PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Patterns of utilisation of specialist care after SARS-Cov-2 infection: a retrospective cohort study
AUTHORS	Formoso, Giulio; Marino, Massimiliano; Formisano, Debora; Grilli, Roberto

VERSION 1 – REVIEW

REVIEWER	Ferguson, Neil Imperial College London
REVIEW RETURNED	17-Jul-2022

GENERAL COMMENTS	This is a nice paper assessing the additional healthcare demand associated with prior COVID diagnosis following initial convalescence. My one substantial concern is about the potential for a correlation between seeking a COVID test and propensity to seek healthcare more generally, which could bias the findings. In addition, given the impact of the pandemic on healthcare generally, there may have been temporal changes in healthcare seeking behaviour over the study period. The authors match on Charlson index, but 90% of cohort subjects had an index of zero. Did the authors compare healthcare utilization before COVID diagnosis in the cases and controls? It would strengthen the paper substantially if they did so, assuming no differences in healthcare utilization were found. Second - more of a request: while the authors use survival analysis to calculate hazard ratios, they don't report on any trends by time since COVID diagnosis. Again, it would be informative to see whether there is any evidence of a decline in hazard ratios over time – e.g. comparing the first 3 months with the next 3 months. Third, I may have missed this, but it wasn't clear how the acute period of COVID disease was treated. Were acute COVID-related hospitalisations included? If not, what time window or other criteria were used to exclude them? Otherwise, I only have minor comments. The text needs careful proof-reading – both for English language usage and minor typos. For instance, the numbers in the text don't match Table 2, and Table 3 uses "," rather than "." as the decimal symbol.

REVIEWER	Telle, Kjetil Norwegian Institute of Public Health
REVIEW RETURNED	14-Aug-2022

GENERAL COMMENTS	The authors describe the utilization of health services in a group that tested positive and then negative, for SARS-CoV-2, and compare their utilization with that of a group that never tested positive. The data are very rich and interesting, and the authors' way of utilizing the longitudinal nature of the data can provide very important insights. Main concern As far as I understand, who are included in the "recovered group" is very simple: First having a positive PCR-test for SARS-CoV-2 and then being tested again with PCR and now the test is negative. I like the simplicity of this, but if I have not misunderstood something essential here, it rises one serious problem of interpretation. People who test positive but recover quickly and fully, and thus do not take a new test, will not be included in the "recovered group". This could be very important for interpreting the results, as it excludes the presumably large group who have no complications after recovery – or, put differently, the "recovered group" tend to comprise exactly the patients with most post-covid complications. Thus, the results cannot illuminate the question of utilization for a random person who gets covid, it will tend, however, to describe utilization for the persons who do not really recover (or catch it again or something else) and thus, likely, tend to use the services after infection. To repeat myself; who are included in the "recovered group" depends on the outcome measure of the analysis: those who utilize the services are more likely to be included in the "recovered group" (since they are more often tested again), and thus, it would be no surprise if the analysis finds that they use more services. To avoid this methodological concern, the authors could instead include everyone who tested positive, and follow their utilization after the positive test (in fact, plotting their utilization both after and before is very informative) irrespective of any subsequent testing. Sub-group analysis could be performed for the patients who were hospitalized in association with the positive tests and those who were not, for example. Whether this change in method affects results would depend on the extent that taking a new test is more common among those with subsequent complications and health care utilization. Also, I presume it will not be that much work for the authors to change their analysis so that the "recovered group" includes those who are not tested again after the positive test (and also include those who test positive again after the first positive test - though for these individuals some may argue that they be censored from the time of the new positive test, but I would not. Alternatively, individuals with a new positive test could be considered a new record in the analysis, given that the time period between the two tests are, say, at least 30 or 60 days to be sure it is a new infection.). Related to this, and provided that I have not misunderstood what the authors have done, it seems to me that the following patient would be included in the "recovered group": testing positive for SARS-CoV-2, being hospitalized for covid-19, put on ventilator, testing negative while still on ventilator, remain hospitalized for the entire follow-up period. I guess there are very few – if any – such cases in the sample, but I still think the way the sample is constructed implies that one should not call this group "recovered". You might find it of interest to look at the method we used in these related papers: - https://www.bmj.com/content/376/bmj-2021-066809

