London School of Hygiene & Tropical Medicine
(University of London)

Department of Health Services Research and Policy
15-17 Tavistock Place, London WC1B 3DP

Faculty of Public Health and Policy

Dr Sophie Cook
Editor
The British Journal of Medicine

11th January 2021

Re: BMJ Analysis Manuscript submission: '*Put to the Test: An evaluation of how rapid testing technologies can be deployed to fight COVID-19*'

Dear Dr. Cook,

We would like to thank the reviewers for their helpful comments, and much appreciate the time taken to review this analysis article. We have addressed these comments in turn, the responses and evidence to which is copied below.

We hope that this satisfies the reviewers comments, and that this article can go forward for publication. Please do not hesitate to contact us if there are additional comments that need addressing.

Reviewer: 1

Recommendation:

Comments:
Put to the test: An evaluation of how new technologies can be deployed to fight COVID-19 is a very timely description of the English government's testing plans.
There has been a great deal of confusion about new testing technologies their costs and benefits. This evaluation provides a very helpful framework to consider potential benefits and harms in controlling the spread of COVID-19.
The paper covers a range of important areas for consideration including testing strategies, interpretation of test results including those of novel tests.
I strongly support the publication of this paper based on the added value of describing the proposed tests and testing regimes however the paper provides more than this.
It tackles directly the issue of mass testing and describes the recent approach taken in Liverpool.
As testing is central to international responses to the COVID-19 pandemic and governments have invested enormous resources to scale-up capacity this paper adds valuable insight into what tests and how they can or could be used. It also challenges the limitations of testing and includes important factors such as ethics and behavioural change. I think the most useful challenge the paper makes is in relation to views that testing can only be effective if it's part of a Holistic Public Health Approach with local contact tracing and support to enable individuals to isolate without financial or other detriments.
This paper also provides some interesting information for Test-to-Protect, Test-to-Release and Test-to Enable which I have not seen presented in this way before I believe it would be valuable to see this being published. There is also a similar approach to Asymptomatic Testing for International Arrivals which is an area of interest for many and would be useful to have these views published.
What I would have like to have seen in this paper was some cost benefit analysis however this was not what the authors set out to do - stating the large costs the government had already committed - but perhaps could be considered in a future publication.
As indicated above I support the publication of this paper.

No comments to address

Reviewer: 2

Recommendation:

Comments:
This is a really important paper and I think analysis of the role of different testing strategies and testing tools is vital as governments across the world negotiate how best to manage detecting COVID within the community and how to start to reopen or keep open as much of the economy as possible. For that reason I support a piece such as this in summarising and

analysing the key approaches that can be taken, and I think it can be an important contribution to the literature. However, I think it needs a bit of restructuring before publication.
- I was confused when reading it that you talk of strategies and then talk of tools (ie. PCR or Ag-LFT and then back to strategies. I wonder if it's better to start with one and then the other to make a more logical argument. i.e. that govs have many approaches they can take for their testing strategy, and depending which they choose, and what the policy aim of testing is, then they can chose the relevant test to support that strategy.

Updated to "We outline the different tests that can facilitate this (Appendix 1) and the benefits and risks associated with implementing them, (Appendix 2), focussing finally on how novel tests, particularly Ag-LFTs, can support the different testing strategies adopted internationally to respond to the COVID-19 pandemic (Figure 1, Appendix 3)."

- I think this is the over-arching argument you are trying to make in this paper (please correct me if i've misunderstood) but I don't think you make the case compelling enough - I think you might need to start with a blunt statement such as that "governments chose testing for multiple reasons, and this paper analyses those reasons and shows which tools are best for the job" to really make it clear what this paper is about. This can then be reiterated through a restructure and into the conclusion.

We believe the above response and edits addresses this.

In summary, I have less concern about the content (albeit it this is slightly beyond my knowledge base as a social scientist), but think it can be argued more convincingly.

Reviewer 3

Substance
A1) A number of issues receive little attention, obviously because of the wordcount limit of 2000 words. Anyhow I would have liked to see them mentioned, if only as 'bypassed for reasons of space.' I am thinking of following collateral concerns: cost and logistics; booking, delays and queuing (waiting time is a problem when people queue under the winter sky); epidemiological monitoring (including quantification of immunity and vaccination effects); problems at frontiers; and global coordination. Later also coordination with vaccination policies and the successive phases of a vaccination programme.

We have updated the last line of conclusion to include vaccination programme -
"A holistic public health approach joined up across towns, cities and regions (and coordinated with vaccination programmes), is key to sustainable recovery from the COVID-19 pandemic."
We have also included in SMART section
"Logistic factors are important, such as arrangements for booking and queuing for tests, so the choice of test policy will often be limited by cost and available workforce and capacity of the community to access booking systems and testing sites."

