BMJ Open

Association between socioeconomic status and self-reported, tested and diagnosed COVID-19 status during the first wave in the Northern Netherlands: a general population-based cohort from 49 474 adults

Yinjie Zhu ![ORCID],[1] Ming-Jie Duan,[1] Hermien H. Dijk,[2] Roel D. Freriks,[2] Louise H. Dekker,[1,3] Jochen O. Mierau,[2,3] Lifelines Corona Research initiative[2,4,5,6]

ABSTRACT

Objectives Studies in clinical settings showed a potential relationship between socioeconomic status (SES) and lifestyle factors with COVID-19, but it is still unknown whether this holds in the general population. In this study, we investigated the associations of SES with self-reported, tested and diagnosed COVID-19 status in the general population.

Design, setting, participants and outcome measures Participants were 49 474 men and women (46±12 years) residing in the Northern Netherlands from the Lifelines cohort study. SES indicators and lifestyle factors (i.e., smoking status, physical activity, alcohol intake, diet quality, sleep time and TV watching time) were assessed by questionnaire from the Lifelines Biobank. Self-reported, tested and diagnosed COVID-19 status was obtained from the Lifelines COVID-19 questionnaire.

Results There were 4711 participants who self-reported having had a COVID-19 infection, 2883 participants tested for COVID-19, and 123 positive cases were diagnosed in this study population. After adjustment for age, sex, lifestyle factors, body mass index and ethnicity, we found that participants with low education or low income were less likely to self-report a COVID-19 infection (OR [95% CI]: low education 0.78 [0.71 to 0.86]; low income 0.86 [0.79 to 0.93]) and be tested for COVID-19 (OR [95% CI]: low education 0.58 [0.52 to 0.66]; low income 0.86 [0.78 to 0.95]) compared with high education or high income groups, respectively.

Conclusion Our findings suggest that the low SES group was the most vulnerable population to self-reported and tested COVID-19 status in the general population.

INTRODUCTION

The ongoing COVID-19 pandemic has already infected 62 363 527 people worldwide (as of 1 December 2020).[1] Many studies have reported a relationship between lifestyle factors (e.g., smoking[2] and unhealthy diets[3]) or comorbidities associated with lifestyle

Strengths and limitations of this study

► This study added evidence to the socioeconomically patterned COVID-19 status in a general population instead of in clinical settings.
► This study innovatively included a broader range of COVID-19 status, including self-reported and tested COVID-19 status, to better understand COVID-19-related socioeconomic factors.
► This study might fail to identify the ethnic differences because the study population was predominantly white (more than 98%).
► During the first wave of the COVID-19 pandemic in the Northern Netherlands, a number of cases identified were relatively low compared with the rest of the country, which might limit the power of the analysis about diagnosed COVID-19 status.

factors (e.g., obesity[4] and diabetes mellitus[5]) and worse outcomes for COVID-19. Since lifestyle factors are often associated with socioeconomic status (SES),[6] several studies have revealed that people with low SES are more susceptible to COVID-19 infection, hospitalisation and mortality,[7–11] the interplay between SES, lifestyle and COVID-19 status becomes a valid query and it is likely that SES is a determinant of COVID-19.

Therefore, a rising concern among policymakers and public health practitioners is that the COVID-19 pandemic might exacerbate the persistent socioeconomically patterned health inequalities over the life course.[12 13] Still, an in-depth investigation of the lifestyle and socioeconomic determinants of COVID-19 in the general population is needed because current COVID-19 studies mainly relied on individuals presenting with symptoms in a

clinical setting and focused on post-COVID-19 infection status (e.g., hospitalisation and intensive care).[14 15] As such, we do not have enough knowledge on the validity of these early results for the general population.

This study aimed to investigate the extent to which SES and lifestyle factors were associated with a broader range of COVID-19 status, that is, self-reported COVID-19, tested COVID-19 and diagnosed COVID-19 in the general population.

METHODS

Study design and participants

To assess COVID-19 status in the general population, we used data from the Lifelines COVID-19 Cohort. The Lifelines COVID-19 Cohort is a questionnaire-based study in a part of the Dutch Lifelines cohort.[16 17] To study the relationship between COVID-19 and SES, we linked data from the Lifelines COVID-19 questionnaire with data from the general Lifelines cohort, which contains a wide range of background data, such as demographics, lifestyles and SES.[18 19] For the present study, 49 474 participants from both the Lifelines cohort and the Lifelines COVID-19 Cohort, who had available and reliable data on demographics, SES and lifestyle, were included in the analysis (online supplemental figure 1).

The Lifelines cohort study is a multidisciplinary prospective population-based cohort study based on a unique three-generation design. The participants were recruited from the three Northern provinces of the Netherlands between 2006 and 2013 and the first group of participants were recruited via local general practitioners. The participants could indicate whether their family members were interested to be recruited as well. In addition, individuals who were interested in the study could also participant through online self-registration. Participants with insufficient knowledge of the Dutch language, with severe psychiatric or physical illness and those with limited life expectancy (<5 years) were excluded from the study. The Lifelines cohort study employs a broad range of investigative procedures in assessing the biomedical, sociodemographic, behavioural, physical and psychological factors that contribute to health and disease. A detailed description of the Lifelines cohort study can be found elsewhere.[18 19] Before study entry, a signed informed consent form was obtained from each participant. Adult participants (≥18 years) were asked to complete several questionnaires regarding various aspects, including demographics, SES and lifestyle.

