD; Carla Pettenati[2] MD; Massimo Andreoni[6] MD; Raffaele Antonelli Incalzi[7] MD; Massimo Galli[3] MD; Sabrina Molinaro[1] PhD

[1]National Research Council Pisa IT
[2]National Research Council Milano IT
[3]Infectious Diseases Unit Milano IT
[4]National Research Council Padova IT
[5]University of Padova Padova IT
[6]Infectious Diseases Clinic Rome IT
[7]Unit of Geriatrics Rome IT

Corresponding Author:
Sabrina Molinaro PhD
National Research Council
Via Moruzzi 1 Pisa
Pisa
IT

Abstract

Background: Confirmed COVID-19 cases have been registered in more than two hundred countries and regions and of July 28 over 16 million cases of COVID-19, including 650805 deaths, have been reported to WHO. The number of cases changes quickly and varies depending upon which source you use to track, so in the current epidemiological context, the early recognition is critical for the rapid identification of suspected cases (with SARS-CoV-2 infection-like symptoms and signs) to be immediately subjected to quarantine measures. Although surveys are widely used for identifying COVID-19 cases, outcomes and associated risks, no validated epidemiological tool exists for surveying SARS-CoV-2 infection in the population so far.

Methods: Our study is the phase II of the EPICOVID19 Italian national survey, launched in April 2020 including a national convenience sample of 201121 adults, who voluntarily filled the EPICOVID19 questionnaire. The phase II questionnaire was mailed to all subjects who underwent tests for COVID-19 by nasopharyngeal swab (NPS) and who accepted to be involved in the second phase of the study, focused on the results reported for NPS and/or serological IgG/IgM tests. We evaluated the capability of the self-reported symptoms collected through the EPICOVID19 questionnaire to discriminate the COVID-19 among symptomatic subjects, in order to identify possible cases to undergo instrumental measurements and clinical examinations. We defined a method for the identification of a total score and validated it with reference to the serological and molecular clinical diagnosis, using four standard steps: identification of critical factors, confirmation of presence of latent variable, development of optimal scoring algorithm and validation of the scoring algorithm.

Findings

2703 subjects [66% response rate] completed the Phase II questionnaire. Of 2703 individuals, 694 (25.7%) were NPS(+) and of these 84 (12.1% of the 694 NPS(+)) were asymptomatic. In the individuals who performed serological testing, of the 472 who did IgG(+) and 421 who did IgM(+), 22.9% and 11.6% tested positive, respectively. Among IgG(+) 1 of 108 subjects was asymptomatic (0.9%) while 5/49 subjects among IgM(+) were asymptomatic (10.2%). Compared with NPS(-), among NPS(+) subjects there was a higher rate for Fever (421 [60.7%] vs 391[19.5%]; p<0.0001), Loss of Taste and/or Smell (365 [52.6%] vs 239 [11.9%]; p<0.0001) and Cough (352 [50.7%] vs 580 [28.9%]; p<0.0001). Also for other symptoms the frequencies were significantly higher in NPS(+) subjects than in NPS(-) ones (p<0.001). Among groups with serological tests, the symptoms with higher percentages in the subjects IgG(+) were Fever (65 [60.2%] vs 43[11.8%]; p<0.0001) and Pain in muscles, bones, joints (73 [67.6%] vs 71 [19.5%]; p<0.0001). For the COVID-19 self-reported symptoms items, exploratory (proportion variance

explained [89.9%]) and confirmatory factor analysis results (SMSR 0.072; RMSEA 0.052) highlights the presence of one latent variable (factor) underlying the symptoms. We define the one-factor solution as EPICOVID19 diagnostic scale and optimal score for each items was identified: Respiratory problems (1.03), Chest pain (1.07), Loss of Taste and/or Smell (0.97) and Tachycardia (palpitations) (1.05) were the most important symptoms. The cut-off score was 2.56 (Sensitivity 76.56%; Specificity 68.24%) in NPS(+) and 2.59 (Se 80.37; Sp 80.17) in IgG(+) subjects.
Interpretation
We developed a short diagnostic scale to detect subjects with symptoms potentially associated with COVID-19 among a wide population. Early recognition screening and rapid diagnosis are essential to prevent transmission and provide supportive care in a timely manner and our score supports the potential for identifying individuals who need to seek immediate clinical evaluation. Although these results are referred to the Italian pandemic period, this short diagnostic scale could be optimised and tested as a screening tool in other similar pandemic contexts.

(JMIR Preprints 27/08/2020:23897)
DOI: https://doi.org/10.2196/preprints.23897

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

EPICOVID19: Psychometric assessment and validation of a short diagnostic scale for a rapid Covid-19 screening based on reported symptoms

Luca Bastiani[1], Loredana Fortunato[1], Stefania Pieroni[1], Fabrizio Bianchi[1], Fulvio Adorni [2], Federica Prinelli[2], Andrea Giacomelli[3], Gabriele Pagani[3], Stefania Maggi[4], Caterina Trevisan[5], Marianna Noale[4], Nithiya [6] Jesuthasan, Aleksandra Sojic[6], Carla Pettenati[6], Massimo Andreoni[7], Raffaele Antonelli Incalzi[8], Massimo Galli[9], Sabrina Molinaro[1]

[1] Luca Bastiani, PhD, luca.bastiani@ifc.cnr.it, National Research Council, Institute of Clinical Physiology, Via G. Moruzzi 1, 56124 Pisa (PI), Italy

[1] Loredana Fortunato, MSc, loredana.fortunato@ifc.cnr.it, National Research Council, Institute of Clinical Physiology, Via G. Moruzzi 1, 56124 Pisa (PI), Italy

[1] Stefania Pieroni, MSc, stefania.pieroni@ifc.cnr.it, National Research Council, Institute of Clinical Physiology, Via G. Moruzzi 1, 56124 Pisa (PI), Italy

[1] Fabrizio Bianchi, PhD, fabriepi@ifc.cnr.it, National Research Council, Institute of Clinical Physiology, Via G. Moruzzi 1, 56124 Pisa (PI), Italy

[2] Fulvio Adorni, MPH, fulvio.adorni@itb.cnr.it, National Research Council, Institute of Biomedical Technologies, Via Fratelli Cervi 93, 20090 Segrate (MI), Italy

[2] Federica Prinelli, PhD, federica.prinelli@itb.cnr.it, National Research Council, Institute of Biomedical Technologies, Via Fratelli Cervi 93, 20090 Segrate (MI), Italy

[3] Andrea Giacomelli, MD, andrea.giacomelli@unimi.it, Infectious Diseases Unit, Department of Biomedical and Clinical Sciences L.Sacco, Università di Milano, ASST Fatebenefratelli Sacco, Milan, Italy Via G.B. Grassi 74, 20157 Milano, Italy

[3] Gabriele Pagani, MD, gabriele.pagani@unimi.it, Infectious Diseases Unit, Department of Biomedical and Clinical Sciences L.Sacco, Università di Milano, ASST Fatebenefratelli Sacco, Milan, Italy Via G.B. Grassi 74, 20157 Milano, Italy

[4] Stefania Maggi, PhD, stefania.maggi@in.cnr.it, National Research Council, Institute of Neuroscience, Aging branch, Via Vincenzo Maria Gallucci 16, 35128 Padova

[5] Caterina Trevisan, MD, caterina.trevisan.5@studenti.unipd.it, University of Padova, Department of Medicine (DIMED) – Geriatric Unit, Via Giustiniani 2, 35128 Padova (PD), Italy - National Research Council, Institute of Neuroscience, Via Giustiniani 2, 35128 Padova (PD), Italy

[4] Marianna Noale, MSc, marianna.noale@in.cnr.it, National Research Council, Institute of Neuroscience, Aging branch, Via Vincenzo Maria Gallucci 16, 35128 Padova

[6] Nithiya Jesuthasan, MPH, nithiya.jesuthasan@itb.cnr.it, National Research Council, Institute of Biomedical Technologies, Via Fratelli Cervi 93, 20090 Segrate (MI), Italy

[6] Aleksandra Sojic, PhD, aleksandra.sojic@itb.cnr.it, National Research Council, Institute of Biomedical Technologies, Via Fratelli Cervi 93, 20090 Segrate (MI), Italy

[6] Carla Pettenati, MD, cpettenati@me.com, National Research Council, Institute of Biomedical Technologies, Via Fratelli Cervi 93, 20090 Segrate (MI), Italy

[7] Massimo Andreoni, MD, andreoni@uniroma2.it, Infectious Diseases Clinic, Department of System Medicine, Tor Vergata University of Rome, 00133 Rome, Italy

[8] Raffaele Antonelli Incalzi, MD, r.antonelli@unicampus.it, Unit of Geriatrics, Department of Medicine, Biomedical Campus of Rome, via Alvaro del Portillo, 21, 00128 Rome, Italy

[9] Massimo Galli, MD, massimo.galli@unimi.it, Infectious Diseases Unit, Department of Biomedical and Clinical Sciences L.Sacco, Università di Milano, ASST Fatebenefratelli Sacco, Milan, Italy Via G.B. Grassi 74, 20157 Milano, Italy

[1] Sabrina Molinaro, PhD, sabrina.molinaro@ifc.cnr.it, National Research Council, Institute of Clinical Physiology, Via G. Moruzzi 1, 56124 Pisa (PI), Italy

Corresponding Author:

Sabrina Molinaro, Ph.D.

