

SUPPLEMENTARY MATERIALS

Symptom based early stage differentiation between SARS-CoV-2 versus other respiratory tract infections - Upper Silesia pilot study

Justyna Mika¹, Joanna Tobiasz¹, Joanna Zyla¹, Anna Papiez¹, Małgorzata Bach², Aleksandra Werner², Michał Kozielski³, Mateusz Kania⁴, Aleksandra Gruca³, Damian Piotrowski⁵, Barbara Sobala-Szczygieł⁵, Bożena Włostowska⁵, Paweł Foszner⁴, Marek Sikora³, Joanna Polanska^{1*}, Jerzy Jaroszewicz⁵

¹Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland,

²Department of Applied Informatics, Silesian University of Technology, Gliwice, Poland,

³Department of Computer Networks and Systems, Silesian University of Technology, Gliwice, Poland,

⁴Department of Graphics, Computer Vision and Digital Systems, Silesian University of Technology, Gliwice, Poland,

⁵Department of Infectious Diseases and Hepatology, Medical University of Silesia, Katowice, Poland

*corresponding author: joanna.polanska@polsl.pl

Supplementary figure legends

Fig. S1. The distribution of missing values across sample.

Fig. S2. Frequency of patients observing one symptom or a combination of symptoms regarding the total number of patients from a certain group (infected or not infected with SARS-CoV-2) who answered all eight analysed queries. Dots in the lower part indicate which symptom or a list of symptoms is shown by the bars in the upper part of the plot. For example: almost 4% of patients with negative COVID-19 diagnosis reported only high temperature (and not cough, muscle aches, headache, sore throat, loss of taste/smell, dyspnea or dizziness), whereas from the group of patients with positive diagnosis, about 1% of patients reported only high temperature. Significant odds ratio were coloured green. Combinations of symptoms with frequencies < 0.01 were not shown here.

Fig. S3. Heatmaps with correlation coefficients for pairwise comparison of queries. Blood type was ignored due to different interpretation of effect size coefficient. Comparisons with at least small effect size have been coloured, i.e. values of correlation within the range $(-0.1, 0.1)$ were not coloured. Panel a shows heatmaps for men and women. Panel b shows heatmaps for two age subsets (younger and older than 46 years).

Fig. S4. Frequency of patients observing one symptom or a combination of symptoms regarding the total number of patients from a certain group (infected or not infected with SARS-CoV-2) who answered all four analysed queries, considered men and women separately. Dots in the middle part indicate which symptom or combination of symptoms the bars in the upper part of the plot refer to. The lower part of the plot shows odds ratio with 95% confidence intervals for comparison of SARS-CoV-2 diagnosis with a corresponding symptom or symptoms combination. The red line shows the level of odds ratio equal to 1. Significant odds ratio were coloured green. The odds ratio is significant if the confidence interval does not include the value of 1 (red line).

