Dear Dr. Nyberg,

Thank you for sending us your paper. We sent it for external peer review and discussed it at our manuscript committee meeting. We are interested in proceeding with it provided you are willing and able to revise your paper as explained below in the report from the manuscript meeting, so that we can make a final decision.

Please remember that the author list and order were finalised upon initial submission, and reviewers and editors judged the paper in light of this information, particularly regarding any competing interests. If authors are later added to a paper this process is subverted. In that case, we reserve the right to rescind any previous decision or return the paper to the review process. Please also remember that we reserve the right to require formation of an authorship group when there are a large number of authors.

When you return your revised manuscript, please note that The BMJ requires an ORCID iD for corresponding authors of all research articles. If you do not have an ORCID iD, registration is free and takes a matter of seconds.

Sincerely,
Elizabeth Loder, MD, MPH

*** PLEASE NOTE: This is a two-step process. After clicking on the link, you will be directed to a webpage to confirm. ***

https://mc.manuscriptcentral.com/bmj?URL_MASK=5959aad4f37944ac8d23dd9c6e6609dc

**Report from The BMJ’s manuscript committee meeting**

These comments are an attempt to summarise the discussions at the manuscript meeting. They are not an exact transcript.

Present: Helen Macdonald (chair); Jamie Kirkham (statistician); Elizabeth Loder; Di Wang; Wim Weber; John Fletcher; David Ludwig; Tiago Villanueva

Decision: Request revisions before final decision

* You provide many analyses to a battery of different study designs. Our statistician wonders if this is necessary. The results are pretty consistent - adjusted analyses seem to suggest the likelihood of hospitalisation is 50-60% higher for patients infected with a SGTF-associated variant. It’s encouraging that all analyses show similar results but we think you need to specify the primary analysis and provide more rationale on why other approaches were used. We think the stats methods section could be more detailed on the chosen primary analysis at least.

* Why was age categorised? Absolute risks?

* We thought the topic of a currently widespread VOC of covid-19, though identified a couple of months ago, is still of interest for a wide readership. But a major concern is very limited confounding control for
the primary analysis, e.g BMI, underlying conditions like immunosuppression\hypertension\diabetes\renal disease\chronic lung disease, etc. Can you comment?

* Please provide a rationale for the choice of Nov 23,2020~Jan 4,2021(which looks just before the start of wide vaccination program in UK); also why choose hospitalization within 14 days after a positive test?

* We previously published another UK study based on "Pillar 2" community test 1 Oct to 29 Jan. This manuscript is based on Pillar 2 community tests 23 Nov to 4 Jan so there is a lot of overlap in the people in the study and the main difference is the outcome reported; admissions vs deaths. We'd like you to include mortality results in this paper, while making clear that there is overlap with the population reported on in our previous paper.

(Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study

By: Challen, Robert; Brooks-Pollock, Ellen; Read, Jonathan M.; et al.

BMJ-BRITISH MEDICAL JOURNAL Volume: 372 Article Number: n579 Published: MAR 10 2021)

* Might you also provide information on ICU admittance?

* Please also lengthen the sample beyond Jan 4 if at all possible.

* Not all editors were convinced about your argument that hospitalisation as an outcome is less likely to be confounded than mortality. Might you strengthen this?

* Do you have information on whether positively tested patients knew what variant they had ? As the media had many stories about increased severity of the UK variant, this knowledge might in itself lead to increased health-care seeking behavior.

* Can you more clearly explain what Pillar 1 and 2 testing is for non-UK readers ?

In your response please provide, point by point, your replies to the comments made by the reviewers and the editors, explaining how you have dealt with them in the paper.