Epidemiology and Health Research Lab,
Institute of Clinical Physiology, IFC
National Research Council of Italy - CNR
Via G. Moruzzi, 1
56124 Pisa - Italy
Tel. +39 050 315 2094
Fax +39 050 3152095
mail sabrina.molinaro@ifc.cnr.it
web http://www.epid.ifc.cnr.it

EPICOVID19: Psychometric assessment and validation of a short diagnostic scale for a rapid Covid-19 screening based on reported symptoms

Abstract

Background

Coronavirus disease (COVID-19) confirmed cases have been registered in more than two hundred countries and, of July 28, over 16 million cases have been reported to WHO. The present study was conducted while the epidemic peak of COVID-19 was occurring in Italy. The early recognition is critical for the identification of suspected cases to be immediately subjected to quarantine. Although surveys are widely used for identifying COVID-19 cases, outcomes and associated risks, no validated epidemiological tool exists for surveying (SARS-CoV-2) infection in the population so far.

Objective

We evaluated the capability of self-reported symptoms to discriminate the COVID-19, in order to identify possible cases to undergo instrumental measurements. We defined and validated a method for the identification of a cut-off score.

Methods

Our study is the phase II of the EPICOVID19 Italian national survey, launched in April 2020 including a convenience sample of 201121 adults, who filled the EPICOVID19 questionnaire. The Phase II questionnaire, focused on results for nasopharyngeal swab (NPS) and/or serological tests, was mailed to all subjects who previously underwent NPS tests.

Results

Of 2703 subjects who completed the Phase II questionnaire, 694 (25.7%) were NPS(+). In the individuals who performed serological testing, of the 472 who did Immunoglobulin-G (IgG) test and 421 who did Immunoglobulin-M (IgM) test, 22.9% (108 out of 472) and 11.6% (49 out of 421) tested positive, respectively. Compared with NPS(-), among NPS(+) subjects there was a higher rate for Fever (421 [60.7%] vs 391[19.5%]; *P*<.001), Loss of Taste and/or Smell (365 [52.6%] vs 239 [11.9%]; *P*<.001) and Cough (352 [50.7%] vs 580 [28.9%]; *P*<.001). Among groups with serological tests, the symptoms with higher percentages in the subjects IgG(+) were Fever (65 [60.2%] vs 43[11.8%]; *P*<.001) and Pain in muscles/bones/joints (73 [67.6%] vs 71 [19.5%]; *P*<.001). Analysis of COVID-19 self-reported symptoms items revealed the presence of a one-factor solution, i.e. the EPICOVID19 diagnostic scale. The optimal score was identified: Respiratory problems (1.03), Chest pain (1.07), Loss of Taste and/or Smell (0.97) and Tachycardia (palpitations) (1.05) were the most important symptoms. On the adults aged 18-84 years the cut-off score was 2.56 (Se 76.56%; Sp 68.24%) in NPS(+) and 2.59 (Se 80.37; Sp 80.17) in IgG(+) subjects. On the >=60 years group the cut-off score was 1.28 and the accuracy detected by Antibodies IgG(+) was improved (Se 88.00; Sp 89.58).

Conclusion

We developed a short diagnostic scale to detect subjects with symptoms potentially associated with COVID-19 among a wide population, which supports the potential for identifying individuals who need to seek immediate clinical evaluation. Although these results are referred to the Italian pandemic period, this short diagnostic scale could be optimised and tested as a screening tool in other similar pandemic contexts.

Introduction

SARS-CoV-2 has led to a global pandemic; on July 28 over 16 million cases and 650805 deaths across more than 200 countries were reported by WHO and Johns Hopkins Center for Health Security [1-2]. Italy was the first European country to be hit hard by the COVID-19 epidemic and it was also the European country in which the highest number of COVID-19 deaths have been recorded (24,780 as of 27 April 2020) [3]. Besides this immediate human toll there are readily acknowledged and potentially long-lasting effects on global economies, politics, health and privacy policies at many levels that will extend beyond the development of vaccines and treatments. The rapid spread of the disease, COVID-19, and its seemingly high degree of variability in its presentation among individuals has led to a level of clinical and scientific focus not previously seen and encompassing both traditionally reviewed and pre-print publications and resources. Collaborative groups are being formed at the local, regional, national and international levels to address both patient data collection/ aggregation and analysis in ways that may change the way research is carried out in the future [4]. To enable these efforts to be both effective and productive is the need for the data to be evaluated as to its suitability for inclusion in these activities while still recognizing that what we understand about COVID-19 is much less than what we do not understand [5].

Because of the far-reaching scope of the pandemic, we are already confronting:
1. Need to implement individual testing at a level far above current capacities to optimize individual treatment, assess disease spread, anticipate potential strains on healthcare resources and personnel [6].
2. Need for improvements in available testing, both NPS and antibody detection, i.e. accuracy, specificity and sensitivity to enable reliable evaluation and interpretation of data for use in clinical care and policy decisions [7].
3. Need to harmonize clinical observations and definitions to support development of guidelines, prognostic and diagnostic indicators and to develop a comprehensive understanding of the disease and critical factors that differentiate patient susceptibility, presentation of the disease and response to treatment [8-9].
The use of online surveys can greatly enhance access to broader populations in a cost-effective manner and optimize both screening for individuals who may need immediate care as well as provide an approach to 3) above. A cross-sectional national survey, EPICOVID19, was launched on April 13, 2020 and received more than 200000 responses [10]. The survey, which represents the Phase I of the study, was promoted using social media (Facebook, Twitter, Instagram, Whatsapp), press releases, internet pages, local radio and television stations, and institutional websites that called upon volunteers to contact the study website. The inclusion criteria were: age of >18 years; access to a mobile phone, computer, or tablet with internet connectivity; and on-line consent to participate in the study.
The present study was conducted while the epidemic peak of COVID-19 was occurring in Italy. The aim of our study is to assess the capability of the self-reported symptoms collected through the EPICOVID19 questionnaire to discriminate the COVID-19 among symptomatic subjects, in order to identify possible COVID-19 cases to undergo instrumental measurements and clinical examinations (Phase II of the study). The final objectives are to propose a method for the development of a total score for the self-reported symptoms in the EPICOVID19 questionnaire and to validate the scoring method with reference to the molecular and serological clinical diagnosis.

Methods

Study design and participants

Our study is the phase II of the EPICOVID19 Italian national survey [9] (appendix p 1-8), launched in April 2020 including a convenience sample of 201121 adults, who filled the EPICOVID19 questionnaire. The Figure 1 shows the overview of EPICOVID19 two-phase study. The Phase I questionnaire investigated six areas through 38 questions (1: Socio-demographic characteristics; 2: Clinical evaluation; 3: Personal characteristics and health status; 4: Housing conditions; 5: Lifestyle; 6: Behaviours after the lockdown).
The Phase II questionnaire was mailed to all subjects who underwent testing for COVID-19 by NPS, and who volunteered to be involved in the follow up study in their Phase I response. Phase II focused on results reported for NPS and/or serological IgG/IgM tests and self-reported symptoms with the aim to better identify both symptomatic and asymptomatic cases of SARS-CoV-2 infections [10].