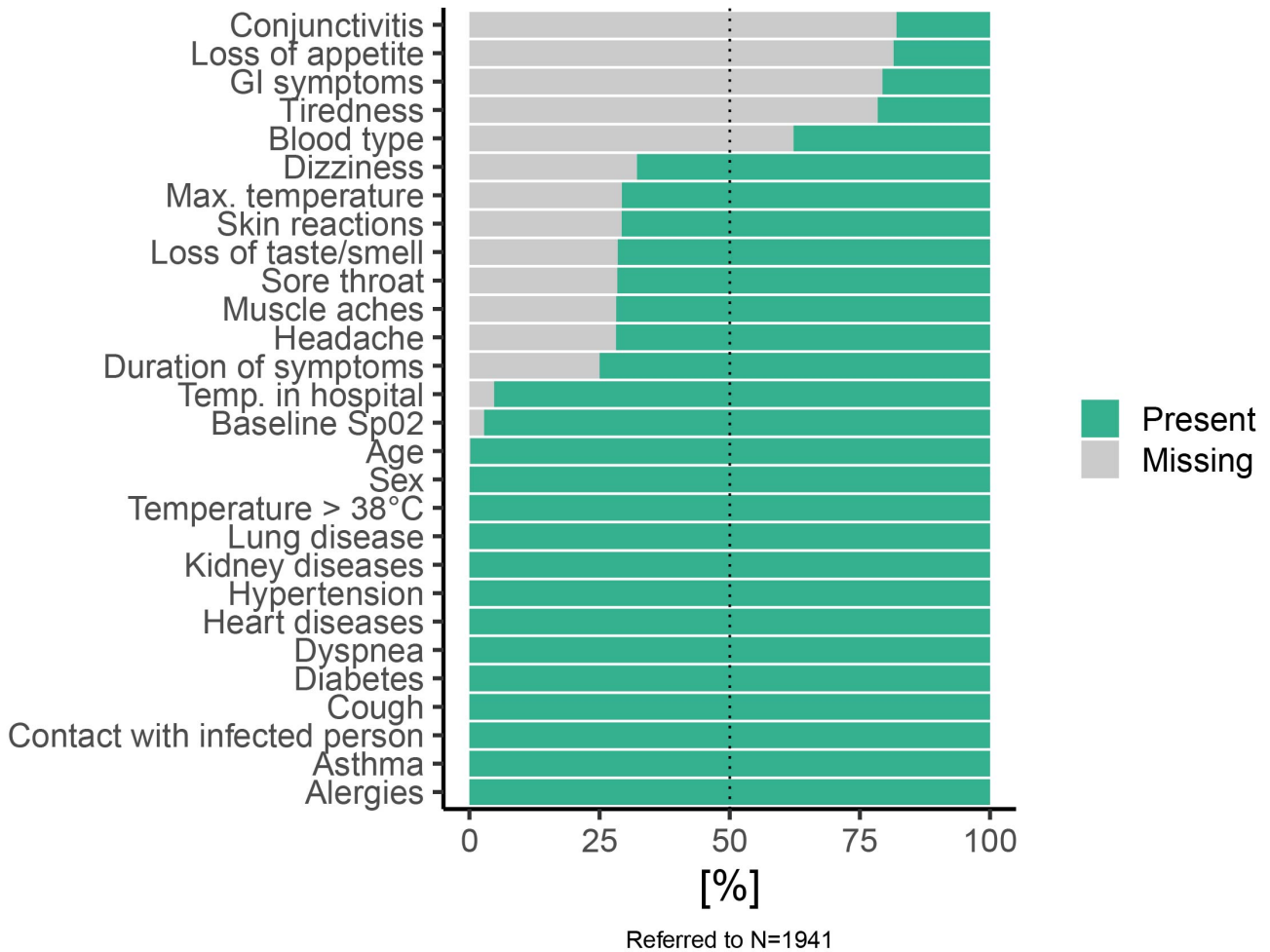
Fig. S5. Venn diagrams showing co-occurrence of four main symptoms among people infected with SARS-CoV-2. Colour scale shows the frequency of symptoms ranging from small frequency (white) to high frequency (red). Panel a – Venn diagrams for men and women separately. Panel b – Venn diagrams for younger and older subsets of patients.

Supplementary tables:

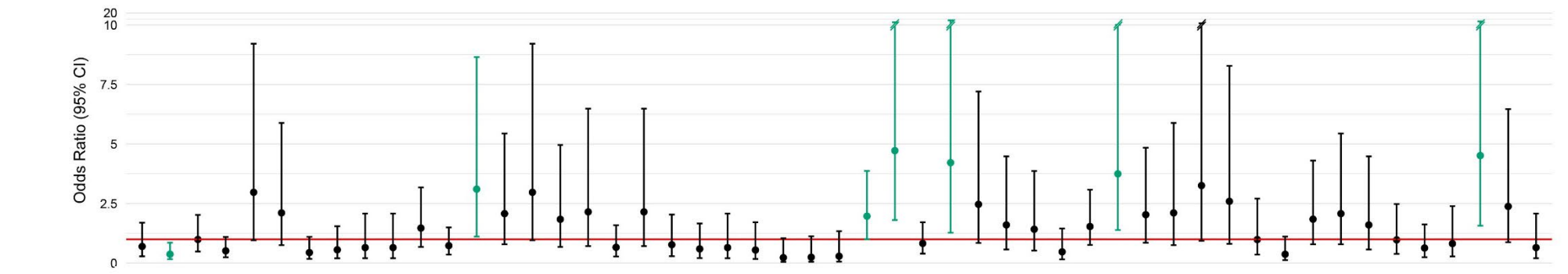
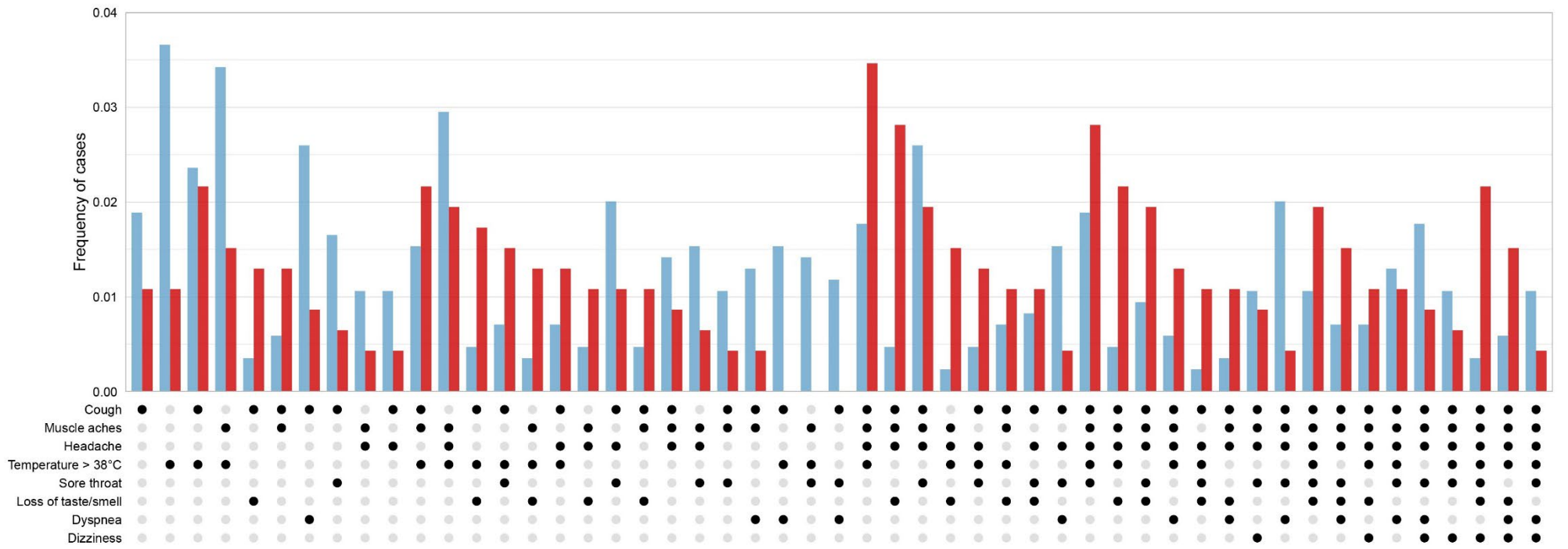
Table S1. The detailed characterisation of the sample.

Table S2. The quality indices for logistic-regression-based predictors of the early-stage SARS-CoV-2 infection for the chosen subsets of the nine main symptoms (10-fold cross-validation).

Table S3. The sex- and age-dependent prevalence of studied symptoms among patients positive and negative for SARS-CoV-2.



SARS-CoV-2(-) N=847 SARS-CoV-2(+) N=462



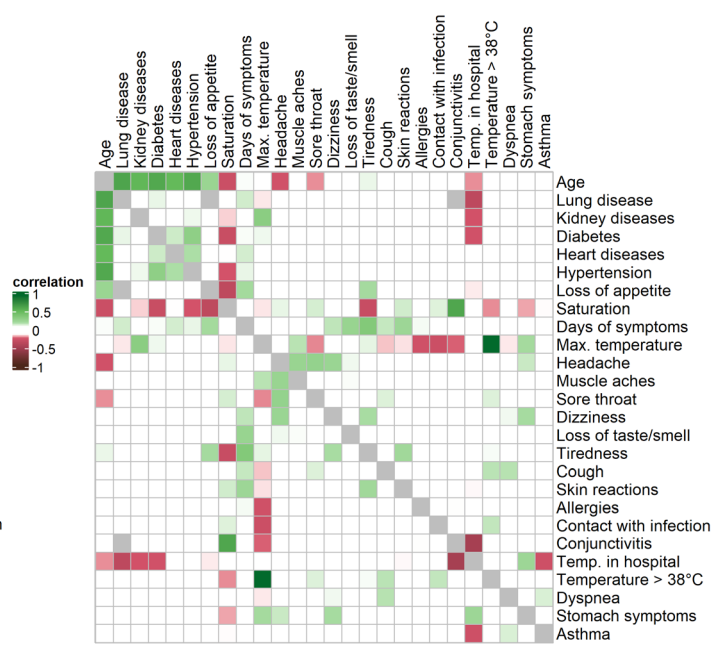
N is the number of patients who answered the questions.