	- https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0257926 - https://www.medrxiv.org/content/10.1101/2021.02.16.21251807v2 - https://www.medrxiv.org/content/10.1101/2022.07.08.22277413v1 Other concerns I have a similar concern about the comparison group. It is not clear to me if the comparison group is constructed from all individuals with at least one PCR-test whose tests were never positive, or from the larger group of any individual in the province who never had a positive PCR-test. Only if everyone in the province have been tested at least once, these two groups of individuals would be the same. Again, my concern is related to those being tested tending to do so because they feel unwell (or because they are to be admitted to a health care institutions). In my opinion there are at least two interesting comparison groups here: - Those who take a test that is negative on the same week (or day?) as a person in the "recovered group" (this could be achieved by including the test week as a matching variable) - Those who do not take a test on the same week as a person in the "recovered group" (presumably in better health than the individuals who take a test) Please clarify how those who die are handled (they should be censored from time of death and not dropped entirely, but what is done in the analysis is not completely clear to me). I'm not sure I understand why you use the episodes as numerator and time as denominator (wouldn't it be easier to interpret if the numerator and denominator were more similar/the same, like weeks with episode over total weeks)? Also, wouldn't it be easier to interpret if you used costs in monetary units (euro), possibly in log, instead of quartiles (and then you could also simplify by using a linear model)? The research question invites using a Kaplan-Meier plot (or similarly rates per week), such a plot would be very informative for a wide range of readers. Is the figures in the Costs sections in euro or number of consultations? Am I correct in thinking that the estimated extra costs for those with covid are extremely low (2.2 million euro)? Maybe so low that the cost-analysis could be downplayed considerably in the paper? I do not see that the stratification by Charlson is adding sufficient insight to be included in the paper. Stratification on vaccine is very interesting! This should be expanded, or, maybe better, done more carefully in a separate paper (i.e. drop it from this paper and note in discussion that it is an important avenue for further research in future?). Are utilization of primary care/general physicians included in the data? The data on drug prescription is VERY interesting. And I encourage the authors to do an additional paper focusing on this. To be able to follow utilization (though you only have prescriptions, that is still a very good indicator of utilization) of drugs before and after covid-19 for the whole population in the region would add substantially to our knowledge of how covid-19 affects health and health-care needs/use. Especially so if the

	authors can also break down the analysis to some drugs of particular interest, like pain killers, (oral) antiviral drugs, sleep-related drugs, and probably other too (I'm not a medical doctor/clinician). In my opinion, a well-crafted such analysis should be of interest to a very good general medical journal, like the BMJ. I wish the authors the best with the important ongoing work with these very good data, Kjetil Telle Director of Health Services Research Norwegian Institute of Public Health

VERSION 1 – AUTHOR RESPONSE

Reviewers' comments	**Authors' replies**
Ferguson	
My one substantial concern is about the potential for a correlation between seeking a COVID test and propensity to seek healthcare more generally, which could bias the findings … The authors match on Charlson index, but 90% of cohort subjects had an index of zero. Did the authors compare healthcare utilization before COVID diagnosis in the cases and controls? It would strengthen the paper substantially if they did so, assuming no differences in healthcare utilization were found	We agree. Our hypothesis is that it was the covid that brought the positives to a subsequent greater use of the services, but we cannot assume (only through the adjustment for age, sex and Charlson Index) that they were comparable even before: those who do not have a positive test may be more careful in lifestyles (and more likely to be able to avoid covid, as well as be more likely to be visited) or vice versa avoid tampons and visits. Therefore, we did an additional analysis (available in Table 4) confirming that the two cohorts did not differ in terms of the more frequently reported ambulatory visits, so that their health-seeking behavior could be considered comparable. We integrated the discussion accordingly.
while the authors use survival analysis to calculate hazard ratios, they don't report on any trends by time since COVID diagnosis. Again, it would be informative to see whether there is any evidence of a decline in hazard ratios over time – e.g. comparing the first 3 months with the next 3 months	Thank you. We performed additional analyses according to this suggestion, also considering either controls who had a swab test and those who never had one, according to suggestions from the other reviewer (see Figure 1)
it wasn't clear how the acute period of COVID disease was treated. Were acute COVID-related hospitalisations included? If not, what time window or other criteria were used to exclude them?	We did not use a time window but started the follow-up from the negative test. The assumption is that those who have been negativized cannot be in an acute phase (or that it is unlikely to do a further test during the acute phase to check if one has been negativized). In the discussion we now acknowledge a risk of misclassification, although we consider it very small and not higher than the risk of missing cases which could occur within a time window
The text needs careful proof-reading – both for English language usage and minor typos. For instance, the numbers in the text don't match Table	Thanks, amended