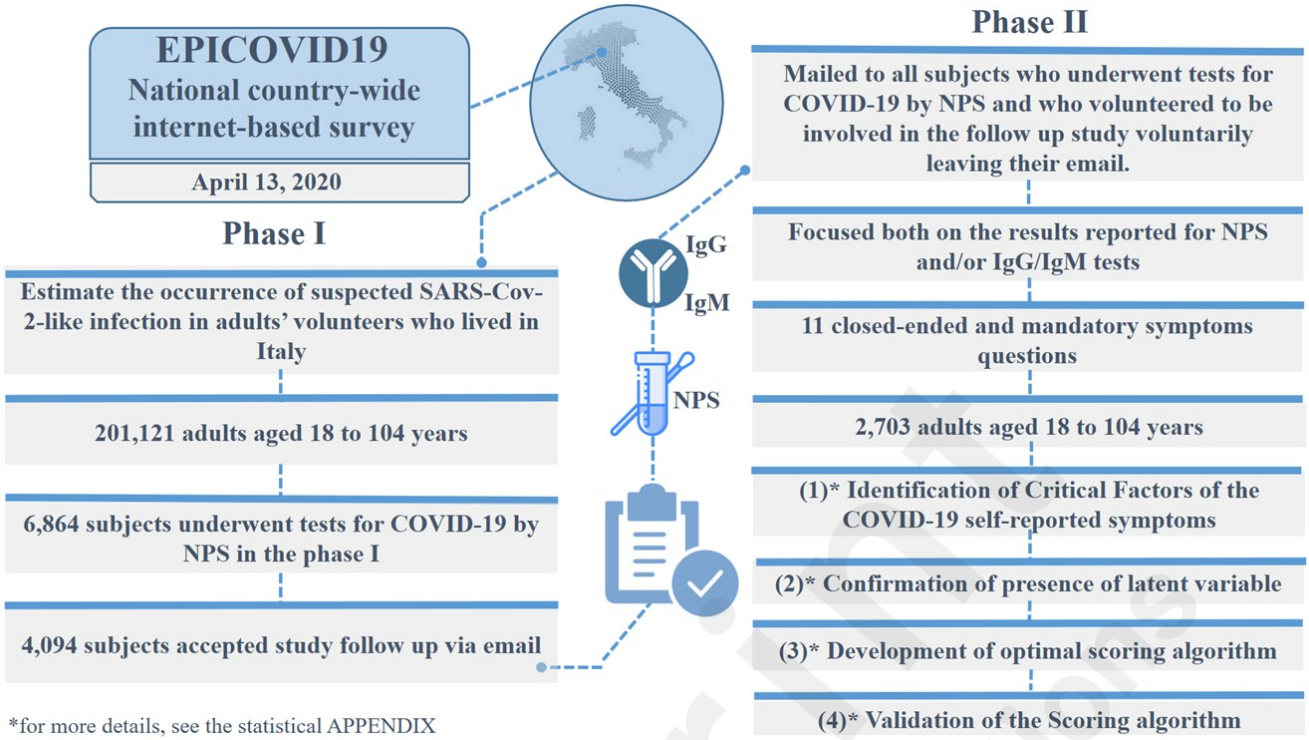


Figure 1. Overview of EPICOVID19 two-phase study

Phase II was implemented using an open source statistical survey framework, LimeSurvey (vers. 3.17). This is a PHP based framework, distributed under the GNU General Public License.

In Phase II, responses to 11 questions were required that covered administration of NPS and/or serological tests and on the time elapsed between observed/reported symptoms and clinical examination (NPS and/or antibodies IgG and IgM) (appendix p 1-8).

Of the 6864 subjects who underwent testing for COVID-19 by NPS in the phase I, 4094 subjects were invited by e-mail to complete the Phase II questionnaires online. Of these, 38 email invitations were not delivered, e.g. wrong address, mailbox full, host or domain name not found, etc.; 101 individuals did not participate by refusing to provide consent; and 1252 individuals who received the email did not proceed to complete the questionnaire.

The online survey was implemented using questions with close-ended answers, in order to facilitate the questionnaire compilation and to avoid errors in digitizing answer values. On 2nd May, at the end of the italian lockdown period, the survey was closed and all collected data were exported to be analysed with statistical tools. The base data for the statistical analysis was structured as a table containing one row for each survey participant and as many columns as the collected responses. The questionnaire is available in the appendix (appendix p 1-16).

A total of 2703 subjects (66% response rate) completed the phase II survey. Considering the 6864 subjects who reported to have performed the NPS test in the Phase I survey, we compared the characteristics of respondents (2703) and non-respondents (4161). Respondents and non-respondents to the Phase II survey appeared similar with respect to gender, age and the perception of their own health as well as their self-reported comorbidities; the details comparing these two groups of subjects are included in appendix p 9. The resulting data of 2703 subjects who completed the Phase II were linked to the results of the self-reported symptoms of Phase I EPICOVID19 questionnaire, which included questions on the presence of 11 symptoms.

Statistical analysis

We analysed the self-reported symptoms collected in the survey to define a method for the construction of a total score and to validate the scoring method with reference to the serological and molecular clinical diagnosis, using four standard questionnaire validation steps:

(1) Identification of Critical Factors: We determined the factorial structure of the COVID-19 self-reported symptoms items using exploratory factor analysis (EFA) followed by confirmatory factor analysis (CFA). EFA and parallel analysis (PA) were performed to identify the performance of specific symptoms (loadings) and to define the number of factors underlying these loadings.
(2) Confirmation of presence of latent variable: We carried out CFA via Structural Equation Modelling (SEM) to confirm the presence of one latent variable (factor) underlying the 11 symptoms chosen to identify COVID-19. Several goodness-of-fit criteria were used: (i) Standardized Root Mean Square Residual (SRMR) and (ii) Root Mean Square Error of Approximation (RMSEA) (not higher than 0.10); (iii) Comparative fit index (CFI) and (iv) Tucker-Lewis index (TLI) (not less than 0.90).
(3) Development of optimal scoring algorithm: We developed an optimal scoring algorithm using Homogeneity Analysis (HOMALS) and Multiple Correspondence Analysis (MCA). Through the HOMALS procedure we replaced specific dichotomous responses, i.e. Yes/No, with categorical quantifications: the resulting score is the sum of the subject's symptom responses, once they are re-coded by category quantifications.
(4) Validation of the Scoring algorithm: We validated the score using an external objective criterion based on receiver operating characteristics (ROC) analysis, in order to evaluate the COVID-19 symptoms score performance in distinguishing symptomatic individuals, both in the complete sample (aged between 18 and 84 years) and into two specific age groups (<60 and >= 60 years). With the aim of discriminating COVID-19 cases, the sensitivity, the specificity, and the Youden's index were computed with two reference standards: a) subjects with positive NPS vs subjects with negative NPS; b) subjects with serological IgG(+) vs subjects with IgG(-). The overall predictive performance was evaluated by the area under the curve (AUC).
All statistical analyses were carried out using R software (version 3.6.3), IBM® SPSS® 23 Stata Statistical Software (Release 15). The details of the performed statistical analysis are reported in appendix p 10-12.

Results

Study design and participants

The characteristics of the 2703 subjects, as supplied by those who completed the phase II survey, their NPS, IgG and IgM results are shown in table 1. The sample was predominantly women (68.1%) with the average age being 49 ± 15.0 and 52 ± 14.1 years for women and male individuals, respectively. Of the 2703 respondents, 151 (5.6%) had a lower educational status, 837 out of 2703 (31.0%) medium educational status and the majority had a high educational status (1715 out of 2703, 63.4%). The chronic conditions most reported by participants were hypertension (361 out of 2703, 13.4%), following the immune system diseases (266 out of 2703, 9.8%) and the depression and/or anxiety diseases (194 out of 2703, 7.2%). The chronic conditions reported less frequently were liver (21 out of 2703, 0.8%) and kidney diseases (22 out of 2703, 0.8%). All the details are reported in appendix p 13.