Comments from Reviewers

Reviewer: 1

Comments:

Nyberg, Twohig, Harris, Seaman, Flannagan, Allen, Charlett, De Angelis, Dabrera, Presanis

Reviewer: Ruth Keogh, LSHTM

Summary

Thank you for the opportunity to review this paper. The paper estimates the association between SARS-CoV-2 variant B.1.1.7 and risk of hospitalisation within 14 days of a positive test. Several studies have found the new variant B.1.1.7 to be associated with an increased risk of mortality, but the evidence on its impact on hospitalisation is more limited. This study uses Covid-19 community testing data from tests performed in Lighthouse laboratories, which are able to identify S-gene target failure (SGTF), a proxy for the new variant B.1.1.7. The test data are linked to hospitalisation records from two sources. The research question is important and the authors have access to data suitable for addressing the question robustly, which I think they have done using the analyses performed (some clarifications are needed – see below). The paper conveys the overall message clearly. The description of the data used is particularly clear. I have some reasonably major comments about the methods used. The preprint of Patone et al, who looked at the association between this variant and critical care admission, should be referred to.
In this study, several analyses are performed using the same data, based on different study designs within the cohort. My overall feeling is that the several study designs/analyses considered are not needed, and my suggestion is to consolidate the analyses into one. Please see below for more detailed comments on the different analyses. A discussion of using different study design/analyses in this context could be interesting, but I feel that would be more appropriate for an epidemiological journal, and a more detailed discussion of the relative merits of the different designs and analyses would be needed.

The paper focuses on estimation of odds ratios, i.e. relative associations. I would like to also see some information added about how the findings translate into absolute risks and numbers hospitalised. Can this study provide estimates of how many additional hospitalisations resulted in over the study period?

Other major comments are (with more details below):
- Information is needed about how deaths among individuals not hospitalised were treated in some of the analyses. Information is also needed about where the data on deaths comes from.
- I found some of the comments about confounding unclear, and was not convinced of the arguments that this study better controls for confounding than some earlier studies.

Detailed comments

Abstract
In the abstract the Design is stated as being “Retrospective cohort, analysed using a matched nested case-control study design.”. This doesn’t reflect that several types of sub-study are considered.
Main outcome measures: “first positive SARS-CoV-2 test.” Do you know it is the first test?

“What is already known on this topic”

Second bullet point – I am not sure this is a fair statement. Please see comments on the Discussion section. Please clarify what you mean by “confounding due to increased hospital burden”. What is the confounder?

Third bullet point: less likely than what?

“What this study adds”

First bullet point: “hospitalisation within 14 days of a positive test” (not just “ever” hospitalised)

1. Background

I suggest adding a reference to the preprint of Patone et al (https://www.medrxiv.org/content/10.1101/2021.03.11.21253364v1), who investigated the association between B.1.1.7 and admission to critical care.

At the end of the background section, the aim is stated as being an investigation of the association between SGTF and hospitalisation. I think you could be stronger than this and state that your aim was to investigate whether there is a causal relationship, which I think is true. An “association” could be purely descriptive.

2. Methods

Please provide some justification for the use of test data from 23 Nov to 4 January. In particular, why did you start from 23 Nov?

Given that hospitalisations are not captured until complete, how can you be sure that you captured all hospitalisations. Perhaps this was the reason for restricting to tests up to 4th Jan?
Were the hospitalisations for any reason or were you able to restrict to hospitalisations due to Covid-19. If you were not able to restrict to Covid-19 hospitalisations please provide some discussion of how this may have impacted your results.

Please provide some information about why you focus on hospitalisations up to 14 days after the test. Why not longer?

I would like to see a statement near the start of the statistical methods section about what it is you are aiming to estimate, and stating that you wish to control for confounding by a number of variables. Please provide justification for the choice of adjustment/matching variables

The primary method of analysis is described as a matched nested case-control analysis. To my understanding, a nested case-control study typically describes a case-control study within a cohort where incidence density sampling has been used. Here you have used standard matched case-control study within the cohort. This is fine, but I think the term “nested” should not be used. The approach is described as “assumption-free”. This is too strong a statement – you have made assumptions, for example, by categorising variables to perform the matching. If your aim is to estimate a causal effect then you also make an assumption that you have adjusted for all confounders.

