

# Supplementary appendix

This appendix formed part of the original submission

## Phase I Epicovid-19 questionnaire and Phase 2 ValidSymptoms Section of Epicovid-19 questionnaire

### COVID-19 NATIONAL EPIDEMIOLOGICAL SURVEY (EPICOVID19)

#### Phase I questionnaire

##### SOCIO-DEMOGRAPHIC DATA

*\*mandatory questions*

1. Sex at birth\* M  F

2. Year of birth \*

3. Ethnicity\*

- Europe
- Asia
- North America
- Central/South America
- Africa
- Oceania
- Other

4. Postal code\*

5. Municipality\*

6. Education\*

- None
- Primary school
- Middle high school
- High school
- University degree
- Post graduate degree (PhD, vocational master, medical specialization)

7. Occupational status\*

- Employed
- Student
- Unemployed

Retired

Other

**8. Last occupation\***

Armed forces occupations

Managers

Intellectual and scientific professionals (e.g. clinicians, engineers, researchers, teachers)

Technicians and associate professionals (e.g. healthcare technicians, IT technicians)

Clerk

Service and sales workers (e.g. merchants, sales persons)

Craft and related trades workers

Plant and machine operators, assemblers, and drivers of vehicles

Other

## CLINICAL EVALUATION

### 9. Did you have one or more of the following symptoms since the 1st of February 2020? \*

- Fever with a temperature greater than 37.5 °C for at least three consecutive days
- Cough
- Sore throat/rhinorrea
- Headache
- Myalgia
- Olfactory or taste disorders
- Shortness of breath
- Chest pain
- Feelings of having a fast-beating
- Gastrointestinal disorders (diarrhoea, nausea, vomiting)
- Conjunctivitis
- Pneumoniae

### 10. If you had at least one of the symptoms above, please indicate in which month they occurred for the first time \*

- February
- March
- April

### 11. Have you ever been diagnosed with one or more of the following conditions?

- Lung diseases (e.g. asthma, obstructive pulmonary disease)
- Heart diseases (e.g. ischemic heart disease, atrial fibrillation)
- Hypertension
- Kidney diseases
- Immune system diseases (e.g. thyroid disease, psoriasis, rheumatoid arthritis)
- Tumours
- Metabolic diseases (e.g. diabetes, obesity, gout)
- Liver diseases (e.g. hepatitis, cirrhosis, liver failure)
- Depression and/or anxiety

### 12. Please indicate other conditions

- Surgical procedures under general anaesthesia during the last year
- Transplants
- Allergies
- Pregnancy
- Non self-sufficient in carrying out daily activities
- Healthcare workers (e.g. clinicians, nurses, rescuer, pharmacist)

### 13. Did you carry out the following vaccinations?\*

- Flu shot during the last autumn  Yes  No
- Anti-pneumococcal in the last 12 months  Yes  No
- Other vaccinations in the last 12 months  Yes  No

**14. Do you regularly take one or more of the following medicines?**

- Aspirin
- Anti-hypertensive
- Hypocholesterolemic drugs
- Anti-diabetics
- Anti-cancer drugs
- Corticosteroids
- Thyroid drugs (e.g. euthyrox)
- Anti-inflammatory drugs
- Anxiety medications and/or sedatives
- Anti-depressant
- Supplements (e.g. vitamins)

*If females*

**15. Have you ever been taking birth pills and/or hormone replacement therapy?\***

- No
- Yes, in the past, for less than 5 years
- Yes, in the past, for more than 5 years
- Yes currently, taking it less than 5 years
- Yes currently, taking it more than 5 years

**16. Indicate the number of completed pregnancies\***

- 0
- 1
- 2
- 3 or more

**17. Have you been in a close contact (direct contact at a distance of less than 2 meters, or in a closed environment such as a house, workplace, transportation vehicles) with confirmed COVID-19 cases, live or deceased?\***