The Lifelines COVID-19 Cohort collected data about COVID-19-related symptoms, current health issue and societal impacts from participants recruited from the Lifelines cohort. It is developed based on a COVID-19 questionnaire to identify genetic and environmental risk factors for COVID-19 and address the medical, social and psychological aspects of the pandemic. A detailed description of the Lifelines COVID-19 Cohort and the COVID-19 questionnaire can be found elsewhere.[16] The COVID-19 questionnaire is sent out weekly since 30 March 2020 for 12 weeks. Of the 139 713 Lifelines participants invited, 74 268 (53.2%) completed at least one of the questionnaires in the 12-week programme.

COVID-19 status

There were three different COVID-19 statuses obtained in our study from the COVID-19 questionnaire: self-reported COVID-19, tested COVID-19 and diagnosed COVID-19. All COVID-19 statuses were coded as binary variables. Self-reported COVID-19 status was obtained by asking 'if you must choose, do you think you have (or have had) a COVID-19 infection?' or 'has a doctor told you that you probably have (had) a coronavirus/COVID-19 infection'; tested COVID-19 status was obtained from the question 'have you been tested for coronavirus (COVID-19)?'; diagnosed COVID-19 status was defined by asking 'do you have or have you had a coronavirus/COVID-19 infection?' or 'what was the results of your corona virus (COVID-19) test?' when they had a COVID-19 test.

Since some of the COVID-related questions were worded such that it is not possible to infer whether the answer referred to the period between the last questionnaire and the current questionnaire or to the beginning of the pandemic and the current questionnaire (e.g., "have you been tested for coronavirus (COVID-19)?"), the 12 questionnaires were condensed into a single observation per individual, indicating for each question whether the individual had at any point answered 'yes' during the study period.

Socioeconomic status

SES was indicated by (a) highest educational level achieved and (b) monthly net household income level separately from the Lifelines cohort study at baseline. Highest educational level achieved was categorised as: (1) low—junior general secondary education or lower (International Standard Classification of Education [ISCED] level 0, 1 or 2); (2) middle—secondary vocational education and senior general secondary education (ISCED level 3 or 4) and (3) high—higher vocational education or university (ISCED level 5 or 6)[20]; household net income level was categorised as: (1) low: <2000 euro/month; (2) middle: 2000–3000 euro/month and (3) high: >3000 euro/month.

Lifestyle factors

Six lifestyle factors (i.e., smoking status, alcohol consumption, diet quality, physical activity, TV watching time and sleep time) were selected from the Lifelines cohort study at baseline. Smoking status was categorised into never, former and current smoker. Alcohol intake and dietary consumption were derived from a validated 110-item semiquantitative food frequency questionnaire (FFQ) that assessed food consumption over the past month.[21] Heavy drinking was defined as >40 g or >20 g average per day alcohol consumption for men and women, respectively.[22] Lifelines Diet

Score was calculated to assess the overall diet quality based on the FFQ. This score ranks the relative intake of nine food groups with positive health effects (vegetables, fruit, whole grain products, legumes/nuts, fish, oils/soft margarines, unsweetened dairy, coffee and tea) and three food groups with negative health effects (red/processed meat, butter/hard margarines and sugar-sweetened beverages). The development of this score is described in detail elsewhere.[23] Nonoccupational moderate-to-vigorous physical activity was calculated in minutes per week from the validated Short Questionnaire to ASsess Health-enhancing physical activity data, which incorporated leisure time and commuting physical activities, including sports, at moderate (4.0–6.4 metabolic equivalent of task [MET]) to vigorous (≥6.5 MET) intensity.[24] TV watching time and sleep time were self-reported. Body mass index (BMI) was calculated by dividing weight in kilograms by the square of height in metres. The BMI was additionally categorised into underweight ($BMI<18.5$ kg/m^2), normal ($18.5≤BMI<25$ kg/m^2), overweight ($25≤BMI<30$ kg/m^2) and obese ($BMI≥30$ kg/m^2).[25]

Statistical analysis

Nominal variables are presented as percentage (%). Continuous variables were shown as mean±standard deviation (SD) or median (interquartile range [IQR]). P values <0.05 were considered statistically significant. To analyse the associations of SES (i.e., education and income) with self-reported and tested COVID-19 status, we used logistic regression models to estimate ORs. Furthermore, we fitted robust Poisson regression models to estimate relative risk for associations between SES and diagnosed COVID-19 since the number of individuals with diagnosed COVID-19 was relatively low in this population. For all regression models, SES indicators were first entered into the model plus age and sex (model 1) and then adjusted for other covariates (model 2: model 1 plus six lifestyle factors; model 3: model 2 plus BMI; model 4: model 3 plus ethnicity). All statistical analyses were conducted using Stata, V.13.1 (StataCorp, Texas, USA) or RStudio, V.4.0 (RStudio PBC, Boston, USA). Sensitivity analyses were conducted by further adjusting for the total number of questionnaires filled by each participant (online supplemental table 1), by comparing the characteristics of study population and excluded participants (online supplemental table 2), and by treating lifestyle factors together as a mediator in the pathway between SES and COVID-19 status (online supplemental table 3).