A2) Medical autonomy vs. concerns for 'thy neighbour' in the face of (non-wartime) global threats: the authors touch upon this but clearly do not commit themselves to a war analogy. Personally, I would vote for certain restrictions of autonomy.

We were unable to address the issues of autonomy due to limitations of space but completely agree that autonomy is a key part of the discussion about preventing transmission.

A3) Many of the pros and cons involved in the choice of test policy depend on cost and manpower comparisons. The paper holds almost no quantitative information in that regard.

We have included this now in response to comment A1 and hope that this addresses the reviewers concerns. We would like to address these issues in more detail, particularly regarding costs but are unfortunately limited by space.

Sensitivity, disease phases, etc.
B1) As regards the 'relative sensitivity,' i.e., that of the 'quick' test when the PCR result is treated as the truth, please warn newcomers that it can only be interpreted as done here (and in the media) if the relative specificity is very very high. Hypothetical example in which the 'quick' test is possibly the better test despite a low relative sensitivity (n = 10,000): Both positive: 80, only PCR positive: 20, only 'quick' test positive: 21 (!), both negative: 9,879; here, relative sens = 0.80, relative spec = 9879/9900 = 0.998, very high but not very very high (for comparison, the reverse relative sens = 80/101 = 0.792 < 0.80).

Given the general readership of a BMJ Analysis and the limited word count, we have tried to keep it readable and straightforward but agree that these are important points.
We have added 'these estimates can only be interpreted in the context of the performance of the gold standard test.'

B2) P5/32-33 and elsewhere: the words 'residual' and 'shedding' require a precise definition: 'residual' = 'post-symptomatic but possibly still infectious' or 'no longer infectious'? 'Shedding' = 'of any viral material' or 'of innocuous material only'? The literature is ambiguous, isn't it?

Updated to post-infectious shedding.

B3) The authors have borrowed Figure 3 from another source. Unfortunately, the texts under the horizontal axis are logically wrong. As shown by the crossings of the dotted lines, the green phase should be called 'Early infection detectable by PCR only (possibly still in part pre-infectious),' and the purple phase: 'Late phase detectable by PCR only (possibly in part already non-infectious).' There are two biological transitions of interest: from pre-infectiousness to infectiousness, and from there to a non-infectious post-phase. Whether they coincide with the vertical dotted lines is an empirical matter; the dotted lines as drawn here reflect test response only, not necessarily the biological epochs. [To depict the two transitions two separate vertical lines would be needed, and they may, one may imagine, fall before, within or after the green phase (and similarly with the purple phase). Better still, one might draw two hump-shaped curves, one, marked 'Infectious virus load,' within the other, marked 'Total virus load'; horizontal dotted lines (representing hypothetical levels of biochemical sensitivity) may then cut, or not cut, the humps in several ways, leading to diverse conclusions.]

We have changed the green and purple phase labels as suggested to 'Early infection detectable by PCR only (potentially pre-infectious).' and 'Late phase detectable by PCR only (potentially post-infectious).'
We fully accept the complexities and nuances a complex topic such as this requires, and do believe the detail in the text does this justice. However, we conclude that it would also be a slight misrepresentation to draw two curves showing 'infectious' and total viral load, as infectious viral load is not a defined amount. The idea of the schematic is to summarise the potential usefulness of a less sensitive test and that a lower sensitivity can be compensated for, and is possibly more effective in terms of finding cases, by more frequent testing. The other purpose of the figure was to show that highly sensitive testing may result in lots of post-infectious positives if undertaking asymptomatic testing, possibly leading to significant unnecessary quarantine during times of high prevalence. We believe that adding too much finer detail may well serve to confuse new readers further, detracting from the main point, whilst also not being quite scientifically accurate.

Presentation
C1) The term mass screening is really mass-repeat screening in the covid context. Readers (journalists!) may misunderstand the term.

We have avoided using the term mass screening to avoid confusing such a programme with organised mass screening (for example cancer screening programmes with national registries etc), but have changed the text so that it now refers to 'mass repeated-testing' at the bottom of page 5.

C2) Similarly, the wording on p. 5(bottom)-6(top) is academically precise and elegant, but hardly to be understood by junior media people or by the parties affected (nursing home leaders, public transport managers, …). Writing a contribution of this kind for the BMJ means writing for health policy people in Belgrad and Bogotá. Plain words, please.
Talking about "Specific testing strategies enabling social fabric, service stability or economic recovery may be targeted, for example to care home visiting, emergency services, public transport, or international arrivals.