a

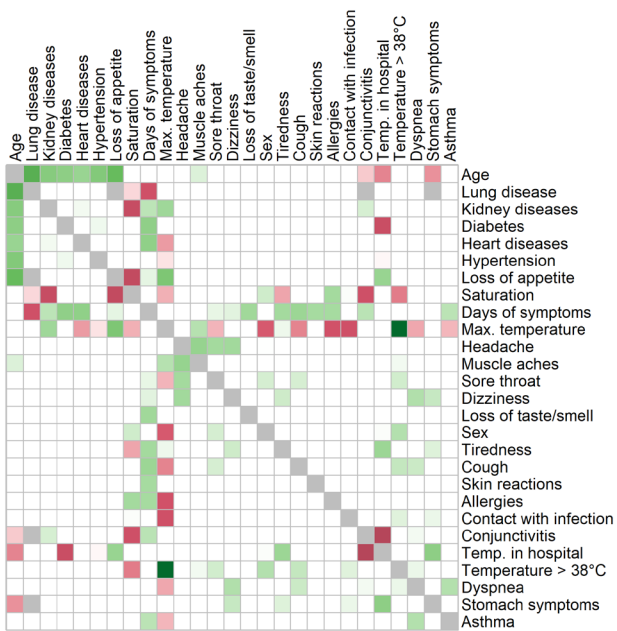
MEN (N=279)



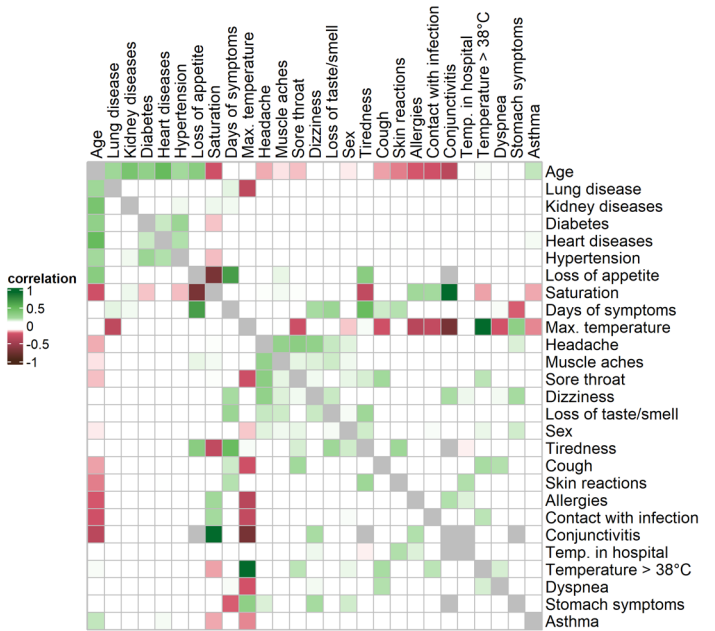
WOMEN (N=205)

**b**

AGE ≤ 46 years (N=305)