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

RESULTS

Participant characteristics are shown in table 1. The average age of the study population was 46±12 years old and more women were recruited in this study (59.7%). More than 92% of the population characterised their ethnicity as White. The prevalence of overweight and obese was 40.0% and 14.5%, respectively (table 1). Out of the 49 474 participants who were recruited in both Lifelines cohort and Lifelines COVID-19 cohort, 4711 participants self-reported a COVID-19 infection, while 2833 participants reported having had a COVID-19 test and 123 participants reported having had a positive outcome of the COVID-19 test (table 1). Self-reported, tested and diagnosed COVID-19 participants were more likely to be female, to have attained middle or high education, to have high income and to have never smoked (table 1).

The association between education and different COVID-19 status is presented in table 2. Participants with low and middle education level were less likely to self-report COVID-19 (OR [95% CI]: low 0.78 [0.71 to 0.86]; middle 0.90 [0.84 to 0.97]) and be tested for COVID-19 (OR [95% CI]: low 0.58 [0.52 to 0.66]; middle 0.72 [0.66 to 0.79]) compared with highly educated participants after adjustment for age, sex, lifestyle factors, ethnicity and BMI (table 2). In addition, compared with high education participants, participants with middle education level had almost two-fold higher risk to be infected with COVID-19 (OR [95% CI]: 1.73 [1.19 to 2.50]) (table 2, model 1) once they had been tested for COVID-19. The association was slightly attenuated after adjustment for lifestyle factors (OR [95% CI]: 1.69 [1.17 to 2.44]) (table 2, model 2) and lifestyle factors seemed to slightly mediate the association between self-reported COVID-19 status (proportion mediated: 9.9%) (online supplemental table 3), however, further adjustment for BMI and ethnicity did not change the results.

The association between income and different COVID-19 status is presented in table 3. Low or middle income groups were less likely to self-report COVID-19 (OR [95% CI]: low 0.86 [0.79 to 0.93]; middle 0.80 [0.74 to 0.87]) and to be tested for COVID-19 (OR [95% CI]: low 0.86 [0.78 to 0.95]; middle 0.86 [0.78 to 0.94]) compared with high-income group after adjustment for all covariates (table 3, model 4). Moreover, lifestyle factors only slightly mediated the association between self-reported and tested COVID-19 statuses (proportion mediated: 6.7% and 9.5%, respectively) (online supplemental table 3). Nevertheless, no risk differences were found between different income levels and diagnosed COVID-19 status, once tested.

Table 1 Demographics, socioeconomic status and lifestyle of the study population

	COVID-19 status			
	Self-reported (n=4711)	Tested (2883)	Diagnosed (123)	Total response (n=49 474)
Sex, male%	37.4	30.4	30.1	40.3
Age, mean±SD	42±11	44±12	44±9	46±12
Ethnicity, white%	89.4	92.8	94.3	92.0
BMI, mean±SD	25.9±4.3	25.7±4.3	25.7±3.8	26.0±4.2
Underweight,%	0.7	1.0	0.8	0.7
Normal,%	46.7	48.0	47.2	44.9
Overweight,%	37.2	36.6	40.7	40.0
Obese,%	15.4	14.5	11.4	14.5
Education,%				
Low	18.0	16.4	14.6	23.5
Middle	40.8	37.4	48.0	39.2
High	41.2	46.2	37.4	37.3
Income, %				
Low	30.1	28.3	23.6	28.1
Middle	29.5	30.7	33.3	33.7
High	40.4	41.0	43.1	38.2
Smoking, %				
Current	20.5	15.1	15.5	16.9
Former	33.0	36.1	40.7	36.3
Never	46.5	48.9	43.9	46.8
TV watching time ≥4 hours/day,%	15.2	15.1	18.7	16.7
Sleep time <7 or>9 hours/day, %	16.3	14.4	13.8	14.9
MVPA <150 min/week,%	38.8	37.5	29.3	37.9
LLDS	24.3±6.1	25.1±6.0	25.4±5.9	24.7±6.0
Alcohol, g/day	4.9 (1.3–11.0)	3.9 (0.9–10.0)	3.5 (0.9–9.0)	4.9 (1.2–11.1)
Heavy drinker, %	3.1	2.8	2.4	2.6

BMI, body mass index; LLDS, lifelines diet score; MVPA, non-occupational moderate-to-vigorous physical activity; SD, standard deviation.