Statistical analysis
In the total sample of 2703 individuals, 694 (25.7%) were NPS(+) and of these 84 (12.1% of the 694 NPS(+)'s) were asymptomatic. For the subgroup of individuals who performed serological testing, of the 472 who did IgG test and 421 who did IgM test, 22.9% (108/472) and 11.6% (49/421) tested positive, respectively. Among IgG(+) 1 of 108 subjects was asymptomatic (0.9%) while 5/49 subjects among IgM(+) were asymptomatic (10.2%). For subjects with NPS(+), the average number of days between initial symptoms and the day of the swab execution was 9.3 ± 9.4 (median 7 days, interquartile range (IQR) 3-7). For subjects IgG(+) the average number of days between initial

symptoms and the day of the serological test execution was 36.1 ± 15.1 (median 36.5 days, IQR 28-47). For subjects IgM(+) the average number of days from initial symptoms to the day of the serological test execution was 26.1 ± 17.9 (median 28.0 days, IQR 4-40). The frequency of the eleven symptoms reported by the three groups (NPS, IgG, IgM tested) was similar among men and women. In NPS(+) group, women reported higher percentages than male for the Sore throat and/or cold and/or Tachycardia (Palpitations) symptoms only. In the IgG(+), males reported higher frequencies for headaches only, while in the IgM(+) group the females showed the lower frequency of symptoms related to Conjunctivitis.

The frequency of symptoms among NPS(+) subjects (table 1) ranged from a low rate of observation, e.g. Tachycardia (palpitations) (S9: 17.3%) and Conjunctivitis (S11: 16.0%), to a high observation rate, e.g. Fever (S4: 60.7%) and olfactory and taste disorders (OT-D) (S6: 52.6%). For all symptoms, apart from Headache, the frequencies were significantly higher in NPS(+) subjects than in NPS(-) ones (P<.001). For the subgroup of individuals who also performed serological tests, the symptoms with higher percentages among the subjects tested positive were Fever (IgG(+) 60.2%; IgM(+) 57.1%) and Pain in muscles, bones, joints IgG(+) 67.6%; IgM(+) 55.1%). In the IgG serological group, no difference was observed in the percentages of Sore throat and/or cold (S3) symptoms, while regarding Respiratory difficulty (S7), Chest pain (S8) and Gastrointestinal symptoms (S10), the percentages in the IgM group were the same.

	Tested[a] for SARS-CoV-2								
	NASOPHARYNGEAL SWAB (2703)			ANTIBODIES IgG (472)			ANTIBODIES IgM (421)		
	Tested positive	Tested negative	*P* value	Tested positive	Tested negative	*P* value	Tested positive	Tested negative	*P* value
Number	694 (25.7%)	2009 (74.3%)		108 (22.9%)	364 (77.1%)		49 (11.6%)	372 (88.4%)	

Women (%)	440 (63.4%)	1401 (69.7%)	.001	61 (56.5%)	258 (70.9%)	.005	25 (51.0%)	260 (69.9%)	.008
Age (years) mean & standard deviation	55.5 ±18.06	47.55 ±12.81	<.001	48.8 ±11.74	45.5 ±11.49	.009	50.6 ±10.56	45.8 ±11.69	.008
Answered questions on symptoms (n)									
(S1) Fever, with temperatures above 37.5 ° C for at least three consecutive days	421 (60.7)	391 (19.5%)	<.001	65 (60.2%)	43 (11.8%)	<.001	28 (57.1%)	68 (18.3%)	<.001
(S2) Cough	352 (50.7%)	580 (28.9%)	<.001	63 (58.3%)	76 (20.9%)	<.001	26 (53.1%)	95 (25.5%)	<.001
(S3) Sore throat and/or Cold	232 (33.4%)	756 (37.6%)	.048	46 (42.6%)	132 (36.3%)	.233	16 (32.7%)	135 (36.3%)	.617
(S4) Headache	313 (45.1%)	703 (35.0%)	<.001	61 (56.5%)	96 (26.4%)	<.001	23 (46.9%)	117 (31.5%)	.031
(S5) Pain in muscles, bones, joints	360 (51.9%)	572 (28.5%)	<.001	73 (67.6%)	71 (19.5%)	<.001	27 (55.1%)	98 (26.3%)	<.001
(S6) Loss of taste and / or smell	365 (52.6%)	239 (11.9%)	<.001	66 (61.1%)	29 (8.0%)	<.001	21 (42.9%)	55 (14.8%)	<.001
(S7) Respiratory difficulty (sense of breathlessness at rest),	179 (25.8%)	249 (12.4%)	<.001	21 (19.4%)	28 (7.7%)	<.001	7 (14.3%)	37 (9.9%)	.350
(S8) Chest pain (sternum pain)	136 (19.6%)	251 (12.5%)	<.001	26 (24.1%)	25 (6.9%)	<.001	7 (14.3%)	37 (9.9%)	.350
(S9) Tachycardia (Palpitations)	120 (17.3%)	237 (11.8%)	<.001	24 (22.2%)	27 (7.4%)	<.001	10 (20.4%)	31 (8.3%)	.007
(S10) Gastrointestinal complaints (diarrhoea, nausea, vomiting)	289 (41.6%)	452 (22.5%)	<.001	54 (50.0%)	65 (17.9%)	<.001	17 (34.7%)	87 (23.4%)	.084
(S11) Conjunctivitis (red eyes)	111 (16.0%)	221 (11.0%)	<.001	24 (22.2%)	35 (9.6%)	.001	11 (22.4%)	40 (10.8%)	.018

a. Numbers are mean ± SD for continuous variables (independent t-test) and frequency (%) for the categorical ones (Chi square test)

Table 1. Self-reported characteristics from the survey analysed using tests' results for SARS-CoV-2 infection

The EFA performed by Principal-Component Factors (PCF) and Horn's PA methods statistics pointed out one factor. Eigenvalues, descriptive indices, and goodness-of-fit indices of cumulative percentages of explained data variability obtained through EFA are displayed in table 2. PCF highlight only one factor with 89.9% of proportion of explained variability, while Horn's PA method identifies two factors with eigenvalues greater than 1.0, with 49.8% and 10.3% of proportion of explained variability, respectively.

Exploratory factor analysis	
Principal-Component Factors	Horn's Parallel Analysis

Factor	Eigenvalue	Proportion of explained variability	Cumulative explained variability	Eigenvalue	Proportion of explained variability	Cumulative explained variability
1	5.00	89.9%	89.9%	5.48	49.8	49.8%
2	-	-	-	1.14	10.3	60.1%

Table 2. Descriptive and Goodness of fit dimensionality indices from EFA of the eleven EPICOVID19 symptoms items in 2703 subjects, using PCF and Horn's PA methods with eigenvalue 1.

Based on a priori determined cut-off value the factor-loading greater than 0.35 was maintained. The factors loading rule of one factor solution extracted by PCF is available in appendix p 13. Then, the dimensionality indices of the one-factor solution, as the high variance explained by the factor (89.9%), confirms the presence of one latent variable underlying COVID-19 symptoms items. Therefore, we define the one-factor solution as the "EPICOVID19 diagnostic scale" (EPICOVID19 DS). On the basis of the CFA, results confirmed the latent construct as uni-dimensional and how the variables contributed to EPICOVID19 DS. The figure 2 shows the values of the standardized factor loadings for the one factor model. The magnitude of each factor loading higher than 0.4 indicates the importance of the corresponding item to EPICOVID19 DS. For example, Pain in muscles, bones, joints were the most important variables with a value equal to 0.814. The other variables with optimal specific validity index were Respiratory difficulty (sense of breathlessness at rest (0.688), Loss of taste and/or smell (0.724) and Gastrointestinal complaints with item-factor correlations equal 0.737. The lowest values were observed for the Sore throat and/or cold and Conjunctivitis items, 0.537 and 0.557, respectively. The goodness of fit (SMSR, RMSEA) of the EPICOVID19 DS was acceptable because two indexes were lower than 0.10 (SMSR 0.072; RMSEA 0.052; CFI 0.977; TLI 0.971). Finally, we compute CFA indexes to measure internal validity of the model (appendix p 14).