- Yes  No

**18. Have you been in a close contact (direct contact at a distance of less than 2 meters, or in a closed environment such as a house, workplace, transportation vehicles) with suspected COVID-19 cases, live or deceased?\***

- Yes  No  I don't know

**19. Did you contact the emergency number and/or the general practitioner to report any symptoms of suspected infection by COVID-19?\***

- No
- No but I went to the hospital on my own initiative
- Yes, and they suggested to me isolation
- Yes, and they did not suggest to me isolation
- Yes, and I was sent to the hospital

**20. Have you been tested for COVID-19? \***

- Yes, with a positive result
- Yes, with a negative result
- Yes, but I do not know the result
- No, I did not perform any test

**21. Have you been hospitalized due to COVID-19, either as a suspected or as a confirmed case?\***

Yes  No

**22. Please indicate other elements that might be of relevance for COVID-19**

---

---

### PERSONAL CHARACTERISTICS AND HEALTH STATUS

**23. How would you describe your health in general?\***

Very bad  Bad  Adequate  Good  Very good

**24. Do you fear getting infected with the coronavirus (COVID-19)?\***

No  
 Just a little bit  
 Neutral  
 Quite enough  
 Yes, a lot

**25. Do you fear your family being infected with the coronavirus (COVID-19)?\***

No  
 Just a little bit  
 Neutral  
 Quite enough  
 Yes, a lot

### HOUSING CONDITIONS

**26. Your home is located in:\***

City centre with more than 100.000 inhabitants  
 Suburbs of cities with more than 100.000 inhabitants  
 Small town  
 Countryside

**27. Your home is located in an area where the road traffic is:\***

Intense (living near a busy road)  
 Moderate  
 Low

**28. How many rooms there are in your home (excluding the bathroom and auxiliary spaces)?\***

One  Two  Three  More than three

**29. Besides you, how many people live in your household?\***

None  One  Two  More than two

**30. Are there in the same household elderly persons or anyone with immunocompromising or chronic disease conditions?\***

Yes  No

## LIFESTYLE

**31. How many people have you been in contact with in average, prior to the Governmental restrictions on lockdown?\***

Less than 10  Between 10 and 100  More than 100

**32. Do you smoke?\***

I have never smoked or I smoked less than 100 cigarettes in my lifetime

I am a former smoker (I have smoked at least 100 cigarettes in my lifetime and I do not smoke anymore)

Yes, smoking less than 10 cigarettes per day

Yes, smoking between 10 and 20 cigarettes per day

Yes, smoking more than 20 cigarettes per day (more than one pack per day)

*If former smoker*

**33. How many year have you smoked?\***

*If current smoker*

**34. How many years?\***

**35. Prior to the lockdown by the Governmental restrictions, for how long did you follow a regular routine of moderate or intense physical activity (swimming, racing)?\***

Not doing any physical activity or doing it for less than 10 minutes per week

Between 10 minutes and two hours and half per week

More than two hours and half per week

## BEHAVIOURS FOLLOWING THE LOCKDOWN

**36. Please indicate your employment status following the restrictions introduced by the Italian Government since March, 9th 2020 \***

Continued going to work at my workplace

Working from home

I had to stop working due to the emergency

I am not employed

**37. Since March, 9th 2020, how many times do you go out during a week? \***

Never

1

2-3

4-5

6 or more

**38. Since March, 9th 2020, do you use public transport go to work or to provide supplies\***

No, never

Yes, 1-3 times per week

Yes, 4-6 times per week

Yes, 7 or more times per week

**Any additional information and comments**

---

---

---

**Phase II questionnaire**

**Laboratory test results**

Please answer the questions below regarding the results of the laboratory tests you have carried out