DISCUSSION

This paper has provided evidence and insight into socioeconomic disparities in self-reported COVID-19, tested COVID-19 and diagnosed COVID-19 in a general population. We found that low and middle SES groups, especially participants with low education, were less likely to self-report COVID-19 and to be tested for COVID-19. Moreover, participants with middle education were more likely to be infected with COVID-19 compared with those who obtained high education.

Our study highlighted the importance of mitigating the socioeconomically patterned effect of the COVID-19 pandemic by showing a clear relationship between SES and COVID-19 status, which enables and aids policymakers to adequately respond to the pandemic and develop public health measurements. It is worrisome that low SES group is likely to be neglected or undetected during the pandemic because they are less likely to self-report and be tested for COVID-19. Given the fact that they are also at high risk of noncommunicable diseases,[26] once infected

with COVID-19 or other infectious diseases in the future, individuals with low SES might suffer disproportionately from the disease development compared with the high SES group, which may further exacerbate health inequality.[27] Therefore, national or regional commitment of public health measures, guidelines or interventions are needed to promote health equity before, during and after COVID-19 pandemic by identifying, targeting and engaging the low SES group.

One of the potential mechanisms of the inverse association between SES and COVID-19 status is poor health literacy. Health literacy refers to 'a person's knowledge, motivation and competences to access, understand, appraise and apply health information in order to make judgements and take decisions in everyday life'.[28] It is known that low SES is the most important determinant of health literacy.[29] During the COVID-19 pandemic, people with high SES may be more aware of the symptoms and health consequences of COVID-19, and can, therefore, better interpret and communicate and make a decision about their own health condition.[30 31] In other words, low SES people

Table 2 Association of education with different COVID-19 status

Education	COVID-19 status					
	Self-reported		Tested		Diagnosed	
	OR (95% CI)	p	OR (95% CI)	p	RR (95% CI)	p
Model 1						
Low	0.81 [0.75–0.89]	**<0.001**	0.58 [0.52–0.65]	**<0.001**	1.18 [0.67–2.07]	0.6
Middle	0.93 [0.87–0.99]	**0.03**	0.73 [0.67–0.80]	**<0.001**	1.73 [1.19–2.50]	**0.004**
High	Ref					
Model 2						
Low	0.79 [0.72–0.86]	**<0.001**	0.59 [0.53–0.66]	**<0.001**	1.11 [0.61–2.00]	0.7
Middle	0.91 [0.85–0.97]	**0.006**	0.74 [0.68–0.81]	**<0.001**	1.69 [1.17–2.44]	**0.005**
High	Ref					
Model 3						
Low	0.77 [0.71–0.85]	**<0.001**	0.59 [0.53–0.66]	**<0.001**	1.11 [0.61–2.00]	0.7
Middle	0.90 [0.84–0.96]	**0.002**	0.74 [0.68–0.81]	**<0.001**	1.69 [1.16–2.45]	**0.006**
High	Ref					
Model 4						
Low	0.78 [0.71–0.86]	**<0.001**	0.58 [0.52–0.66]	**<0.001**	1.10 [0.60–2.01]	0.8
Middle	0.90 [0.84–0.97]	**0.004**	0.72 [0.66–0.79]	**<0.001**	1.77 [1.21–2.60]	**0.003**
High	Ref					

Model 1: adjusted for education, age and sex. Model 2: adjusted for education, age, sex, and six lifestyle factors (smoking status, TV watching time ≥ 4h/day, sleep time <7 or >9 h/day, MVPA<150 min/week, LLDS, and heavy drinker). Model3: adjusted for education, age, sex, six lifestyle factors (smoking status, TV watching time ≥ 4h/day, sleep time <7 or >9 h/day, MVPA<150 min/week, LLDS, and heavy drinker), and BMI. Model 4: adjusted for education, age, sex, six lifestyle factors (smoking status, TV watching time ≥ 4h/day, sleep time <7 or >9 h/day, MVPA<150 min/week, LLDS, and heavy drinker), BMI, and ethnicity.
BMI, body mass index; LLDS, lifelines diet score; MVPA, non-occupational moderate-to-vigorous physical activity; OR, odds ratio; RR, risk ratio.

might be less sensitive and aware of their COVID-19-related symptoms, or they might lack the ability to assess whether they could have had COVID-19 infection. Since they might lack health literacy, low SES people might be less likely to seek medical help even when they have symptoms because of the lack of practicing health-seeking behaviour and critical health resources.[12 13] On the other hand, potentially low SES individuals had more difficulties obtaining a test even when they had sufficient health literacy. Our results also suggested that education level is a better determinant for COVID-19 status than income[29] because the association gradient was larger for low education participants than for low income participants. Further research is needed to fully understand the observed link between socioeconomic disparities and COVID-19.

Furthermore, there could be other explanations of the inverse association of SES and COVID-19 status. Despite the fact that the low SES group is at higher risk of noncommunicable disease, they also suffer from a higher general disease burden both mentally and financially during the pandemic.[26] Therefore, they might be less likely to notice the relevant symptom due to the already existing disease burden. Additionally, the first wave of COVID-19 pandemic in the Netherlands covered the winter holiday, and people with high income and education were more likely to go for a relatively expensive skiing holiday compared with the low SES group. The skiing area in Italy has always been a popular destination for Dutch people and Northern Italy was the first outbreak centre during the first wave of this pandemic,[32] so that people with high income or high education who came back from the skiing holiday in Italy were more likely to self-report and be tested for COVID-19.