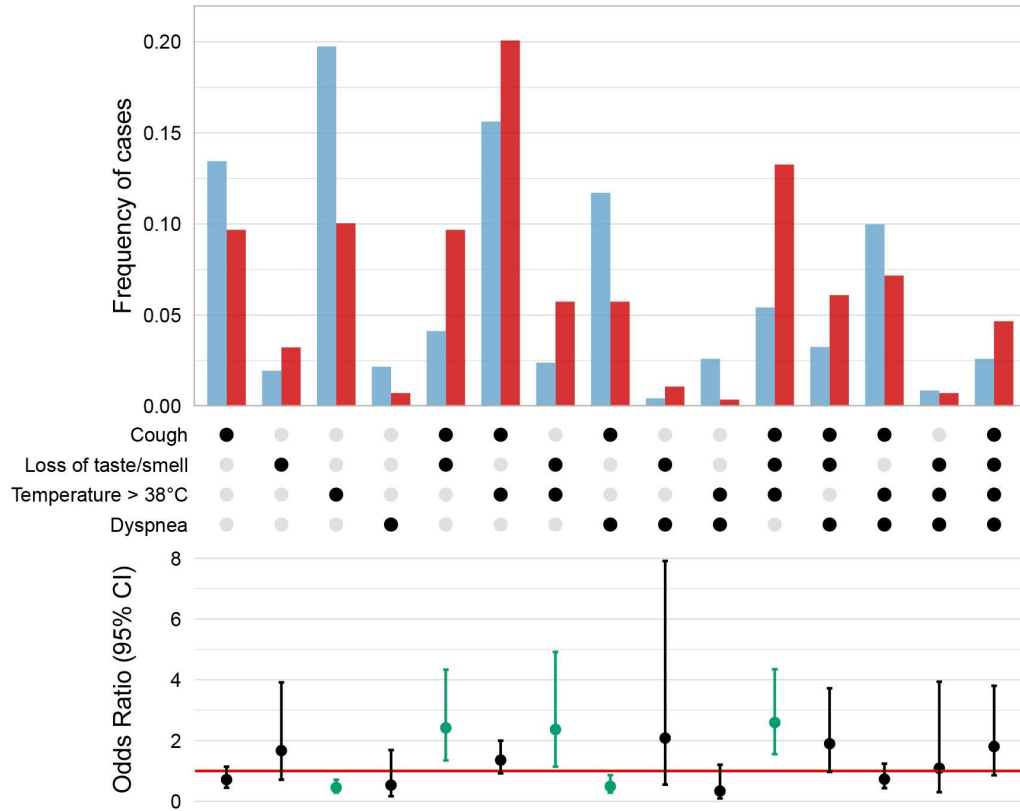


AGE > 46 years (N=178)