**A1. Have you made one or more swabs with a POSITIVE result? \***

Yes  No

**A2. Indicate the date of the FIRST POSITIVE Swab**

\_\_\_\_/\_\_\_\_/\_\_\_\_

**A3. With reference to your FIRST POSITIVE SWAB, enter the viral load, if available**

\_\_\_\_\_

**A4. Indicate the number of elapsed days from the beginning of the symptoms to the day of the FIRST POSITIVE SWAB**

\_\_\_\_\_

**A5. Have you made one or more swabs with a NEGATIVE result? \***

Yes  No

**A6. Indicate the date of the LAST NEGATIVE SWAB**

\_\_\_\_/\_\_\_\_/\_\_\_\_

**A7. Have you performed the serological test too? \***

Yes  No

**A8. Indicate the outcome of the serological test for Immunoglobulin G (IgG) Antibodies:**

Positive

Negative

**A9. Indicate the outcome of the serological test for Immunoglobulin M (IgM) Antibodies:**

Positive

Negative

**A10. Indicate the number of elapsed days from the beginning of the symptoms to the day in which the serological test was performed:**

\_\_\_\_\_

**A11. Any additional information and comments**

---

---

---

**Thank you for your cooperation!**

## APPENDIX

### Respondents and non-respondents to the Phase I and Phase II survey

	Respondents Phase I and Phase II	Respondents only to Phase I	p value
	(n° 2703)	(n° 4161)	
<b>Age</b>	49.6, ±14.7	46.2, ±13.1	0.014
<b>Positive Nasopharyngeal swab</b>	706 (26.1%)	862 (20.7%)	<0.0001
<b>Female</b>	1841 (68.14%)	2669 (64.1%)	0.001
<b>Ethnicity</b>	2666 (98.60%)	4133 (99.3%)	0.005
<b>Lower educational status</b>	151 (5.6%)	17 (0.4%)	<0.0001
<b>Medium educational status</b>	837 (31.0%)	1019 (24.5%)	
<b>High educational status</b>	1715 (63.4%)	3125 (75.1%)	
<b>Self-reported comorbidities</b>			
<b>Lung diseases (e.g. asthma, obstructive pulmonary disease)</b>	196 (7.3%)	297 (7.1%)	0.859
<b>Heart diseases (e.g. ischemic heart disease, atrial fibrillation)</b>	103 (3.8%)	120 (2.9%)	0.034
<b>Hypertension</b>	361 (13.4%)	567 (13.6%)	0.748
<b>Kidney diseases</b>	22 (0.8%)	33 (0.8%)	0.925

<b>Immune system diseases (e.g. thyroid disease, psoriasis, rheumatoid arthritis)</b>	266 (9.8%)	365 (8.8%)	0.134
<b>Tumours</b>	88 (3.3%)	121 (2.9%)	0.413
<b>Metabolic diseases (e.g. diabetes, obesity, gout)</b>	124 (4.6%)	161 (3.9%)	0.145
<b>Liver diseases (e.g. hepatitis, cirrhosis, liver failure)</b>	21 (0.8%)	29 (0.7%)	0.703
<b>Depression and/or anxiety</b>	194 (7.2%)	265 (6.4%)	0.190

## APPENDIX

### Supplementary appendix

This appendix formed part of the original submission

#### (1) Identification of Critical Factors by Exploratory and confirmatory factor analysis

The factor structure of the COVID-19 self-reported symptoms items was tested using two procedures: Exploratory factor analysis (EFA)<sup>1</sup> and confirmatory factor analysis (CFA)<sup>2</sup>. Exploratory factor analysis (EFA) is a variable reduction technique which identifies the number of latent constructs and the underlying factor structure of a set of variables. Confirmatory factor analysis (CFA) is a statistical technique used to verify the factor structure of a set of observed variables. CFA allows testing the hypothesis that a relationship between the observed variables and their underlying latent constructs exists. In the first step, EFA and PA (parallel analysis) were performed to identify item performance (loadings) and to define the number of factors underlying the scale. EFA was applied using tetrachoric correlation matrix and two solutions were tested with one and two factors. PA was performed (1000 random correlation matrices) to establish the number of factors underlying the scale. Classical criteria from EFA were used: eigenvalue rule according Kaiser's rule of thumb<sup>3</sup>, only a factor with eigenvalue > 1 explained more of the variance, scree plot (n° of factors before the break in the eigenvalue plot), Horn's parallel analysis (number of factors with eigenvalue higher than average random eigenvalues), and factor-loading rule (item-factor correlations >0.35).