Interestingly, lifestyle factors only slightly mediated the pathway between SES and self-reported, and tested COVID-19 status, and did not seem to be a mediator in the pathway between SES and diagnosed COVID-19 status in our study, which is different from a previous study that suggested an unhealthy lifestyle that was a profound risk factor for COVID-19 status.[15] Some other studies also found that nutritional status[33 34] and smoking[2] might also be risk factors for COVID-19 disease severity, disease progression and mortality. Because of the differences of study population and COVID-19 outcome selection, our results were not comparable with previous studies. Nevertheless, this study provided a fresh message that in the general population, SES is a better determinant of self-reported, tested and diagnosed COVID-19

Table 3 Association of income with different COVID-19 status

Income	COVID-19 status					
	Self-reported		Tested		Diagnosed	
	OR (95% CI)	p	OR (95% CI)	p	RR (95% CI)	p
Model 1						
Low	0.89 [0.83–0.96]	**0.002**	0.85 [0.77–0.93]	**0.001**	0.84 [0.54–1.31]	0.4
Middle	0.82 [0.76–0.88]	**<0.001**	0.84 [0.77–0.92]	**<0.001**	1.14 [0.77–1.70]	0.5
High	Ref					
Model 2						
Low	0.88 [0.81–0.95]	**0.001**	0.87 [0.79–0.96]	**0.005**	0.80 [0.51–1.25]	0.3
Middle	0.82 [0.76–0.88]	**<0.001**	0.86 [0.78–0.94]	**0.001**	1.10 [0.73–1.64]	0.6
High	Ref					
Model 3						
Low	0.88 [0.81–0.95]	**0.001**	0.87 [0.79–0.96]	**0.006**	0.80 [0.51–1.25]	0.3
Middle	0.81 [0.76–0.87]	**<0.001**	0.86 [0.78–0.94]	**0.001**	1.10 [0.73–1.64]	0.7
High	Ref					
Model 4						
Low	0.86 [0.79–0.93]	**<0.001**	0.86 [0.78–0.95]	**0.002**	0.74 [0.47–1.18]	0.2
Middle	0.80 [0.74–0.87]	**<0.001**	0.86 [0.78–0.94]	**0.002**	1.05 [0.70–1.57]	0.8
High	Ref					

Model 1: adjusted for income, age and sex. Model 2: adjusted for income, age, sex, and six lifestyle factors (smoking status, TV watching time ≥ 4h/day, sleep time <7 or >9 h/day, MVPA<150 min/week, LLDS, and heavy drinker). Model3: adjusted for income, age, sex, six lifestyle factors (smoking status, TV watching time ≥ 4h/day, sleep time <7 or >9 h/day, MVPA<150 min/week, LLDS, and heavy drinker), and BMI. Model 4: adjusted for income, age, sex, six lifestyle factors (smoking status, TV watching time ≥ 4h/day, sleep time <7 or >9 h/day, MVPA<150 min/week, LLDS, and heavy drinker), BMI, and ethnicity.
BMI, body mass index; LLDS, lifelines diet score; MVPA, non-occupational moderate-to-vigorous physical activity; OR, odds ratio; RR, risk ratio.

status. In addition, unlike other COVID-19 studies that found, obesity and ethnicity were both risk factors of COVID-19 infection, hospitalisation and mortality,[11 15 35–38] our results showed that obesity and ethnicity did not contribute substantially to the disparities in COVID-19 status, which indicates that obesity and ethnicity were not in the pathway between SES and self-reported, tested and diagnosed COVID-19 status.

This study has several strengths. First, this general population-based study about COVID-19 may to some extent mitigate the selection bias compared with studies conducted based in clinical settings and can allow us to thoroughly understand the relation between SES and COVID-19 in the general population. Second, we innovatively included a broader range of COVID-19 status, including self-reported and tested COVID-19 status, to better understand COVID-19-related health literacy and accessibility to health services. Still, this study failed to link the objective measures of COVID-19 infection status, which could cause some bias. Another limitation of the current study is the fact that the study population was predominantly white (more than 98%). Additionally, the Netherlands has a well-developed social security system. This may limit its generalisability to populations of other ethnicity, and in a different social context. One caveat of our study is that the relative low number of cases in the region, which is the nature of the COVID-19 outbreak in the Northern Netherlands (online supplemental table 4), might influence

the statistical power of the study. However, we still find statistically significant results for SES and COVID-19 status. Another limitation is from the condensed observations about COVID-19 status since certain types of individuals might be more likely to fill in more questionnaires and are, therefore, more likely to report a positive COVID-19 status in at least one of the questionnaires. If these individuals predominantly belong to a specific SES group, this would bias the results. However, by further adjusting for the total number of questionnaires filled by each participants, we did not observe any difference from the current results (online supplemental table 1). Additionally, although we only observed minor effect of lifestyle factor on the association between SES and COVID-19 status, the true effect of lifestyle factors could be more pronounced since the low SES group tends to report their lifestyle according to the social desire. Finally, approximately 17% of the participants were dropped because of missing SES data, while 16% of the participants were dropped because of incomplete lifestyle information, which might bias the results given that people with low SES and worse lifestyle practices could be less likely to participant in the COVID-19 questionnaire. Nevertheless, both the study population and excluded participants have similar characteristics (online supplemental table 2). Consequently, it is less likely that there was a selection bias resulting from missing variables.