We have amended this sentence and paragraphs. The phrasing referred to now is now split into two sections - Test-to-protect reads as:
'The recent policy of bi-weekly testing of front-line NHS staff with Ag-LFTs recognises that frequent testing can compensate for reduced sensitivity. Specific testing strategies may also be focused on protecting groups most susceptible to infection and transmission, such as key workers during the current lockdown, enabling essential service continuity, and possibly reducing overall transmission.'
And test-to-enable reads as:
'Test-to-enable policies seek to lift the current restrictions on social contact that are causing wider public health and economic harms, in a risk responsive way. For example, once current lockdown restrictions are eased, specific test-to-enable strategies may be able to reduce the harms of social isolation by enabling care home visiting, or supporting workplaces in fragile local economies to operate with risk mitigation.'

C3) If the tables and figures of the appendices present themselves as they do with my browser on my home (!) PC, they are hard to read. Some of the text needs enlargement, other parts a reduction. As the authors recognize, a professional hand is needed.

We do agree that the tables and the figures do require some editorial support to improve the presentation but we have simplified these a little further.

C4) References 14 and 16 are identical. Different paper may be meant.

Corrected.

C5) Appendices, Figure 1 and Appendix Table 1: both legends start with 'Principle,' a noun. It should be the adjective 'Principal.'

Corrected.

Reviewer: 4

Recommendation:

Comments:
This Analysis paper provides a really nice characterisation of potential testing strategies that could be deployed to reduce transmission of covid-19 infection in the community with a particular focus on use of rapid antigen tests for testing asymptomatic individuals. The paper is highly relevant and well thought through and will certainly be of interest to a general readership. I have a few comments on specifics of the paper that the authors might want to consider.
Pg 4 line – the title does not reflect the paper's focus on antigen tests nor their use for asymptomatic testing strategies.

The paper was intended to provide an overview of all the testing strategies, specifically detailing those deployed in various countries, which do tend to focus on asymptomatic testing using Ag-LFT's. However, although the focus is on antigen tests, we have tried deliberately not to restrict the paper to one about antigen tests alone, and we do outline the benefits and risks of many other tests and the details of a SMART testing strategy. In this latest draft we also outline proposals to reconsider symptomatic testing strategies. We do hope that this explanation addresses the reviewers concerns.
After consideration, we have updated the title to

Put to the test: An evaluation of how rapid testing technologies can be deployed to fight COVID-19

Pg 4 line 24-26 - The authors make the point that real word evaluations of rapid tests are needed in order to better understand accuracy and how tests might be used more widely, however this misses a crucial step in the evaluation process, which is to first establish accuracy in the populations in which these tests are intended to be deployed. Appendix 2 gives examples of the accuracy of 3 different Ag tests established in independent test evaluations but fails to mention that these were established predominantly in symptomatic individuals within the first few days of symptom onset and cannot be assumed to apply in asymptomatic cohorts. Antigen tests are not all equal, and it is vital that test accuracy is well understood before tests are selected for implementation in large scale evaluations such as we have seen in Liverpool.

We thank the reviewer for this helpful comment and agree that this needs to be made clearer. We have clarified as follows:

Added to the introduction -
'Such real-world evaluations of Ag-LFTs are needed to understand how these models work in different populations and settings, how they influence behaviour, and the contribution of Ag-LFT in overall strategies, where they have the potential to interrupt transmission while reducing the harms from restrictions.'

Amended Appendix 1 (previously Appendix 2) -
'The term 'clinical sensitivity/specificity' refers to the real-world identification of infections, rather than the analytical properties under laboratory conditions.
The term 'relative sensitivity/specificity' refers to test performance when compared to the 'gold standard' test, rRT-PCR. These estimates of accuracy for alternative tests can only be interpreted in the context of the performance of the 'gold standard' test, rRT-PCR.
* Note that this is information taken from the OptiGene COVID-19 Direct Plus RT-LAMP KIT-500 Direct RT-LAMP test manual. These tests have been piloted in selected UK hospitals by DHSC and there is more recent real-world data for this assay, but it is as yet unpublished.
**Reported SAMBA II results are after discrepant analysis (i.e re-testing) of initially false positive and false negative results and therefore likely to inflate accuracy measures.
*** Mean of FIND evaluations from Brazil, Germany, and Switzerland.
Although peak viral load between individuals is comparable,[11] clearance rates are different - It should be noted that data for the Innova Antigen test from the DHSC/PHE/Oxford evaluations includes some testing of asymptomatic individuals, which is likely to impact on reported relative sensitivity, compared to the evaluation of the PANBIO Covid-19 Ag Rapid Test which was on almost all symptomatic individuals within the first few days of symptom onset, and SD Biosensor which was on mostly symptomatic individuals. It is also important to note that antigen tests and population groups are not all equal, and it is therefore vital that test accuracy is understood in each population it is used in (for example asymptomatic/pauci/symptomatic, and by age and background prevalence) before any large-scale roll-out.
All results here should be treated with caution – Manufacturer's instructions for use may over-estimate accuracy compared to real-world test use. Although, where possible, data presented is from real-world pilot evaluations, results may not be directly applicable to specific real-world scenarios.
Caution must also be given for new variants, such as B.1.1.7 (VOC-2012/01), which may affect test accuracy. Whilst some S gene PCR assays, including the Thermo Fisher assay used in the UK Lighthouse Laboratories, are affected, many assays