I found it a little surprising that you also considered a stratified cohort analysis, an “exploratory” cohort analysis and a matched cohort analysis, alongside the matched case-control analysis. I agree it can sometimes be reassuring to see the same result from different study designs, but you did not provide reasoning behind this decision, and I didn’t feel that it was justified here. If you wish to retain the different approaches then I would like to see some justification as to why you consider the different approaches, e.g. are they making different assumptions. I feel it would be more sensible to consider which approach best answers your question and to focus on that. My preference would be a stratified Cox regression used to estimate cumulative incidences, and possibly cause-specific hazard ratios, possibly with regression adjustment for certain covariates that result in uninformative strata. Some specific comments on the different approaches used are given below.

You do not provide any information about how individuals who die without being hospitalised are treated, except for in the description of the Cox regression analysis. Please provide this information, and explain how it influences the interpretation of the odds ratios for hospitalisation.

2.4.1 Matched nested case-control analysis:

Why did you only match each case to one control, when there must often be many potential controls per case? This is inefficient. Were controls reused, or were they sampled without replacement?

Age and IMD were categorised for the matching. There may be residual confounding and you could have additionally adjusted for ‘exact’ age and IMD score in the conditional logistic regression, which is common in analyses of matched studies.

2.4.2 Stratified cohort analysis:

The wording suggests that in a matched case-control study the number of controls per case must be the same - this is not necessary. The stratified cohort analysis makes more efficient use of the data, and does not appear to make any additional assumptions relative to the matched case-control analysis. Though it may be more computationally intensive, especially if some of the strata are large and contain more than one case. It is not clear to me why you would present the matched case-control analysis in addition to the stratified cohort analysis.

2.4.3 Exploratory cohort analysis:

Here an unconditional logistic regression analysis is used, with regression-adjustment for covariates instead of matching. How was age entered into the model – as a linear term? Non-linear relationships should also be explored. Based on the results presented in section 3 it appears that age was categorised into 10-year age groups. It would be preferable not to categorize age in this way, to avoid unnecessary loss of information. Here you investigated interactions between SGTF and covariates. However, this could also have been done in the previous two analyses. While matched analyses preclude estimation of main effects for the
matching/stratification variables, one can still estimate interactions between matching/stratifying variables and the exposure. It seems odd that you only look at interactions in this unmatched analysis.

Cox regression analyses:
You state: "Individuals who died prior to hospitalisation were censored in the main analysis; we assessed the potential impact of mortality as a competing risk using the Fine and Gray sub-distribution hazard model.". It is unclear to me what you did based on this wording. Did you perform one Cox regression analysis in which individuals who died without being hospitalised were censored at death, and one in which individuals who die are left in the risk set (Fine and Gray method).
I do not think it is really appropriate to say that the second allows an assessment of the "impact of mortality" – the two approaches estimate different quantities.
This is the only analysis in which hospitalisation beyond 14 days was considered. Why was this not considered in the other approaches?
Please give some information about where the data on mortality comes from, as this is not currently mentioned in the subsections giving details about the data.

2.4.4 Matched cohort analysis:
Please explain the justification for this analysis. Matched cohort studies are most appropriate I believe when the exposure is rare. That is not really the situation here, as the outcome is much rarer.
Why was age categorised into 10-year groups here, compared with 5-year age groups in the other analyses.
It is not stated what analysis method was used here. I assume an unconditional unadjusted logistic regression? However, the results suggest that some regression adjustment was performed.

3. Results

In Table 1 I would also like to see a description of the whole cohort, as well as separated by SGTF status.

I notice that the first time the term "confounding" is used is in section 3.2.3. Please see earlier comments about being clear from the start about the aims of the study.

Section 3.2.3

Table 2 shows odds ratio estimates from univariable and multivariable analyses. Since your focus (I think) is on the OR for SGTF I would avoid showing the ORs for the other covariates (certainly in the main text), which are being used to adjust for confounding, rather than being of interest in themselves. This tables invites overinterpretation and the "table 2 fallacy".

For the time-to-event analyses you present both cause-specific and sub-distribution HRs. These are presented as though they have the same interpretation, which they do not. Although in this case the estimates are almost identical. You state that there were 345 deaths within 14 days of the test – is this referring to all deaths, or deaths that occurred without the person going to hospital?