In conclusion, our findings of SES and self-reported, tested and diagnosed COVID-19 status indicate that individuals with lower SES may be more vulnerable and neglected during

COVID-19 pandemic, possibly because of disease burden, poor health literacy and access to healthcare, suggesting that health services and health promotion interventions should particularly focus on targeting individuals with lower SES to get them better prepared for future health emergencies.

Author affiliations
[1]Department of Internal Medicine, Division of Nephrology, University Medical Center Groningen, Groningen, The Netherlands
[2]Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands
[3]Aletta Jacobs School of Public Health, University of Groningen, Groningen, The Netherlands
[4]Department of Epidemiology, University Medical Center Groningen, Groningen, The Netherlands
[5]Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands
[6]Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, Groningen, The Netherlands

ORCID iD
Yinjie Zhu http://orcid.org/0000-0001-8059-6446

## REFERENCES

1. WHO. WHO coronavirus disease (COVID-19) Dashboard, 2020. Available: https://covid19.who.int/?gclid=EAIaIQobChMIzsuel8uQ7AIVgrh3Ch2FbgICEAAYASAAEgL5y_D_BwE
2. Vardavas C, Nikitara K. COVID-19 and smoking: a systematic review of the evidence. *Tob Induc Dis* 2020;18.
3. Butler MJ, Barrientos RM. The impact of nutrition on COVID-19 susceptibility and long-term consequences. *Brain Behav Immun* 2020;87:53–4.
4. Tartof SY, Qian L, Hong V, *et al*. Obesity and mortality among patients diagnosed with COVID-19: results from an integrated health care organization. *Ann Intern Med* 2020;173:773–81.
5. Cuschieri S, Grech S. COVID-19 and diabetes: the why, the what and the how. *J Diabetes Complications* 2020;34:107637.
6. Glorioso V, Pisati M. Socioeconomic inequality in health-related behaviors: a lifestyle approach. *Qual Quant* 2014;48:2859–79.
7. Batty GD, Deary IJ, Luciano M, *et al*. Psychosocial factors and hospitalisations for COVID-19: prospective cohort study based on a community sample. *Brain Behav Immun* 2020;89:569–78.
8. McCarthy CP, Murphy S, Jones-O'Connor M, *et al*. Early clinical and sociodemographic experience with patients hospitalized with COVID-19 at a large American healthcare system. *EClinicalMedicine* 2020;26:100504.
9. Khalatbari-Soltani S, Cumming RC, Delpierre C, *et al*. Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. *J Epidemiol Community Health* 2020;74:620–3.
10. Baena-Díez JM, Barroso M, Cordeiro-Coelho SI, *et al*. Impact of COVID-19 outbreak by income: hitting hardest the most deprived. *J Public Health* 2020;42:698–703.
11. Lieberman-Cribbin W, Tuminello S, Flores RM, *et al*. Disparities in COVID-19 testing and positivity in New York City. *Am J Prev Med* 2020;59:326–32.
12. Burström B, Tao W. Social determinants of health and inequalities in COVID-19. *Eur J Public Health* 2020;30:617–8.
13. Schulz AJ, Mehdipanah R, Chatters LM, *et al*. Moving health education and behavior upstream: lessons from COVID-19 for addressing structural drivers of health inequities. *Health Educ Behav* 2020;47:519–24.
14. Wynants L, Van Calster B, Collins GS, *et al*. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020;369:m1328.
15. Hamer M, Kivimäki M, Gale CR, *et al*. Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: a community-based cohort study of 387,109 adults in UK. *Brain Behav Immun* 2020;87:184–7.
16. Mc Intyre K. The lifelines COVID-19 cohort: a questionnaire-based study to investigate COVID-19 infection and its health and societal impacts in a Dutch population-based cohort. *medRxiv* 2020:2020.06.19.20135426. doi:10.1101/2020.06.19.20135426
17. Varga TV, Bu F, Dissing AS. Loneliness, worries, anxiety, and precautionary behaviours in response to the COVID-19 pandemic: a longitudinal analysis of 200,000 Western and Northern Europeans. *The Lancet Regional Health - Europe* 2021;9:100020.
18. Klijs B, Scholtens S, Mandemakers JJ, *et al*. Representativeness of the lifelines cohort study. *PLoS One* 2015;10:e0137203.
19. Scholtens S, Smidt N, Swertz MA, *et al*. Cohort profile: lifelines, a three-generation cohort study and Biobank. *Int J Epidemiol* 2015;44:1172–80.
20. International standard classification of education. *United nations educational, scientific, and cultural organization*, 1997.
21. Siebelink E, Geelen A, de Vries JHM. Self-Reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. *Br J Nutr* 2011;106:274–81.
22. Dawson DA. Defining risk drinking. *Alcohol Res Health* 2011;34:144–56.
23. Vinke PC, Corpeleijn E, Dekker LH, *et al*. Development of the food-based lifelines diet score (LLDS) and its application in 129,369 lifelines participants. *Eur J Clin Nutr* 2018;72:1111–9.
24. Wendel-Vos GCW, Schuit AJ, Saris WHM, *et al*. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *J Clin Epidemiol* 2003;56:1163–9.
25. WHO. *Obesity : preventing and managing the global epidemic : report of a WHO consultation, in WHO Technical Report Series*. Geneva, Switzerland: World Health Organization, 1999.