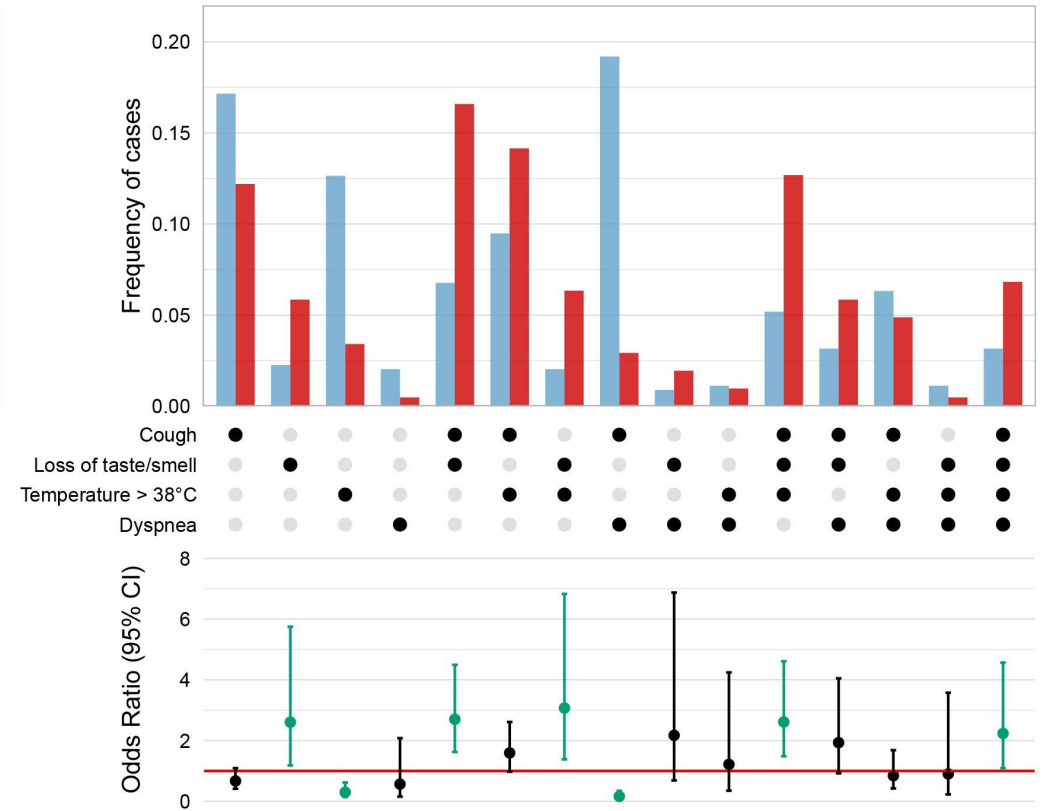
MEN

SARS-CoV-2(-) N=461 SARS-CoV-2(+) N=279

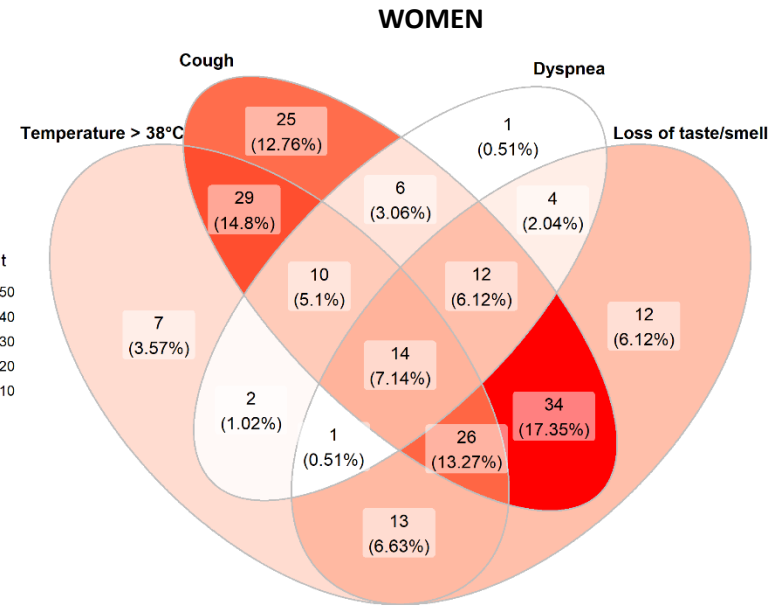
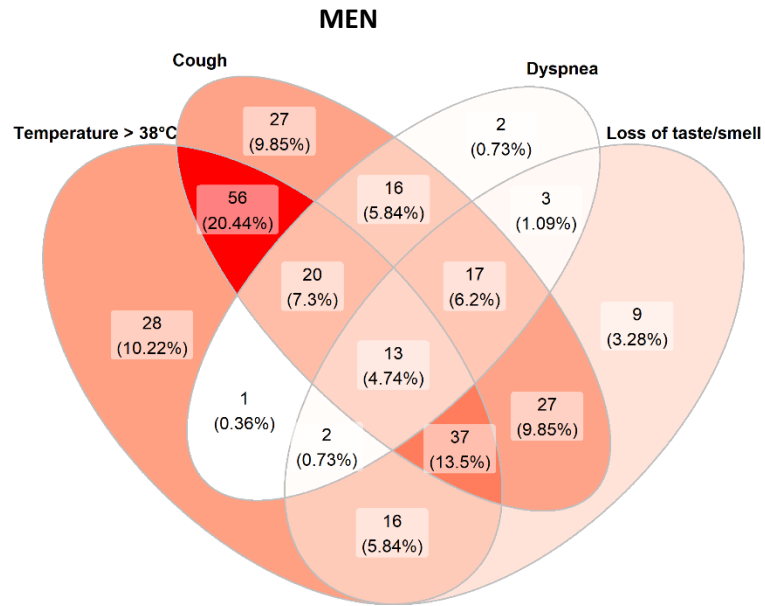
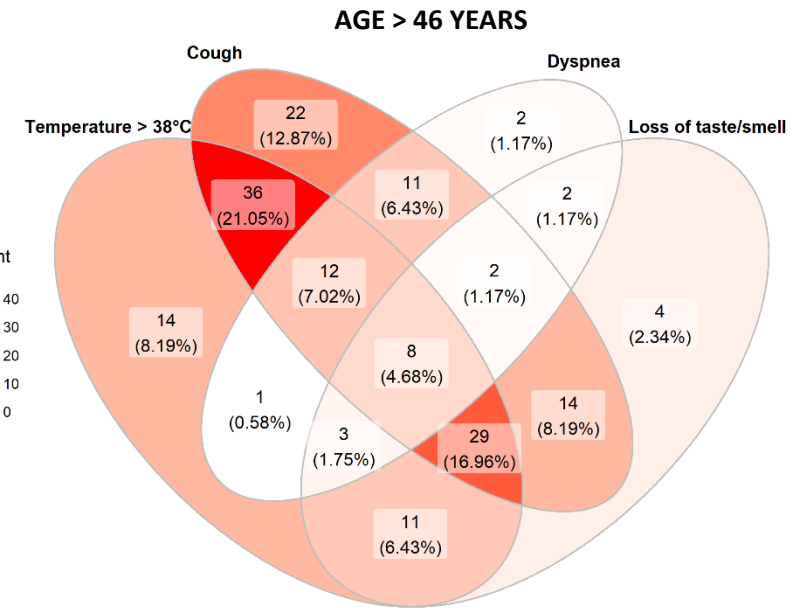
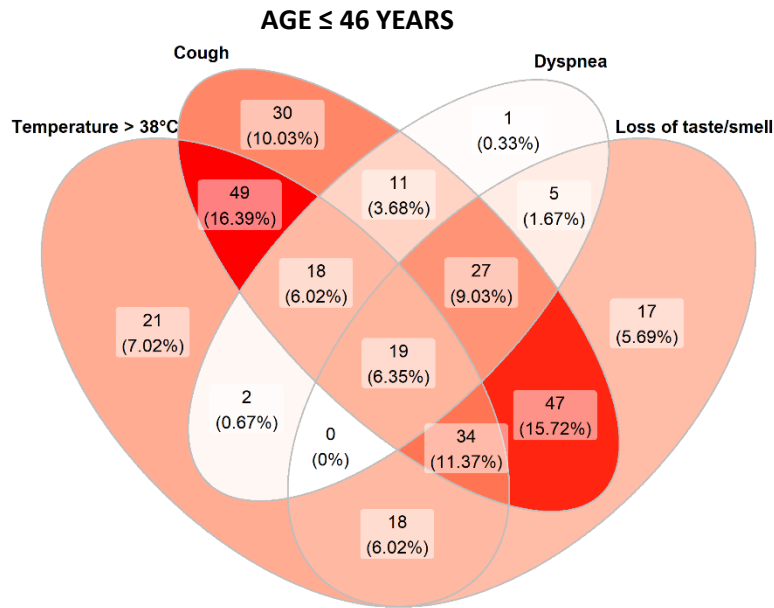