target for multiple genes and should still be able to identify cases. S gene target failure (SGTF) in Lighthouse Laboratories is in fact being used as a proxy to indicate carriage of VOC-2012/01).[14]'

Pg 4 line 46-50 – There is an implication here that false negatives on Ag tests only occur at lower viral loads, however false negatives on Ag tests also arise at higher viral loads (e.g. <=25 Ct) although at a lower rate. As Ag tests are increasingly being promoted as tests of so called 'infectiousness', it is important that we do not lose sight of the fact that they are not perfect even at high viral load.

Added to Novel Test Section (P4)
'However, as explained above, sensitivity is user and operator-dependent[10], and self-swabbing in real-world conditions is likely to miss more infections than swabbing in controlled conditions. In Liverpool, data suggests Ag-LFTs missed around a third of substantially infectious individuals (high detected viral load), although further work is required to more precisely determine how this relates to infectiousness.[5] The test must also be read and laboratory scientists are less likely than others to report false negatives, although this can improve with more robust protocols, additional training, and possibly AI-augmented reading.[5,10] Also, antigen tests and population groups are not all the same so test accuracy must be understood for different groups (e.g. asymptomatic/(pauci-)symptomatic, and by age and background prevalence) before large-scale use.'

Pg 4 line 46-50 – It is not just swab collection that incurs false negatives, but test interpretation even by health care workers has been shown to incur more false negatives compared to interpretation by laboratory scientists.

We have now addressed this point above.

Pg 4 line 55-58 - The communication to the public in regard what a negative test result means, especially a negative Ag test, is fundamentally important to any testing strategy that involves asymptomatic individuals. The point that Ag tests are a test of the 'moment', and only reduce risk of infection at the time of the test, could perhaps be made more strongly here.
We agree with this comment and have clarified this at the bottom of page 3. 'Whatever the sensitivity, all tests are just snapshot of the 'moment' the sample was taken, and only reduce the risk of infection at the time the swab was taken.'

Pg 5-6 – As mentioned above, the exposition of possible testing strategies is really well done. The SMART approach looks highly sensible if implemented correctly and with the right test. Given emerging evidence for exceptionally poor performance of Innova in mass testing scenarios, it will be interesting to see how effective a SMART approach can actually be (assuming Innova is the test of choice here). More detail in regard to frequency of testing, how test results and ongoing 'risk' are communicated, and what sort of 'tactical changes' might be implemented would be nice to see.

Added both
"For example, now Liverpool is in national lockdown, testing is targeted at workplaces, to enable continuity of essential services and to protect against transmission in high mixing environments such as supermarkets."

"Although the behavioural responses to large-scale asymptomatic testing in the community are not fully understood, particularly the potential increase in hazardous behaviours following a negative result, ONS survey data in Liverpool showed most (62%) said a negative result would be unlikely to cause them to change their behaviour.[5] However, some said they were more likely to visit friends (9%) or go to work (7%), emphasising the need to communicate the importance of maintaining COVID-safe behaviours."

Appendix Table 2 – could be much clearer about the sources of accuracy data here which range from a systematic review (RT-PCR), to manufacturer instructions for use (Rt-LAMP and NGS) which are notorious for over-estimating accuracy compared to real world test use, to independent test evaluations of Ag tests. All quoted sensitivities and specificities are from symptomatic populations and will not be directly applicable to the scenarios presented in the paper.