26 Collaborators GBDCoD, GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;392:1736–88.

27 Marmot M, Allen J. COVID-19: exposing and amplifying inequalities. *J Epidemiol Community Health* 2020;74:681.

28 Sørensen K, Van den Broucke S, Fullam J, *et al*. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12:80.

29 Stormacq C, Van den Broucke S, Wosinski J. Does health literacy mediate the relationship between socioeconomic status and health disparities? integrative review. *Health Promot Int* 2019;34:e1–17.

30 Zhong B-L, Luo W, Li H-M, *et al*. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. *Int J Biol Sci* 2020;16:1745–52.

31 Alobuia WM, Dalva-Baird NP, Forrester JD, *et al*. Racial disparities in knowledge, attitudes and practices related to COVID-19 in the USA. *J Public Health* 2020;42:470–8.

32 Distante C, Piscitelli P, Miani A. Covid-19 outbreak progression in Italian regions: approaching the peak by the end of March in northern Italy and first week of April in southern Italy. *Int J Environ Res Public Health* 2020;17:3025.

33 Richardson DP, Lovegrove JA. Nutritional status of micronutrients as a possible and modifiable risk factor for COVID-19: a UK perspective. *Br J Nutr* 2021;125:1–7.

34 Zabetakis I, Lordan R, Norton C, *et al*. COVID-19: the inflammation link and the role of nutrition in potential mitigation. *Nutrients* 2020;12:1466.

35 Sattar N, Ho FK, Gill JM, *et al*. Bmi and future risk for COVID-19 infection and death across sex, age and ethnicity: preliminary findings from UK Biobank. *Diabetes Metab Syndr* 2020;14:1149–51.

36 Patel AP, Paranjpe MD, Kathiresan NP, *et al*. *Race, socioeconomic deprivation, and hospitalization for COVID-19 in English participants of a national biobank*. *Int J Equity Health* 2020;19:114.

37 Raisi-Estabragh Z, McCracken C, Bethell MS, *et al*. Greater risk of severe COVID-19 in Black, Asian and Minority Ethnic populations is not explained by cardiometabolic, socioeconomic or behavioural factors, or by 25(OH)-vitamin D status: study of 1326 cases from the UK Biobank. *J Public Health* 2020;42:451–60.

38 Adhikari S, Pantaleo NP, Feldman JM, *et al*. Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. *JAMA Netw Open* 2020;3:e2016938.

Supplementary Figure 1. Study flow chart
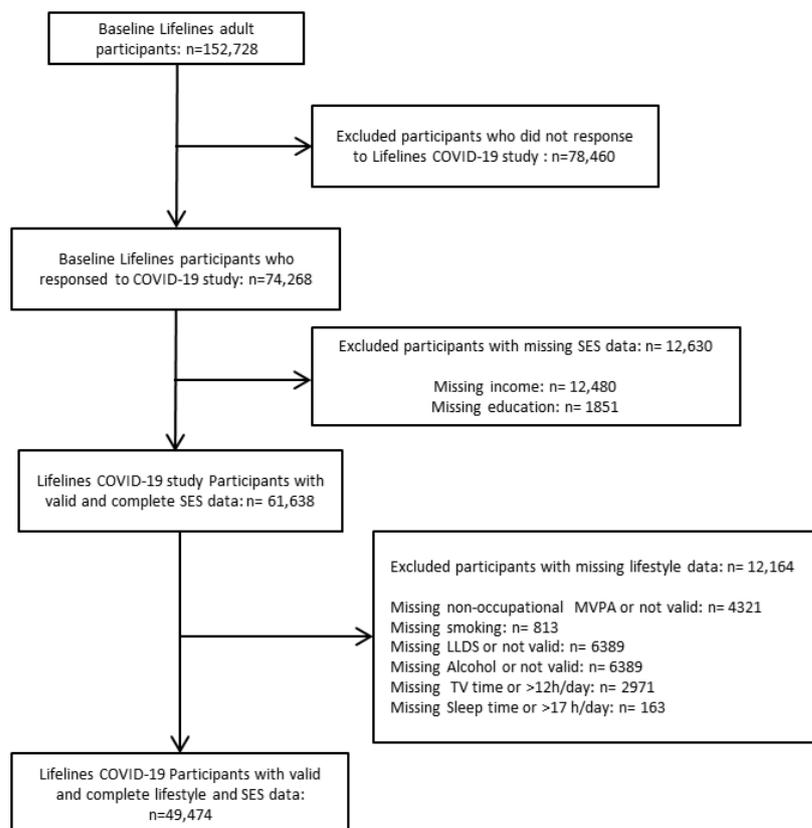
Supplementary Table 1. Association of education or income with COVID-19 status with further adjustment for number of questionnaires filled.