WOMEN

SARS-CoV-2(-) N=443 SARS-CoV-2(+) N=205



N is a number of people who answered the questions

a**b**

Sample N = 3114

Continuous variables:

	N	Mean	SD	Median	Q1 [25%]	Q3 [75%]	Min	Max	no. of missing data
Age	3111	43,29	14,45	42	32	54	1	98	3
Saturation	3006	97,20	1,32	97	97	98	88	100	108
Temp. in hospital	2976	36,47	0,46	36,5	36,3	36,6	33,9	40,2	138
Max. temperature	1916	37,68	1,05	37,7	37	38,5	30,9	41,2	1198
Days of symptoms	1827	5,35	8,03	4	2	7	0	180	1287

Discrete variables:

	Yes	No	N	no. of missing data	Yes [% of N]	No [% of N]
Sex	1559	1554	3113	1	50,08	49,92
Number of infections	759	2355	3114	0	24,37	75,63
Health state	1941	1173	3114	0	62,33	37,67
Contact with infection	1227	1785	3012	102	40,74	59,26
Temperature > 38	931	2183	3114	0	29,90	70,10
Cough	1588	1526	3114	0	51,00	49,00
Dyspnea	724	2390	3114	0	23,25	76,75
Conjunctivitis	9	740	749	2365	1,20	98,80
Muscle aches	958	1288	2246	868	42,65	57,35
Loss of taste/smell	519	1725	2244	870	23,13	76,87
Sore throat	679	1566	2245	869	30,24	69,76
Tiredness	83	749	832	2282	9,98	90,02
Headache	971	1278	2249	865	43,17	56,83
Dizziness	348	1703	2051	1063	16,97	83,03
Skin reactions	60	2163	2223	891	2,70	97,30
Stomach symptoms	46	785	831	2283	5,54	94,46
Loss of appetite	8	762	770	2344	1,04	98,96
Lung disease	47	3067	3114	0	1,51	98,49
Hypertension	533	2581	3114	0	17,12	82,88
Heart diseases	163	2951	3114	0	5,23	94,77
Diabetes	133	2981	3114	0	4,27	95,73
Kidney diseases	34	3080	3114	0	1,09	98,91
Allergies	84	3030	3114	0	2,70	97,30
Asthma	133	2981	3114	0	4,27	95,73