We have updated LamPORE sensitivity/specificity data with data from a large study of symptomatic and asymptomatic testing in NHS labs. All data presented is now from published pilot evaluations (accepting real-world use, especially when in different contexts, may be different).
To address this, we have added into the Table legend

"These estimates of accuracy for alternative tests can only be interpreted in the context of the performance of the 'gold standard' test, rRT-PCR.
* Note that this is information taken from the OptiGene COVID-19 Direct Plus RT-LAMP KIT-500 Direct RT-LAMP test manual. These tests have been piloted in selected UK hospitals by DHSC and there is more recent real-world data for this assay, but it is as yet unpublished.
**Reported SAMBA II results are after discrepant analysis (i.e re-testing) of initially false positive and false negative results and therefore likely to inflate accuracy measures.
*** Mean of FIND evaluations from Brazil, Germany, and Switzerland.

Although peak viral load between individuals is comparable,[11] clearance rates are different - It should be noted that data for the Innova Antigen test from the DHSC/PHE/Oxford evaluations includes some testing of asymptomatic individuals, which is likely to impact on reported relative sensitivity, compared to the evaluation of the PANBIO Covid-19 Ag Rapid Test which was on almost all symptomatic individuals within the first few days of symptom onset, and SD Biosensor which was on mostly symptomatic individuals. It is also important to note that antigen tests and population groups are not all equal, and it is therefore vital that test accuracy is understood in each population it is used in (for example asymptomatic/pauci/symptomatic, and by age and background prevalence) before any large-scale roll-out.
All results here should be treated with caution – Manufacturer's instructions for use may over-estimate accuracy compared to real-world test use. Although, where possible, data presented is from real-world pilot evaluations, results may not be directly applicable to specific real-world scenarios."

Pg 8 – line 26-30 – the reported SAMBA II results are after discrepant analysis (i.e. re-testing) of initially FP and FN results and therefore likely to inflate accuracy measures.

Added -
**Reported SAMBA II results are after discrepant analysis (i.e re-testing) of initially false positive and false negative results and therefore likely to inflate accuracy measures.

Pg 9 line 24-27 – Some tests do use saliva samples but I do not think that accuracy has yet been shown to be the same as for nasopharyngeal swabs.

Added
Some tests use saliva samples which can improve throughput and acceptability (although may reduce accuracy).

Reviewer: 5

Recommendation:

Comments:
This is a timely paper that is of considerable immediate interest and it fully deserves publication. The importance of the paper arises not only from the assessment of the value of testing methodologies deployed in respect of COVID-19 but also from the way in which the authors make the connection between the science behind the testing methodologies and the ways in which the testing methods should be deployed. It is particularly valuable to those making decisions on the approach to test that should be used at a local level and within institutions.

No comments to address.

Editor's comments:

The reviewers are supportive of the paper and include some helpful suggestions to further strengthen the piece. We would like to see the paper revised to incorporate some of these suggestions and also the editors' comments with a view to publication.

*This is a very well-timed piece which clarifies some elements of mass testing. We agree with Clare Wenham's suggestions to restructure this slightly to make this clearer and more compelling.

We agree with this comment and, as per Clare Wenham and the editors suggestion, we have re-structured the article to first focus on novel tests and then testing strategies, and how novel tests may be integrated into an overall public health strategy to mitigate the direct and wider harms of the COVID-19 pandemic. We believe these changes address this comment.

*We also felt the Liverpool example would benefit from being mentioned upfront or integrated into the paper earlier.

We have now integrated the Liverpool example into the introduction, and analyse the Liverpool SMART strategy earlier in the article. We believe these changes address this comment.

*Table 2 is a very helpful summary of the tests. We think this would be better in the main article rather than as an appendix. We would not have capacity to publish this in print but it would be helpful to have this integrated into the article online.

We agree that this table is a helpful summary of testing technologies. However, due to space limitations, we have currently left it as an Appendix Table (Appendix Table 1). We are very happy for this to be described as a Table in the main article alongside the 4 figures already present, however we have currently left it as an Appendix table as we do not have the IT technical skills to format and submit this as one image which is readable. We are happy to work with you further on this, in addition to hopefully working with the infographics team on the 4 figures.

Once again, we would like to thank the reviewers and editors for their helpful and constructive comments, and we hope that our responses are satisfactory.

Yours Sincerely,

Alex Crozier BSc
PhD Candidate,
Division of Biosciences, Medical Sciences Building, University College London, WC1E 6BT

Email:
alexander.crozier.20@ucl.ac.uk
Tel: 07412411054

Dr Selina Rajan MSc
Honorary Research Fellow
Department of Health Services Research and Policy,
London School of Hygiene and Tropical Medicine,
15-17 Tavistock Place, Kings Cross,
London WC1H 9SH, UK

Professor Martin McKee
Professor European Public Health
Department of Health Services Research and Policy
London School of Hygiene and Tropical Medicine