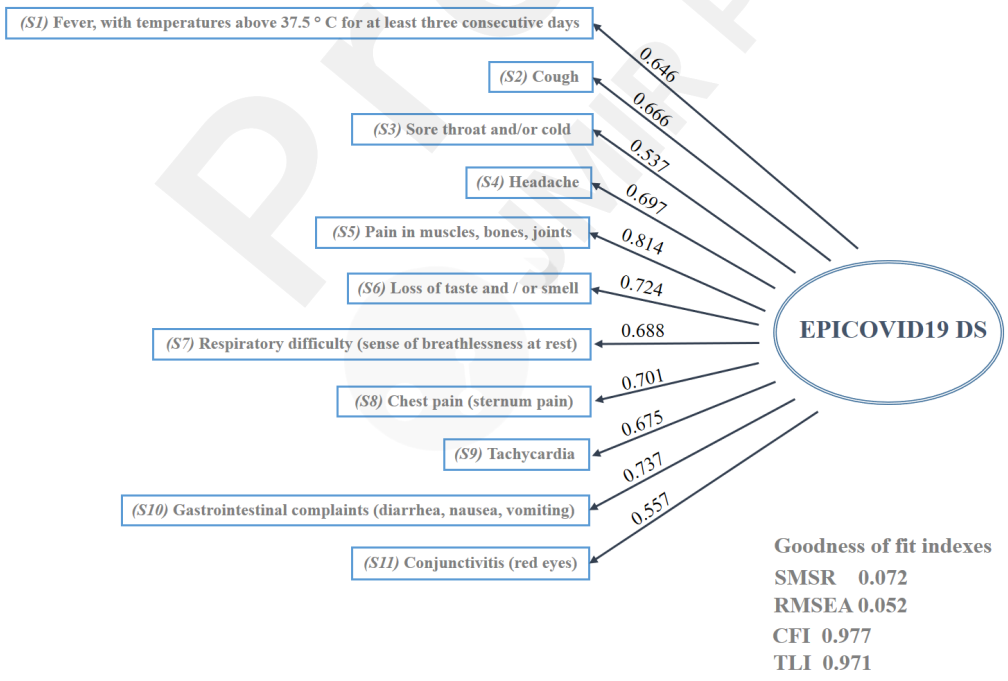


Figure 2. Standardized factor loadings and goodness of fit indexes for one factor model. The goodness of fit indexes (SMSR, RMSEA) of the "EPICOVID19 DS"

Given the successful uni-dimensionality testing of EPICOVID19 DS, optimal scaling via was undertaken. The optimal score proposed was extracted from the HOMALS procedure (single-factor measurement) and for each subject, the computed optimal score was obtained summing the category quantifications of the screening questionnaire item responses. Cronbach (0.88) and Greenacre (78%) indices confirm the uni-dimensionality found through EFA and CFA. The HOMALS optimal category quantifications of the EPICOVID19 symptoms variable are summarized in table 3. Columns show the binary options (YES-NO), while rows show the different symptoms. The HOMALS category quantifications were scaled so that the score obtained from the sum of responses, varies from 0 (if a subject answered NO to all the symptoms) to 10 (if a subject answered YES to all the symptoms). These values are shown in the last column of table 3. An example of the resulting score calculation is reported as follows. If the subject response pattern with respect to symptoms is: YES, NO, YES, NO, NO, YES, YES, NO, NO, NO, YES, the re-coded response pattern is: 0.80, 0, 0.64, 0, 0, 0.97, 1.03, 0, 0, 0, 0.88 and subject optimal score is: 0.8 + 0+ 0.64+0+ 0+ 0.97+ 1.03+ 0+ 0+ 0+ 0.88 = 4.2.

	HOMALS category quantifications		Recoded HOMALS category quantifications	
Symptoms	No	Yes	No	Yes
(S1) Fever, with temperatures above 37.5 ° C for at least 3 consecutive days	-0.362	0.8421	0	0.80
(S2) Cough	-0.426	0.810	0	0.81
(S3) Sore throat and/or cold	-0.358	0.622	0	0.64
(S4) Headache	-0.470	0.780	0	0.83
(S5) Pain in muscles, bones, joints	-0.505	0.959	0	0.97
(S6) Loss of taste and/or smell	-0.326	1.133	0	0.97
(S7) Respiratory difficulty (sense of breathlessness at rest)	-0.246	1.305	0	1.03
(S8) Chest pain (sternum pain)	-0.232	1.388	0	1.07
(S9) Tachycardia (palpitations)	-0.209	1.374	0	1.05
(S10) Gastrointestinal complaints (diarrhoea, nausea, vomiting)	-0.393	1.042	0	0.95
(S11) Conjunctivitis (red eyes)	-0.164	1.170	0	0.88

Table 3. Multiple Correspondence Analysis optimal weights recording of the EPICOVID19 DS

There was no significant difference in the EPICOVID19 DS mean score between men (2.34±2.2) and women (2.49±2.4), while a low negative correlation between the score and the age of the participants was found (Spearman rho: -0.126; <0.001). 64.3% of the sample reported no pre-existing disease, 25.3% had one chronic condition only, while the remaining 10.4% declared two or more conditions. Significant differences in the EPICOVID19 DS mean score were observed among the participants who did not report any disease (2.26±2.3) and those with at least one pre-existing condition (2.75±2.4) (*P* values <.001). Analyzing the EPICOVID19 DS mean score among the healthy group and the one chronic condition group, we observed significant differences in the lung diseases (healthy people: 2.40±2.3; lung: 3.10±2.5; *P* values <.001), immune system diseases (healthy people: 2.39±2.3; immune system: 2.91±2.4; *P* values <.001) and for depression and/or anxiety diseases (healthy people: 2.42±2.4; depression and/or anxiety: 2.79±2.6; *P* values <.05). For the other chronic conditions (heart, hypertension, kidney, tumors, metabolic and liver diseases), no differences in EPICOVID19 DS mean score were found.

The screening properties of the EPICOVID19 DS compared to COVID-19 positive molecular and serological diagnosis are shown in table 4. The best value of the Youden Index was observed for EPICOVID19 DS with respect to COVID-19 diagnosed by NPS(+). A good trade-off between sensitivity and specificity was observed (Se 76.56; Sp 68.24 AUC 77.5; 95% CI: 75.6-79.4). The cut-off score obtained was 2.56. Sensitivity and specificity improved when EPICOVID19 DS was compared to COVID-19 detected by Antibodies IgG(+) (Se 80.37; Sp 80.17 AUC 86.0; 95% CI: 82.3-89.5) and the cut-off obtained was the same as for NPS(+) (2.59). The positive and negative predictive value (PPV and NPV) for the IgG serological test was better than that for the nasopharyngeal swab (PPV IgG 54.43%, NPV IgG 93.27%). Finally, we observed a poor performance of the IgM results and are not presented in the table.

18 to 84 years	Tested for SARS-CoV-2			
	NASOPHARYNGEAL SWAB (2703) (NPS+ 694)		ANTIBODIES IgG (472) (IgG+ 108)	
Statistic	Value	95% CI	Value	95% CI
Sensitivity	76.56%	72.99% to 79.87%	80.37%	71.58% to 87.42%
Specificity	68.24%	66.16% to 70.28%	80.17%	75.69% to 84.14%
Positive Likelihood Ratio	2.41	2.23 to 2.61	4.05	3.23 to 5.08
Negative Likelihood Ratio	0.34	0.30 to 0.40	0.24	0.17 to 0.36
COVID-19 Disease %	23.29%	21.68% to 24.96%	22.77%	19.05% to 26.83%
PPV	42.26%	40.38% to 44.17%	54.43%	48.77% to 59.98%
NPV	90.55%	89.23% to 91.74%	93.27%	90.40% to 95.33%
Accuracy	70.18%	68.39% to 71.93%	80.21%	76.32% to 83.72%
Cut-off	2.59		2.56	

Table 4. Sensitivity and specificity of EPICOVID19 DC compared to COVID-19 positive molecular and serological diagnosis (18 to 84 years)

When the EPICOVID19 DS scoring algorithm was performed on specific age groups, the sensitivity and specificity detected by Antibodies IgG(+) (Se 88.00; Sp 89.58 AUC 93.10; 95% CI: 86.0-99.5) improved greatly among subjects equal or over 60 years and the obtained cut-off was lower (1.28) with respect to the younger ones (< 60 years) (2.71; Se 88.00; Sp 89.58 AUC 93.10; 95% CI: 86.0-99.5). The PPV and NPV for the IgG resulted better in the oldest subjects (>=60 years: PPV IgG 81.48%, NPV IgG 93.48%; <60 years: PPV IgG 51.52%, NPV IgG 94.38%). Furthermore, we observed the same performance of the NPS results among the specific age groups (>=60 years & <60 years) with respect to the overall sample (18 to 84 years). The details of the screening properties of the EPICOVID19 DS compared to COVID-19 positive molecular and serological diagnosis among specific age groups are reported in appendix p 16.