Blood type:	n	% of N
0	470	39,13
A	407	33,89
AB	115	9,58
B	209	17,4
N	1201	100
no.of missing data	1913	

characteristics	All			Women			Men			Young			Old		
	Odds Ratio	95% Confidence Interval - Lower Bound	95% Confidence Interval - Upper Bound	Odds Ratio	95% Confidence Interval - Lower Bound	95% Confidence Interval - Upper Bound	Odds Ratio	95% Confidence Interval - Lower Bound	95% Confidence Interval - Upper Bound	Odds Ratio	95% Confidence Interval - Lower Bound	95% Confidence Interval - Upper Bound	Odds Ratio	95% Confidence Interval - Lower Bound	95% Confidence Interval - Upper Bound
Sex *	0,870	0,710	1,050	-	-	-	-	-	-	0,729	0,567	0,937	1,102	0,806	1,505
Contact with infected person	1,790	1,450	2,210	1,650	1,220	2,240	1,970	1,470	2,660	1,890	1,440	2,460	1,680	1,190	2,370
Temperature > 38oC	1,670	1,370	2,030	1,780	1,340	2,370	1,530	1,170	2,010	1,530	1,190	1,960	1,930	1,400	2,650
Cough	1,260	1,000	1,580	1,120	0,800	1,570	1,410	1,040	1,920	1,140	0,850	1,540	1,480	1,040	2,110
Dyspnea	0,550	0,440	0,680	0,520	0,380	0,710	0,580	0,430	0,770	0,560	0,430	0,730	0,530	0,370	0,770
Conjunctivitis	0,240	0,050	1,140	0,550	0,110	2,910	0,310	0,060	1,660	0,280	0,060	1,410	0,630	0,110	3,670
Muscle aches	1,300	1,030	1,630	1,200	0,850	1,690	1,420	1,040	1,920	1,730	1,280	2,350	0,860	0,600	1,230
Loss of taste/smell	3,330	2,630	4,220	3,990	2,810	5,670	3,000	2,170	4,160	3,900	2,890	5,260	2,560	1,730	3,790
Sore throat	0,850	0,680	1,070	0,660	0,470	0,920	1,130	0,830	1,540	0,700	0,530	0,930	1,210	0,830	1,760
Tiredness	0,540	0,320	0,930	0,730	0,350	1,550	0,490	0,230	1,030	0,570	0,290	1,100	0,500	0,190	1,270
Headache	1,270	1,010	1,590	1,150	0,810	1,640	1,440	1,060	1,940	1,410	1,040	1,900	1,060	0,740	1,530
Dizziness	0,810	0,620	1,060	0,980	0,660	1,440	0,700	0,490	1,010	0,700	0,500	0,980	1,060	0,680	1,640
Skin reactions	1,080	0,630	1,880	1,230	0,590	2,570	1,060	0,490	2,300	1,110	0,550	2,260	1,150	0,520	2,580
GI symptoms	0,290	0,140	0,600	0,470	0,190	1,160	0,250	0,090	0,710	0,290	0,120	0,700	0,400	0,140	1,170
Loss of appetite	1,650	0,510	5,330	1,730	0,470	6,270	1,780	0,310	10,210	1,390	0,360	5,380	1,730	0,360	8,170
Lung disease	0,440	0,180	1,050	0,640	0,210	1,980	0,460	0,150	1,390	1,190	0,290	4,860	0,380	0,140	1,000
Hypertension	0,910	0,700	1,170	0,940	0,640	1,360	0,880	0,620	1,250	0,790	0,490	1,260	0,910	0,650	1,260
Heart diseases	0,640	0,410	1,000	0,490	0,250	0,950	0,900	0,500	1,620	0,860	0,400	1,860	0,560	0,330	0,960
Diabetes	0,750	0,460	1,200	0,560	0,260	1,240	0,930	0,530	1,650	0,790	0,310	2,020	0,730	0,430	1,240
Kidney diseases	0,960	0,420	2,220	1,230	0,410	3,660	0,960	0,330	2,820	1,190	0,290	4,860	0,990	0,400	2,480
Allergies	0,460	0,230	0,910	0,450	0,180	1,090	0,660	0,260	1,690	0,500	0,220	1,140	0,580	0,210	1,590
Asthma	0,680	0,420	1,110	0,810	0,430	1,510	0,630	0,310	1,300	0,460	0,230	0,930	1,210	0,620	2,370

LEGEND:	
	significant odds ratio higher than 1 (potential risk factor)
	significant odds ratio smaller than 1 (potential protective factor)