Education	COVID-19 Status					
	Self-reported		Tested		Diagnosed	
	OR [95%CI]	p	OR [95%CI]	p	RR [95%CI]	p
Model 1						
low	0.78 [0.71-0.86]	**<0.001**	0.58 [0.52-0.66]	**<0.001**	1.10 [0.60-2.01]	0.8
middle	0.90 [0.84-0.97]	**0.004**	0.72 [0.66-0.79]	**<0.001**	1.77 [1.21-2.60]	**0.003**
high			Ref			
Model 2						
low	0.79 [0.72-0.87]	**<0.001**	0.61 [0.54-0.68]	**<0.001**	1.08 [0.60-1.95]	0.8
middle	0.91 [0.84-0.97]	**0.006**	0.73 [0.67-0.80]	**<0.001**	1.72 [1.17-2.52]	**0.005**
high			Ref			
Income	Self-reported		Tested		Diagnosed	
	OR [95%CI]	p	OR [95%CI]	p	RR [95%CI]	p
Model 1						
low	0.86 [0.79-0.93]	**<0.001**	0.86 [0.78-0.95]	**0.002**	0.74 [0.47-1.18]	0.2
middle	0.80 [0.74-0.87]	**<0.001**	0.86 [0.78-0.94]	**0.002**	1.05 [0.70-1.57]	0.8
high			Ref			
Model 2						

low	0.85 [0.78-0.92]	**<0.001**	0.84 [0.76-0.92]	**<0.001**	0.78 [0.49-1.24]	0.3
middle	0.80 [0.74-0.86]	**<0.001**	0.85 [0.77-0.93]	**0.001**	1.04 [0.69-1.55]	**0.9**
high			Ref			

*Model 1: adjusted for education or income, age, sex, six lifestyle factors (smoking status, TV watching time ≥ 4h/day, sleep time <7 or >9 h/day, MVPA<150 min/week, LLDS, and heavy drinker), BMI, and ethnicity. Model 2: adjusted for education or income, age, sex, six lifestyle factors (smoking status, TV watching time ≥ 4h/day, sleep time <7 or >9 h/day, MVPA<150 min/week, LLDS, and heavy drinker), BMI, ethnicity, and number of questionnaires filled.

Supplementary Table 2. Characteristics and COVID-19 status of study population and excluded population.

	Study population n= 49,474	Excluded population n= 24,794
Sex, male %	40.3	37.3
Age, mean (sd)	46 ± 12	45 ± 12
BMI, mean (sd)	26.0 ± 4.2	26.0 ± 4.3
Underweight,%	0.7	0.8
Normal,%	44.9	42.0
Overweight,%	40.0	36.9
Obese,%	14.5	14.0
Ethnicity, white %	92.0	81.3
Self-reported covid-19, n (%)	4711 (9.5)	2534 (10.2)
Tested covid-19, n (%)	2883 (5.8)	1284 (5.2)
Diagnosed covid-19, n (%)	123 (0.3)	50 (0.2)

Supplementary Table 3. Mediation effect of lifestyle factors in the association between SES and COVID-19 status.

	COVID-19 Status		
SES indicators	Self-reported	Tested	Diagnosed
Education			
Proportion mediated (%)*	9.9	ns	ns
Income			
Proportion mediated (%)	6.7	9.5	ns

*The proportion mediated was calculated when the mediation effect of lifestyle factors was significant. As shown in the table, lifestyle factors only slightly mediated the association of SES and self-reported, and tested COVID-19 status because the proportion mediated was really small (<10%). This was in accordance with our regression model2 that the coefficient was slighted attenuated after adjustment for lifestyle factors.

Supplementary Table 4. COVID-19 diagnosed cases for the Netherlands as whole and for the Northern Provinces (Lifelines population). Source: RIVM, downloaded from the CoronaWatchNL Github.

Region	Diagnosed (cummulative on 28th of June, 2020)
Netherlands	50,167
North NL	1529
Groningen	362
Frisland	634
Drenthe	533

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Open

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-7
Bias	9	Describe any efforts to address potential sources of bias	11-12
Study size	10	Explain how the study size was arrived at	supplementary
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	7-8
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	7-8

1

Results

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	supplementary
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	supplementary
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	*8*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-9
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	11

Discussion

Key results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10
Generalisability	21	Discuss the generalisability (external validity) of the study results	9

Other information

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

2