Discussion

Our focus was on developing a tool composed of simple questions related to the COVID-19 symptomatology for the identification of the subjects who are more likely to have SARS-CoV-2 infection in the general population. We validated the EPICOVID19 DS in a sample of voluntary subjects with serological and molecular clinical diagnosis. The optimal score, computed in 2703 adults aged 18 to 84 years, discriminates among symptomatic individuals. Before constructing the score, we performed both exploratory and confirmatory factor analysis to determine the number of factors/dimensions underlying the questionnaire. The results of these analyses supported a one factor model and the uni-dimensionality of the EPICOVID19 questionnaire. The magnitude of all factor loading was satisfactory, showing the highest factor loading value for the Respiratory difficulty, Chest pain, Tachycardia (palpitations) and Loss of taste and/or smell, Gastrointestinal complaints items appeared to be the most essential features of the EPICOVID19 DS. The high value for Chest pain is also explained by the fact that several patients reported it, possibly because of a tracheal pain caused by pneumonia [11-12]. Several clinical studies on hospitalised patients have shown that, at the onset of COVID-19, patients frequently show typical symptoms of viral pneumonia [3]. Symptoms that are less common, but still reported by a substantial number of patients are nasal congestion, sore throat, Gastrointestinal complaints and OT-D [13-15]. Subjects often reported Gastrointestinal complaints, not as isolated symptomatology of SARS-CoV-2 infection but as concurrent symptoms [16]. The lowest factor loading value was observed for the Sore throat and/or cold and Conjunctivitis. These lower values may be related to the fact that conjunctivitis and cold were not the most frequent symptoms of COVID-19 [17]. In line with other recent works [18-19], the features encountered showed various aspects of the COVID-19 diagnosis definition. Indeed, Cough, loss of Taste and/or Smell and Respiratory difficulty were among the most reported symptoms in previous researches and corresponded to the items that gained the most importance in our score [11, 15, 20, 21].

However, the clinical presentation of COVID-19 disease is varied, and discrepancy may exist between symptoms and disease. A recent meta-analysis of symptoms including 50000 COVID-19 patients, found that fever and cough were the most common symptoms [22] (89.1% and 72.2%) and a separate study of hospitalized subjects has suggested that respiratory distress has been reported in the most critical cases of COVID-19 [23]. With the aim of supporting medical decision making, predicted models were developed for detecting people in the general population at risk of being admitted to hospital and for diagnosis of COVID-19 in patients with symptoms, but the results

presented in a recent systematic review describe a poor research performance and a high risk of bias [24]. Using Homogeneity Analysis, we propose a scoring methodology for developing an improved scale. Therefore, we achieved a numerical weight value (optimal quantification) which represents the importance of the binary response categories (YES/NO) for each question of EPICOVID19 DS. As a result, the various binary items of the eleven questions of EPICOVID19 DS contributed with different weights to the overall score. This produced an improved scale, 0 to 10, reflecting the importance for each symptom. Thus, Respiratory problems and Chest pain were the most important symptoms, with score 1.03 and 1.07 respectively. The other symptoms showing an important contribution to the total score were Gastrointestinal complaint (0.95), Loss of Taste and/or Smell (0.97), Tachycardia (palpitations) (1.05). Subsequently, we computed sensitivity (Se) and specificity (Sp) of EPICOVID19 DS compared to COVID-19 positive serological and molecular diagnosis subjects. In NPS(+) subjects the cut-off score was 2.56 with sensitivity equal to 76.56% and specificity equal to 68.24%. In IgG(+) subjects the cut-off score was 2.59 and we obtained a substantial improvement of sensitivity, specificity, PPV and NPV with respect to NPS(+) subjects (Se 80.37; Sp 80.17; PPV 54.43%; NPV 93.27%). When the EPICOVID19 DS scoring algorithm was tested on the older group (>= 60 years), the accuracy detected by Antibodies IgG(+) was improved (Se 88.00; Sp 89.58 AUC 93.10; 95% CI: 86.0-99.5; PPV 81.48%, NPV IgG 93.48%) and the threshold of detection (1.28) was lower than in younger one. Our data are coherent with the findings of the literature, in fact in mid-May 2020, the European all-cause mortality monitoring system showed all-cause mortality above the expected rate in several European countries (Belgium, France, Malta, Spain) and in Italy [25], mainly in the age group of 60 years and above. People over 60, are more vulnerable to SARS-CoV-2 infection and those with pre-existing medical conditions are particularly at risk. Actually, several best practices for older people and their families have been recommended by the WHO, by the Centers for Disease Control and Prevention and by the geriatricians and infectious diseases specialists [26]. The COVID-19 sensitivity and specificity of serological and molecular diagnostic tests is not fully resolved but some studies suggest the sensitivity could be as low as 80% [27-28]. This raises concerns for a high false negative rate, which could result in an increase in the infection spread in the community. There is no absolute answer on the sensitivity and specificity of COVID-19 diagnostic tests because, to determine their accuracy, they must be compared with a "gold standard" test that currently is not identified. Considering estimates of Se and Sp, PPV and NPV can be gained on the basis of the disease prevalence and the rate of illness in the population but, as known, concerning COVID-19 there is significant uncertainty about this prevalence [29]. Statistically it is assumed that PPV varies widely, in a range between 30-50% in areas with low prevalence, as stated in a recent US research about COVID-19 [30]. ~~Early recognition screening and rapid diagnosis are essential to prevent transmission and provide supportive care in a timely manner. Nevertheless, screening is distinguished from further, more detailed diagnostic test assessment. This is of particular relevance as resources for full testing remains as a limited resource and optimizing its use is critical.~~ EPICOVID19 DS could be a preliminary assessment that attempts to detect subjects with symptoms potentially associated with COVID-19 among a wide population. EPICOVID19 DS does not enable a clinical interview to determine the complete symptomatic profile and needs but identifies those who may warrant further assessment. An advantage would be to use this screening in primary care settings, so that GPs can avoid people with suspected COVID-19 in primary care's offices when possible [31], or as a first screening tool and then manage the patient remotely by telephone or video consultations [32]. ~~The EPICOVID19 DS could be applied to the general population. Once the scoring is assigned to each symptom, the EPICOVID19 DS could allow to set different cut-offs according to the subjects involved and to the gold standards used (NPS, serological tests, clinical evaluation by clinicians, etc.).~~ It should be noted that since it is plausible to expect lower prevalence values in the general population than the 22.77% of the present series, the probability of NPV would increase beyond the current 93.27% and consequently the probability of progressing to COVID-19 for subjects who

tested negative (1 - NPV) would be less than the current 6.7%. Although the identified symptoms are not specific for COVID-19, they have been found as valid references in a population setting because they are frequently reported by patients affected by COVID-19. Likely, in a non-pandemic scenario, these symptoms could be assessed with different weights because of their aspecificity, configuring the EPICOVID19 DS as a valid diagnostic support mainly in a pandemic situation. Moreover, the health authority is still unable to monitor through classic tests the spread of SARS-CoV-2 infection, and allowing the circulation of unsuspecting positive subjects could represent a risk for the spread of the infection. The validation of an instrument that can easily identify a suspected case, through a score attributed to each symptom related to COVID-19, can be of great importance in facilitating the containment of the epidemic. The proposed score seems worthy of validation in broader populations in order to confirm its clinimetric properties. In the event of a confirmatory answer, it might qualify as a useful means of selecting people amenable to serological and/or molecular diagnostic tests for COVID-19. ~~The availability and offering of diagnostic tests for the SARS-CoV-2 coronavirus has proven to be one of the keys to the containment of the COVID-19 pandemic. The early identification of positive subjects at molecular and serological tests among people with specific symptoms or considered at risk is crucial to limit the spread of the infection. The tool we validated responds to the need to readily identify a suspect case, through a score attributed to each symptom related to COVID-19. Although the validation was satisfactory, the proposed score seems worthy of further testing in larger populations in order to confirm its clinimetric properties, useful for selecting people susceptible to serological and / or molecular diagnostic tests for COVID-19.~~

~~Although this tool could be a public health prevention measure instrument, directing subjects to a self-assessment without trigger panic, alarmism and concern among the screened population,~~ Some Limitations of our research deserve consideration. First of all, the participation in the study was voluntary and not representative of the general population. This presents potential selection biases that must be taken into consideration. Data were collected in a convenient young-adult and highly educated population sample characterized by low multimorbidity, as resulting from the phase I study [10] and expectable in the case of an on-line questionnaire promoted by e-mail invitations. Further, in the context of a pandemic, our survey might have interested people who had no opportunity to report symptoms to clinicians. Moreover, the effect linked to recall bias cannot be excluded among the participants who tested positive at COVID-19 and/or presenting symptoms related to the SARS-COV2 infection. Another limitation of our study is the resulting small sample size when separating the analysis in the two age groups (<60, >= 60 years). Given these limitations, the adoption of EPICOVID19 DS should be considered with caution and the procedures outlined for its development could be applied iteratively as new data is collected to continue refinement of this potentially valuable clinical decision support tool.