#### (2) Confirmation of presence of latent variable by Confirmatory Factor Analysis (CFA)

A Confirmatory Factor Analysis (CFA) via Structural Equation Modeling (SEM) was performed to confirm the presence of a one latent variable (factors) underlying the COVID-19 self-reported symptoms items. The items (endogenous variables) are (S1) *Fever, with temperatures above 37.5 ° C for at least three consecutive days*, (S2) *Cough*, (S3) *Sore throat and/or cold*, (S4) *Headache*, (S5) *Pain in muscles, bones, joints*, (S6) *Loss of taste and / or smell*, (S7) *Respiratory difficulty (sense of breathlessness at rest)*, (S8) *Chest pain (sternum*

pain), (S9) Tachycardia, (S10) Gastrointestinal complaints (diarrhea, nausea, vomiting), (S11) Conjunctivitis (red eyes). Several goodness-of-fit criteria was use<sup>4</sup>: (i) Standardized Root Mean Square Residual (SRMR); (ii) Root Mean Square Error of Approximation (RMSEA). In general, a model is considered to show good fit when the SRMR and RMSEA are not higher than 0.10. In addition, further CFA indexes as gamma for unidimensionality (AVE; average variance extracted acceptable if > 0.4) and the general validity coefficient (McDonald's Omega acceptable if > 0.8) verify the adequacy of single-factor measurement model and the factor validity index (rho) for the evaluation of the correlation factor with the variables<sup>5</sup>. Finally, the standardized factor loadings (item-factor correlations) (acceptable if > 0.4) proved high-quality specific factor validity for each item and determinacy coefficient (DC) >0.90 as construct validity index.

### (3) Development of optimal scoring algorithm by Homogeneity Analysis

After evaluating the unidimensionality through exploratory and confirmatory factor analyses

an optimal questionnaire score by Homogeneity Analysis (Multiple Correspondence Analysis)<sup>6</sup> was constructed. Through the HOMALS procedure we replaced item dichotomous responses (Yes/No) with category quantifications: the score is the sum of the subject's item responses, once they are recoded by category quantifications. Category quantifications assigned by HOMALS to the dichotomous items are named "optimal" quantifications because they maximize homogeneity among items and minimize measurement error. Subsequently, the COVID-19 self-reported optimal score was rescaled to a range of 0 -10 using:  $10 * (\text{EPICOVID 19 score} - \text{min}) / (\text{max} - \text{min})$ . The result of this equation was used as the EPICOVID 19 self-reported optimal score for each individual, a continuous trait that summarizes the common components of COVID-19 self-reported symptoms items. With more details the subject score is:

$$\text{EPICOVID 19 optimal score} = \begin{cases} 0 & \text{if "NO"} \\ \frac{q_{\text{YES}_j} - q_{\text{NO}_j}}{\sum_{j=1}^p (q_{\text{YES}_j} - q_{\text{NO}_j})} \cdot 10 & \text{if "YES"} \end{cases}$$

the figures  $q^{\text{YES}_j}$  and  $q^{\text{NO}_j}$  are the HOMALS optimal quantifications of the "Yes" and "No" binary item responses, respectively <sup>6</sup>. Additionally, Cronbach Index ( $\alpha$ ) which is a scale reliability coefficient based on the internal consistency, and Greenacre Index ( $\tau$ ) which is the percentage of inter-item association explained by the first dimension, were also computed