2, and Table 3 uses "," rather than "." as the decimal symbol	
Telle	
People who test positive but recover quickly and fully, and thus do not take a new test, will not be included in the "recovered group"… To avoid this methodological concern, the authors could instead include everyone who tested positive, and follow their utilization after the positive test (in fact, plotting their utilization both after and before is very informative) irrespective of any subsequent testing …	Thank you. We now specify that in Italy an exit test was mandatory and that a risk of misclassification is extremely low in this regard. In any case, all people in the "positive test" cohort have been included if they did have a subsequent negative test
it seems to me that the following patient would be included in the "recovered group": testing positive for SARS-CoV-2, being hospitalized for covid-19, put on ventilator, testing negative while still on ventilator, remain hospitalized for the entire follow-up period. I guess there are very few – if any – such cases in the sample, but I still think the way the sample is constructed implies that one should not call this group "recovered". You might find it of interest to look at the method we used in these related papers …	Thanks for highlighting that. Our assumption is that those who have been negativized cannot be in an acute phase (or that it is unlikely to do a further test during the acute phase to check if one has been negativized). In the discussion we now acknowledge a risk of misclassification, although we consider it very small and not higher than the risk of missing cases which could occur within a time window. Thanks also for the references! We used a couple of them to comment on our results
It is not clear to me if the comparison group is constructed from all individuals with at least one PCR-test whose tests were never positive, or from the larger group of any individual in the province who never had a positive PCR-test. Only if everyone in the province have been tested at least once, these two groups of individuals would be the same. Again, my concern is related to those being tested tending to do so because they feel unwell (or because they are to be admitted to a health care institutions). In my opinion there are at least two interesting comparison groups here: - Those who take a test that is negative on the same week (or day?) as a person in the "recovered group" (this could be achieved by including the test week as a matching variable) - Those who do not take a test on the same week as a person in the "recovered group" (presumably in better health than the individuals who take a test)	Thank you, we agree and added analyses related to the two different comparison groups (Figure 1, also showing a decline in HRs after 90 days). These analyses show that HRs related to people who never tested are higher than HRs related to people with at least one negative test. The latter may be at higher risk for clinical sequelae and this may be the reason why they are more likely to be tested. This hypothesis is also supported by a logistic model we performed using the subjects in the control cohort (see Table 7): those with higher Charlson index are more likely to have been tested. At the same time younger people are more likely to have been tested (they are more likely to be socially involved), while older people are less (they are more likely to have been kept isolated in those months)
Please clarify how those who die are handled (they should be censored from time of death and not dropped entirely	Thank you. Yes, they were censored from time of death (we now specify it in the methods section)
I'm not sure I understand why you use the episodes as numerator and time as denominator (wouldn't it be easier to interpret if the numerator and denominator were more similar/the same, like weeks with episode over total weeks)? Also, wouldn't it be easier to interpret if you used costs in monetary units (euro),	Thank you, but we prefer to keep events per person/months since we prefer to focus on how many events occur. As for using a linear model to assess factors associated to costs: we acknowledge that would be a possibility, assuming a linear relationship with the various

possibly in log, instead of quartiles (and then you could also simplify by using a linear model)?	covariates considered, although we wanted to focus on factors associated with having particularly high care costs
The research question invites using a Kaplan-Meier plot (or similarly rates per week), such a plot would be very informative for a wide range of readers	Thank you for the suggestion, but we prefer not to use it, focusing on estimating the risk of whatever event in each cohort (and taking the possibility of repeated events into account).
Is the figures in the Costs sections in euro or number of consultations? Am I correct in thinking that the estimated extra costs for those with covid are extremely low (2.2 million euro)? Maybe so low that the cost-analysis could be downplayed considerably in the paper?	We actually prefer to keep that analysis: for the budget of a local health authority in a province with half a million inhabitants, even 2.2 million euros of difference in such a period of time may be important.
I do not see that the stratification by Charlson is adding sufficient insight to be included in the paper	Thanks. We prefer to keep it since it quantifies the possible cost burden of long-covid and highlights that people with comorbidities are those most affected
Stratification on vaccine is very interesting! This should be expanded, or, maybe better, done more carefully in a separate paper (i.e. drop it from this paper and note in discussion that it is an important avenue for further research in future?).	Thank you. We agree and now specify in the discussion that we plan to expand this analysis in a further paper. We already acknowledged that using longer follow-up periods with higher numbers of vaccinated people would allow a comparison between those who developed COVID and those who do not, warranting the inclusion of boosted people who could not be included in our cohort yet
Are utilization of primary care/general physicians included in the data?	No, unfortunately they are not because they are not traceable. We now acknowledge it in the discussion and changed the paper title accordingly
The data on drug prescription is VERY interesting. And I encourage the authors to do an additional paper focusing on this. To be able to follow utilization (though you only have prescriptions, that is still a very good indicator of utilization) of drugs before and after covid-19 for the whole population in the region would add substantially to our knowledge of how covid-19 affects health and health-care needs/use. Especially so if the authors can also break down the analysis to some drugs of particular interest	Actually, we did not carry out a specific analysis on the use of drugs (the related expenditure contributes to the overall expenditure). Here too, we will follow the suggestion to make a separate paper (thank you!)