It appears that you investigated the proportional hazard assumption (bottom of page 10) – this was not mentioned in the methods section. Please also comment on whether there was evidence for time-varying effects based on a formal test.

Section 3.3.4

Here it became a bit clearer what analysis you performed for the matched cohort study. Are the unadjusted and adjusted analyses estimating different quantities in this setting of a matched cohort study? Please provide clear interpretations (though see other comments about choosing one study design)

4. Discussion
You state that the stratified or matched cohort analyses used a smaller subset of the hospital admissions than the nested case-control analysis. This appears to contradict what was written in the methods section. It is unclear from the results section whether more of the hospitalisations are used in the matched case-control study or in the stratified cohort study. I would expect the stratified cohort analysis to use more of the overall data than the 1:1 matched case-control analysis, and the same number of cases (if the matching/stratification variables are the same). Please clarify.

4th para of discussion. If you retain the matched cohort analysis, I would like to see some more discussion of why it gives higher OR estimates compared to the other analyses. Is it just due to the same size? Or is it in part due to the types of individual for whom it was easiest to find a match, in combination with the presence of covariate-by-exposure interactions?

5th para of discussion. I don’t follow the reasoning around confounding given here. It is stated that earlier studies did not adjust for hospital over-burden. Do you think hospital over-burden in itself is a confounder here? How is it associated with the exposure?

Related to the above comment, I don’t understand this sentence: “hospital over-burden is unlikely to positively confound an association between SGTF status and hospital admission, because it is unlikely that higher hospital over-burden results in increased admission rates for diagnosed individuals”. Please could you provide further explanation. What about negatively confound? How are you imagining hospital burden to be associated with SGTF status? What do you mean by “diagnosed individuals” – diagnosed with Covid-19, or exposed with SGTF?

Top of page 12. You write: “In the absence of this potential confounder in previous analyses, our results corroborate the hypothesis that the B.1.1.7 variant is associated with more severe disease than wildtype variants.”. Which confounder is being referred to here – I assume hospital burden, but please see above comments. When you say “In the absence of this potential confounder” are you referring to the variable not having been controlled for in the previous analyses, or to the variable not actually having been a confounder?

You restricted to pillar 2 tests with known SGTF status, due to having been analysed in one of the TaqPath Lighthouse laboratories. Please comment on assumptions made when you exclude those tests not performed in one of these laboratories, and the any impact this may have on your analysis.

Figure A.1. This figure shows cumulative hazards by 10-year age group. The legend should make clear whether the lines refer to age groups or to specific ages. E.g. is age10=0 referring to age=0 or to age group 0-9? I think it would also be more appropriate and informative to present cumulative incidences instead of cumulative hazards. I don’t think these estimates were mentioned in the methods section. In general I would like to see more clarity about what message you’re trying to convey with this plot. Perhaps the plot would be useful to include earlier in the results section as part of your justification for focusing on hospitalisation within 14 days of the test.

Additional Questions:

The BMJ uses compulsory open peer review. Your name and institution will be included with your comments when they are sent to the authors. If the manuscript is accepted, your review, name and institution will be published alongside the article.

If this manuscript is rejected from The BMJ, it may be transferred to another BMJ journal along with your reviewer comments. If the article is selected for publication in another BMJ journal, depending on the editorial policy of the journal your review may also be published. You will be contacted for your permission before this happens.
For more information, please see our peer review terms and conditions.

Please confirm that you understand and consent to the above terms and conditions:
I consent to the publication of this review

Please enter your name: Ruth Keogh
Job Title: Professor of Biostatistics and Epidemiology
Institution: London School of Hygiene & Tropical Medicine
Reimbursement for attending a symposium?: No
A fee for speaking?: No
A fee for organising education?: No
Funds for research?: No
Funds for a member of staff?: No
Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No
Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests please declare them here:

BMJ are working with ORCID to recognise the importance of the reviewer community. Reviewers are now able to share their activity by connecting their review to their ORCID account to gain recognition for their contributions.