### (4) Validation of the Scoring algorithm by receiver operating characteristics (ROC)

Hence, the score validation was performed by receiver operating characteristics (ROC) analysis was performed to evaluate the COVID-19 self-reported symptoms items effectiveness in distinguishing symptomatic individuals severity. With the aim of discriminating between different severity of symptoms, for each EPICOVID 19 factor score built, sensitivity, specificity, and youden's index were computed with two

reference standard<sup>7</sup>: a) subjects with NPS + vs subjects with NPS - ; b) subjects with serological IgG(+) vs subjects with IgG(-). Overall, predictive performance was measured by the AUC (Area Under the Curve): ROC curves with AUC of 0.5 indicating no predictive diagnosis versus AUC of 1.0 indicating perfect ones. The Youden index ( $J$ ), a statistic that captures the performance of a diagnostic test, computed as  $J = \text{sensitivity} + \text{specificity} - 1$ , was used to identify the optimal cut point<sup>8</sup>. All statistical analyses were done using R software (version 3.6.3)

1. Gorsuch RL. Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research* 1990, **25**, 33-39.
2. Gifi A. Nonlinear multivariate analysis. New York: John Wiley & Sons 1990 (First edition 1981, Department of Data Theory, University of Leiden).
3. Bartholomew DJ, Knott M. Latent variable models and factor analysis. 1999, 2nd ed. New York: Oxford University Press.
4. Hooper D, Coughlan J, Mullen M. Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods* 2008; **6**(1), 53-60.
5. McDonald RP. Theoretical foundations of principal factor analysis and alpha factor analysis. *Br J Math Stat Psychol* 1970; **23**:1-21.
6. Greenacre M. Correspondence analysis in the social sciences. London: Academic Press; 1994.
7. Choi BC. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. *Am J Epidemiol* 1998; **148**: 1127–32
8. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* 2005;**16**:73e81

## APPENDIX

### One factor solution extracted by Principal-Component Factors

Symptoms items	One-factor solution
(S1) Fever, with temperatures above 37.5 ° C for at least three consecutive days	0.653
(S2) Cough	0.702
(S3) Sore throat and/or cold	0.571
(S4) Headache	0.721
(S5) Pain in muscles, bones, joints	0.815
(S6) Loss of taste and / or smell	0.732
(S7) Respiratory difficulty (sense of breathlessness at rest)	0.723
(S8) Chest pain (sternum pain)	0.736
(S9) Tachycardia (Palpitation)	0.707

(S10) Gastrointestinal complaints (diarrhea, nausea, vomiting)	0.762
(S11) Conjunctivitis (red eyes)	0.606

## APPENDIX

### Standardized coefficients and goodness of fit indexes for one factor model.

	Standardized coefficients	Std.Err	z-value	P(> z )
(S1) Fever, with temperatures above 37.5 ° C for at least three consecutive days	0.646	0.020	31.657	<0.0001
(S2) Cough	0.666	0.020	33.816	<0.0001
(S3) Sore throat and/or cold	0.537	0.023	23.752	<0.0001
(S4) Headache	0.697	0.019	37.296	<0.0001
(S5) Pain in muscles, bones, joints	0.814	0.016	51.722	<0.0001
(S6) Loss of taste and / or smell	0.724	0.019	38.974	<0.0001
(S7) Respiratory difficulty (sense of breathlessness at rest),	0.688	0.021	32.710	<0.0001
(S8) Chest pain (sternum pain)	0.701	0.021	33.122	<0.0001

(S9) Tachycardia	0.675	0.023	29.201	<0.0001
(S10) Gastrointestinal complaints (diarrhea, nausea, vomiting)	0.737	0.018	42.009	<0.0001
(S11) Conjunctivitis (red eyes)	0.557	0.027	20.373	<0.0001

AVE; average variance extracted 0.516

McDonald's Omega 0.904

Factor validity index (rho) 0.951

SRMR-Standardized Root Mean Square Residual 0.072

RMSEA-Root Mean Square Error of Approximation 0.052

CFI-Comparative fit index 0.977

TLI-Tucker-Lewis index 0.971