VERSION 2 – REVIEW

REVIEWER	Ferguson, Neil Imperial College London
REVIEW RETURNED	19-Nov-2022

GENERAL COMMENTS	I am satisfied that the issues raised in my original review have been adequately addressed.

REVIEWER	Telle, Kjetil Norwegian Institute of Public Health

REVIEW RETURNED	27-Nov-2022

GENERAL COMMENTS	Review report on bmjopen-2022-063493.R1: Patterns of utilization of specialist care after SARS-CoV-2 infection: a retrospective cohort study In the revision, the authors have clarified several aspects of their analysis. The retrospective design still raises concerns about selection (eg. testing) with associated difficulties in interpreting the results. Moreover, I do not find that the authors state the limitations of their analysis sufficiently clear in the discussion, and especially not since they make several statements alluding to causal questions their retrospective design cannot illuminate, like that they "assess the impact of SARS-CoV-2 infection on use of specific specialist care" (start of Discussion) or that the "findings also reflect the burden of post-covid sequelae" (towards end of Discussion). It is very important that the discussion of limitations of the study are stated much clearer. You state that the exit test was mandatory, but you should use your data to show this: What is the number of inhabitants testing positive who actually had taken a new test (that was negative) following the positive test by eg. one week, one month, three months (or better: a Kaplan-Maier plot)? I would find it very informative, and important for interpretations, if you provided separate results (stratified analysis) for Covid convalescents who had and had not been hospitalized for covid – a variable you do seem to have (cf. Table 7). I find this to be far more important than the separate results for Charlson=0 and Charlson>=1 (e.g. table 2). The "placebo-test" for selection in Table 5 showing the outpatient specialist visits in the year before the infection (instead of after as in the main analysis) is a very nice idea. However, it should be done in exactly the same way as the main analysis (HR column with CI in Tabel 2; i.e. you should run the same regression with the same controls except Charlson since it is constructed using these "placebo-outcomes"), and definitely for all the outcomes used in the main analysis (not only for the four included in Table 5). Table 3 is also nice, but I do not understand why results in the last column ("Overall") is not identical to the corresponding results in Table 2? Also you do need to include the number of observations in the various groups in the table. In the discussion you seem to state that these results (from Table 3) show that those in control group with at least one negative test "may be of higher risk of clinical sequelae" – which I do not understand since none of the individuals in the control group could have covid-sequelae (they did not have covid)? Why did you not include deaths in Table 1 (and 2)? Please use the same term throughout when you mean the same ting: I presume "Non surgical h admissions" in Table 2 is the same variable as "H admissions" in Table 3 (if not, explain why). I the data statement you say that "Data will be made available upon reasonable request" – does this mean that the authors are legally allowed to decide to send these individual-level records to researchers elsewhere in the world? Might it not be possible with

	all this individual level information to backward-identify some of the individuals in the dataset?

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2 - Dr. Kjetil Telle	**Authors' replies**
In the revision, the authors have clarified several aspects of their analysis. The retrospective design still raises concerns about selection (eg. testing) with associated difficulties in interpreting the results. Moreover, I do not find that the authors state the limitations of their analysis sufficiently clear in the discussion, and especially not since they make several statements alluding to causal questions their retrospective design cannot illuminate, like that they "assess the impact of SARS-CoV-2 infection on use of specific specialist care" (start of Discussion) or that the "findings also reflect the burden of post-covid sequelae" (towards end of Discussion). It is very important that the discussion of limitations of the study are stated much clearer.	We agree that our retrospective design raises concerns, that we should carefully avoid statements implying causal relations and that study limitations should be examined more closely. In order to better support our findings, we added sensitivity and subgroup analyses to check how previous hospital and ER admissions and time to recovery can influence the likelihood of post-covid use of health services. Comments on study limitations have been added accordingly.
You state that the exit test was mandatory, but you should use your data to show this: What is the number of inhabitants testing positive who actually had taken a new test (that was negative) following the positive test by eg. one week, one month, three months (or better: a Kaplan-Maier plot)?	Thank you for helping us to be more specific about this issue. The exit test was actually mandatory within 21 days from testing positive. In case a person still had a positive test after 21 days, end of isolation was also allowed regardless of a positive test if people were asymptomatic. Specifically, 77% had a negative exit test within 21 days and 23% were asymptomatic and allowed to exit isolation after the 21st day. The risk of having "non-recovered" positives in the COVID database is extremely low: a surveillance system with daily phone calls and interviews with all cases cared for in outpatient settings was into place. We added a subgroup analysis (in the new table 5) to check whether HRs and CIs for each outcome were different by time to recovery quartiles and amended the discussion on this issue
I would find it very informative, and important for interpretations, if you provided separate results (stratified analysis) for Covid convalescents who had and had not been hospitalized for covid – a variable you do seem to have (cf. Table 7). I find this to be far more important than the separate results for Charlson=0 and Charlson>=1 (e.g.	Thank you. Added these results in table 2