In conclusion, the proposed EPICOVID19 DS seems worthy of further testing in different scenarios and populations in order to achieve a comprehensive knowledge of its clinimetric properties in both low prevalence and high prevalence settings as well as its aptitude to capture disease severity. This will allow define the boundaries of its use and the optimal indications to assist clinicians in the early recognition of COVID-19.

Ethical approval

EPICOVID19 phase II study was approved by the Ethical Committee of the Istituto Nazionale per le Malattie Infettive I.R.C.C.S. ‘L. Spallanzani’ as an amendment of the EPICOVID19 epidemiological study (approval No. 93 of the trial register). Data transfer was safeguarded by means of password protection and encrypting/decrypting policy; all data were handled and stored in accordance with the EU GDPR 2016/679 [33]. Informed consent was accessible on the home page of the platform and

participants were asked to review before starting the compilation, thus explaining the purpose of the study and which data were to be collected and how data were stored.

Email address is the personal data provided on a voluntary basis in Phase I. In our study it was only used: 1) to send email invitation to the Phase II survey; 2) to link the information related to the results (swabs and/or antibodies IgG, IgM) to the information on symptoms collected during the phase I survey. In the participation mail, the person was able to participate by clicking on the provided link to the survey; to not participate ignoring the invitation; to communicate to a specific address, i.e. valid-epicovid19@ifc.cnr.it, the request of deletion of the email address from the database.

Acknowledgments

We would like to thank Prof. Mario Grassi of the University of Pavia for his suggestions in statistical analysis and Dr. Michael N. Liebman for his support in the final revision of the manuscript contents and their formalization in English. The authors would like to thank all the participants who took part in this study and made it possible.

Contributors

SM, LB, FA and FP were responsible for the study concept and design. LB, LF and SP were responsible for the literature search. FA, LF and SP were responsible for acquisition of data. LB, FA, FB, FP and SM were responsible for analysis and interpretation of data. LB, LF, SM, SP and FB were responsible for drafting the manuscript. MG, AG, RAI, CP, MA and GP were responsible for critical revision of the manuscript for important intellectual content. LB, SM were responsible for statistical analysis. GP, SM, CT, MN, NJ, AS, CP, MA critically revised the manuscript for important intellectual content. All authors participated in data interpretation, read, and approved the final version. The corresponding author, SM, attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Declaration of interests

We declare no competing interests.

Reference

1 WHO. Coronavirus (COVID-19) events as they happen. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed July, 2020).

2 Johns Hopkins University and Medicine. COVID-19 map. Johns Hopkins Coronavirus Resource Centre. https://coronavirus.jhu.edu/map.html (accessed July, 2020)

3 Giacomelli A, Ridolfo AL, Milazzo L, et al. 30-day mortality in patients hospitalized with COVID-19 during the first wave of the Italian epidemic: A prospective cohort study. *Pharmacol Res*. 2020; **158**:104931. PMID: 32446978

4 COVID-19 Clinical Research Coalition. Electronic address: nick.white@covid19crc.org. Global coalition to accelerate COVID-19 clinical research in resource-limited settings. *Lancet*. 2020; 395(10233):1322-1325. PMID: 32247324

5 Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19 - studies needed. *N Engl J Med.* 2020; **382**: 1194–96. PMID: 32074416

6 WHO. Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection. March 17, 2020. https://apps.who.int/iris/handle/10665/331656 (accessed June, 2020).

7 Green et al. Molecular and antibody point-of-care tests to support the screening, diagnosis and monitoring of COVID-19, April 2020, Oxford University.

8 WHO. Global surveillance for COVID-19 caused by human infection with COVID-19 virus. Interim guidance 20 March 2020.

9 Reeves et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. *J Am Med Inform Assoc*. 2020 Jun 1; **27**(6):853-859. PMID: 32208481

10 Adorni F, Prinelli F, et al. Self-reported symptoms of SARS-CoV-2 infection in a non-hospitalized population: data from the large Italian web-based EPICOVID19 cross-sectional survey. 2020 Jul 10. PMID: 32650305

11 Long Q, Liu B, Deng H, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med*. 2020 Jun; **26**(6):845-848. PMID: 32350462

12 Chan JFW, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020; **395**: 514–23. PMID: 31986261

13 Giacomelli A, Pezzati L, Conti F, et al. Self-reported olfactory and taste disorders in SARS-CoV-2 patients: a cross-sectional study. Clin Infect Dis. 2020;ciaa330. PMID: 32215618

14 Pallanti Importance of SARs-Cov-2 anosmia: From phenomenology to neurobiology. May 2020; Compr Psychiatry. 2020 Jul; **100**:152184. PMID: 32422426

15 Menni et al. Quantifying additional COVID-19 symptoms will save lives. *Lancet* 2020 Jun 20; **395**(10241):e107-e108. PMID: 32505221

16 Cholankeril et al. High Prevalence of Concurrent Gastrointestinal Manifestations in Patients with SARS-CoV-2: Early Experience from California. *Gastroenterology* 2020 Apr 10; **S0016-5085**(20)30471-6. PMID: 32283101

17 Sarma P, Kaur H, Kaur H, et al. Ocular Manifestations and Tear or Conjunctival Swab PCR Positivity for 2019-nCoV in Patients with COVID-19: A Systematic Review and Meta-Analysis (3/30/2020). SSRN Electronic Journal. 2020 Apr 15. DOI:10.2139/ssrn.3566161

18 Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020 Feb 15; **395**(10223):497-506. PMID: 31986264

19 Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020 Apr 30; **382**(18):1708-20. PMID: 32109013

20 Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province. *J Med Internet Res*. 2020;22(6):e19636. Published 2020 Jun 30. PMID: 32544071

21 Bénézit F, Le Turnier P, Declerck C, et al. Utility of hyposmia and hypogeusia for the diagnosis of COVID-19. *Lancet Infect Dis.* 2020 Apr 15. PMID: 32304632

22 Sun P, Qie S, Liu Z, Ren J, Li K, Xi J. Clinical characteristics of hospitalized patients with SARS-CoV-2 infection: A single arm meta-analysis. *J Med Virol*. 2020 Feb 28. **92**(6):612-617. PMID: 32108351

23 Jain V, Yuan J-M. Systematic review and meta-analysis of predictive symptoms and comorbidities for severe COVID-19 infection. MedRxiv 2020; published online March 16. DOI: https://doi.org/10.1101/2020.03.15.20035360

24 Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020 Apr 7; **369**:m1328 PMID: 32265220

25 Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA. 2020 May 12;**323**(18):1775-1776. JAMA. 2020 Apr 28;323(16):1619. PMID: 32203977.

26 World Health Organization. (2020). Preventing and managing COVID-19 across long-term care services: policy brief, 24 July 2020. World Health Organization. https://apps.who.int/iris/handle/10665/333074. License: CC BY-NC-SA 3.0 IGO

27 Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* 2020; **323**(18):1843-1844 PMID: 32159775

28 Yong SEF, Anderson DE, Wei WE, et al. Connecting clusters of COVID-19: an epidemiological and serological investigation. *Lancet Infect Dis*. 2020 Apr 21. **20**(7):809-815. PMID: 32330439

29 Charles F. Manski Bounding the Predictive Values of COVID-19 Antibody Tests. MedRxiv 2020; published online May 18. DOI: https://doi.org/10.1101/2020.05.14.20102061

30 Mathur G, Mathur S. Antibody testing for COVID-19: Can it be used as a screening tool in areas with low prevalence? Am J Clin Pathol 2020.;**154**:1-3.

31 Razai MS, Doerholt K, Ladhani S, et al. Coronavirus disease 2019 (covid-19): a guide for UK GPs. *BMJ* 2020 Mar 5; **368**:m800.