Only the Journal title will be uploaded into the reviewer’s ORCID record, along with the date the record was uploaded; there is no identification of the article’s title or authors. Records are uploaded once a decision (accept, reject, or revision) has been made on the article.

Would you like to be accredited by ORCID for this review?: Yes
Reviewer: 2
Comments:
1. The paper by Nuyber et al is very interesting since the B.1.1.7 variant has become rapidly the most prevalent in many countries. The study is well conducted and, even if SGTF is not as accurate as genome sequencing, data on such a large population are undoubtedly relevant. However, in my opinion, the interpretation of the results is somewhat misleading and issues around potential residual confounding have been too easily dismissed. In general, the paper may benefit from a revision by an infectious disease specialist.

Main points

2. The authors’ conclusion that “The higher severity may be specific to older age groups” seems misleading from looking at the data. Indeed, if it is true that people of approximately 20 years of age do not show an increase risk of hospitalization but the relative risk starts to increase to 1.67 at 40 years and it remains pretty stable after that age. The fact that even people in their 40th, 50th and 60th are at higher risk of hospitalization is extremely important because: i) this was not seen during the first wave ii) these ages meet the criteria for invasive mechanical ventilation and the ICU occupancy is a major problem during this pandemic; iii) the emotional impact on health care workers of dealing with younger patients is certainly greater; iv) there is a greater socio-economic impact of admitting to hospital actively working persons. Also, I suggest to include in Table 3 the odds in the age strata in the two exposure groups to see the absolute risk by age group.

3. Residual confounding is likely to be an issue and retrospective case-control studies are particularly prone to this kind of bias. In particular, the authors completely dismiss residual confounding due to the lack of controlling for comorbidities and obesity. Indeed, they claim that that possible backdoor confounding pathway is already blocked by controlling for age, sex, ethnicity and deprivation. In reality, the risk of SARS Cov-2 acquisition largely depends on comorbidities especially those causing immunosuppression (cancer, HIV, chemotherapy etc). Thus, it is possible that SGTF variants were more likely transmitted to this vulnerable population who also have a greater intrinsic risk of hospitalization. People with cancer are generally registered with GP databases (for medical exemptions), unclear why at least a linkage with these data was not possible for this analysis. Because OR are not collapsible (and therefore OR may change from unadjusted to adjusted analysis for this reason other than confounding) it would be interesting also to see the risk and the risk difference calculated.

Additional Questions:
<strong>em>The BMJ</em> uses compulsory open peer review. Your name and institution will be included with your comments when they are sent to the authors. If the manuscript is accepted, your review, name and institution will be published alongside the article.</strong>

If this manuscript is rejected from <em>The BMJ</em>, it may be transferred to another BMJ journal along with your reviewer comments. If the article is selected for publication in another BMJ journal, depending on the editorial policy of the journal your review may also be published. You will be contacted for your permission before this happens.

For more information, please see our <a href="https://www.bmj.com/about-bmj/resources-reviewers" target="_blank">peer review terms and conditions</a>. 
<strong>Please confirm that you understand and consent to the above terms and conditions.</strong>: I consent to the publication of this review

Please enter your name: Cristina Mussini

Job Title: Professor of Infectious Diseases

Institution: University of Modena and Reggio Emilia

Reimbursement for attending a symposium?: No

A fee for speaking?: Yes

A fee for organising education?: No

Funds for research?: Yes

Funds for a member of staff?: Yes

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <a href="http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests" target="_new">please see BMJ policy</a> please declare them here: None

<em>BMJ are working with <a href="https://orcid.org/" target="_blank">ORCID</a> to recognise the importance of the reviewer community. Reviewers are now able to share their activity by connecting their review to their ORCID account to gain recognition for their contributions.

Only the Journal title will be uploaded into the reviewer’s ORCID record, along with the date the record was uploaded; there is no identification of the article’s title or authors. Records are uploaded once a decision (accept, reject, or revision) has been made on the article.</em>

Would you like to be accredited by <a href="https://orcid.org/" target="_blank">ORCID</a> for this review?: No