table 2).	
The “placebo-test” for selection in Table 5 showing the outpatient specialist visits in the year before the infection (instead of after as in the main analysis) is a very nice idea. However, it should be done in exactly the same way as the main analysis (HR column with CI in Tabel 2; i.e. you should run the same regression with the same controls except Charlson since it is constructed using these “placebo-outcomes”), and definitely for all the outcomes used in the main analysis (not only for the four included in Table 5).	Thank you, this comment is extremely important and led us to perform a sensitivity analysis which may help enlighten our conclusions. Although adjustment for Charlson index should help to limit risks of confounding, we initially did the “placebo test” as requested (from the other reviewer) to strengthen the hypothesis of an association between covid and subsequent encounters. We had focused on outpatient visits, which we consider particularly relevant in terms of long-covid sequelae. Following your suggestion, we checked for possible imbalances in each outcome, and found that three of them were imbalanced at baseline: non-surgical hospital admissions, admissions for respiratory disease (a numerically relevant part of non-surgical admissions) and accesses to emergency room. This was somehow expected, considering that respiratory patients may be at higher risk of getting covid (see here). We provide a separate table with these data for your information (see next page), that we would not include in the paper (to avoid presenting too many data) but partly cite in the text. We then performed a sensitivity analysis to check whether conditions leading to recent hospital admissions and/or accesses to emergency room could influence the results. This sensitivity analysis and the previously mentioned subgroup analysis (both in the new table 5) could help the interpretations of results. We amended the discussion accordingly.
Table 3 is also nice, but I do not understand why results in the last column (“Overall”) is not identical to the corresponding results in Table 2? Also you do need to include the number of observations in the various groups in the table. In the discussion you seem to state that these results (from Table 3) show that those in control group with at least one negative test “may be of higher risk of clinical sequelae” – which I do not understand since none of the individuals in the control group could have covid-sequelae (they did not have covid)?	Thank you. The mismatch was due to erroneously including all hospital admissions (and not just non-surgical hospital admissions) in table 3. Now data in table 2 and 3 are congruent (although we avoid repeating overall HRs in table 3). We now include all the outcomes in table 3 and the size of the two groups. We also amended the discussion (we were actually not referring to covid sequelae but to clinical problems in general – sicker people in the control group may be more likely to be tested)
Why did you not include deaths in Table 1 (and 2)? Please use the same term throughout when you mean the same ting: I presume “Non surgical h admissions” in Table 2 is the same variable as “H admissions” in Table 3 (if not, explain why).	Thank you. Amended as requested (see previous comment)

I the data statement you say that “Data will be made available upon reasonable request” – does this mean that the authors are legally allowed to decide to send these individual-level records to researchers elsewhere in the world? Might it not be possible with all this individual level information to backward-identify some of the individuals in the dataset?	According to Italian law, if there is potential for the re-identification of individuals, only anonymized data can be made publicly available (https://www.garanteprivacy.it). Thus, the data underlying this study are available on request to researchers who meet the criteria for access to confidential data. In order to obtain data, approval must be obtained from the Area Vasta Emilia Nord (AVEN) Ethics Committee, who would then authorize us to provide aggregated or anonymized data. Data access requests should be addressed to the Ethics Committee at CEReggioemilia@ausl.re.it as well as to the corresponding author. We tried to make these points clearer in the data availability statement.

Table. % of hospital, emergency room and outpatient visits accesses for convalescent covid-19 patients and the control group in the year preceding the index date

	SARS-CoV-2 cohort	Control cohort
Non surgical h admissions	8.8%	3.1%
H admissions for respiratory disease	5.0%	0.2%
H admissions for heart disease	0.5%	0.4%
Accesses to Emergency Room	23,7%	14.1%
Pneumology	1.6%	1.6%
Cardiology	5.3%	5.5%
Neurology	3.3%	3.3%
Rheumatology	1.8%	1.9%
Gastroenterology	1.1%	1.2%
Mental health	0.7%	0.9%
Dermatology	7.1%	7.5%
Endocrinology	5.1%	4.7%