32 Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. *BMJ* 2020 Mar 25; **368**:m1182. PMID: 32213507

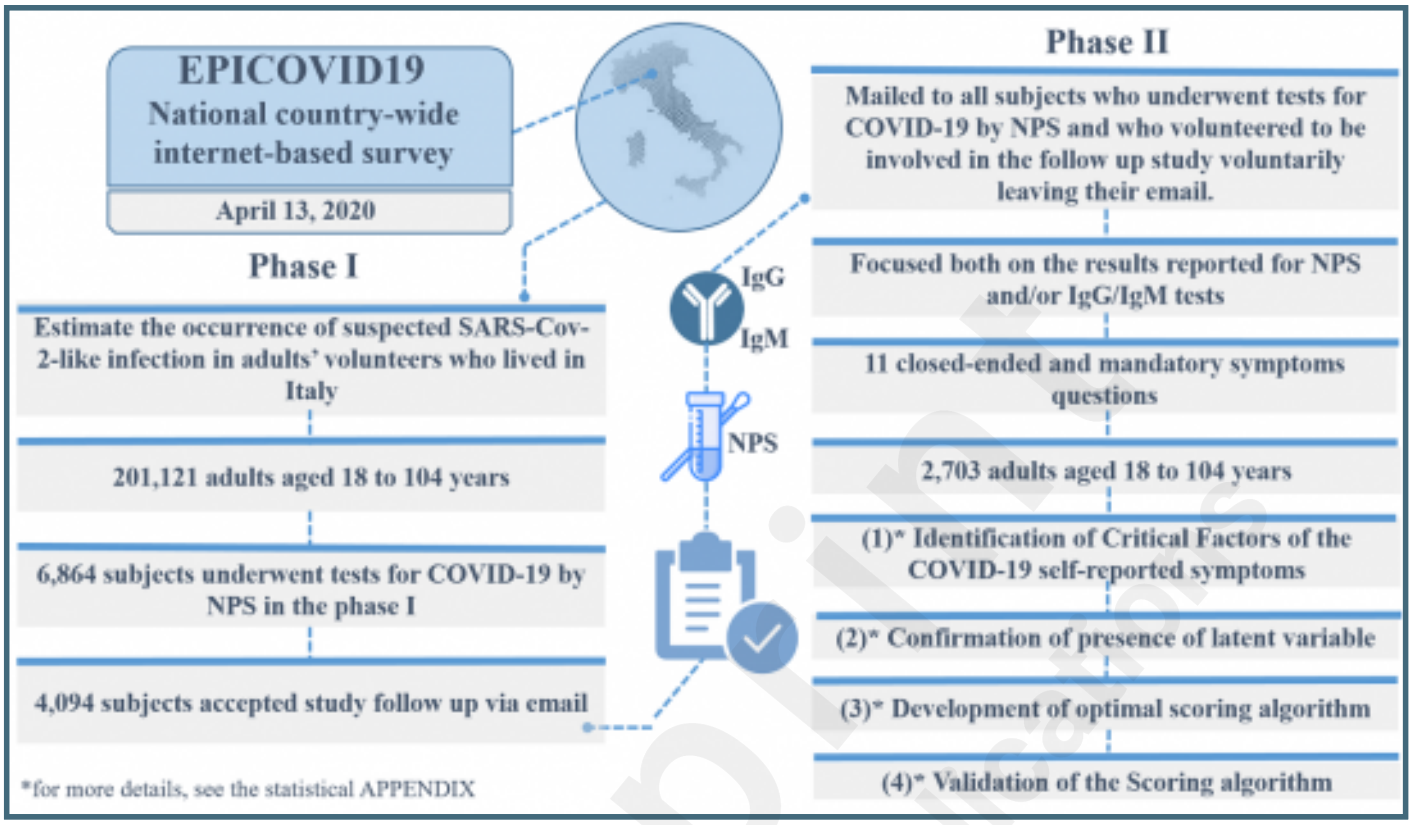
33 Official documents of the EU Regulation 2016/679 General Data Protection Regulation. https://gdpr-info.eu/. (accessed September, 2020).

Supplementary Files

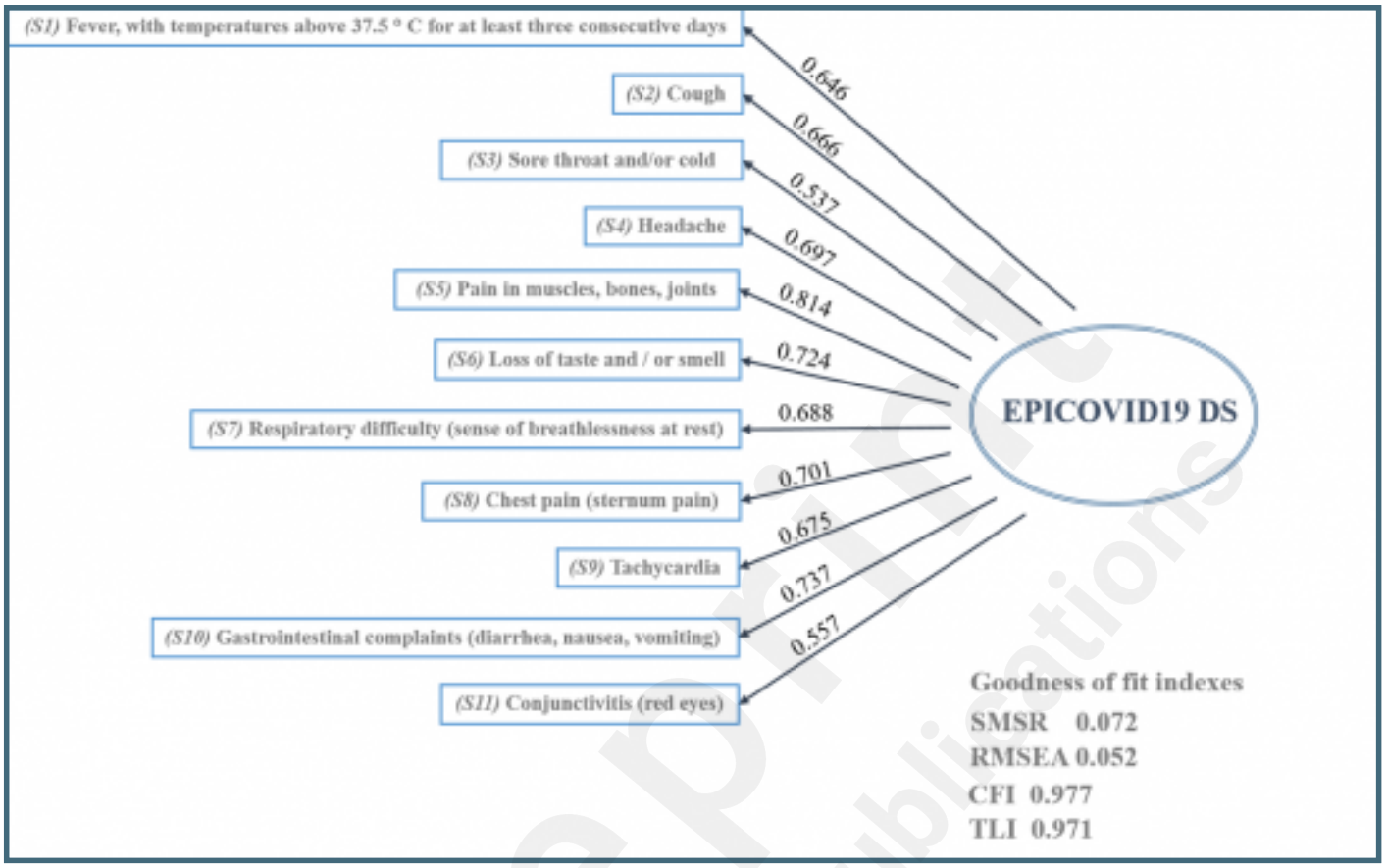
Untitled.
URL: https://asset.jmir.pub/assets/58b40e785332f6edeffee9ee15639a07.docx

Figures

Overview of EPICOVID19 two-phase study.



Standardized factor loadings and goodness of fit indexes for one factor model. The goodness of fit indexes (SMSR, RMSEA) of the "EPICOVID19 DS".



Multimedia Appendixes

Supplementary appendix. Phase I Epicovid-19 questionnaire and Phase 2 Valid Symptoms Section of Epicovid-19 questionnaire and Statistical appendix.
URL: https://asset.jmir.pub/assets/c860c52a720a7d62706f0b89dafe8593.docx