PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Diagnostic accuracy of an app-guided, self-administered test for influenza among individuals presenting to general practice with influenza-like illness: Study protocol
AUTHORS	Lyon, Victoria; Zigman Suchsland, Monica; Chilver, Monique; Stocks, Nigel; Lutz, Barry; Su, Philip; Cooper, Shawna; Park, Chunjong; Lavitt, Libby Rose; Mariakakis, Alex; Patel, Shwetak; Graham, Chelsey; Rieder, Mark; LeRouge, Cynthia; Thompson, Matthew

VERSION 1 – REVIEW

REVIEWER	José Tomás Prieto Centers for Disease Control and Prevention, United States.
REVIEW RETURNED	17-Dec-2019

GENERAL COMMENTS	Summary: Lyon and colleagues present a protocol for a human subjects research study that aims to determine the accuracy of an influenza self-test whose instructions are presented through a smartphone application. The authors describe with eloquence the significance of the study, the objective, the methods, the study instruments, and the potential implications of this work. The study will start in July 2019. This reviewer believes that additional clarifications are needed, particularly about the main measurement, accuracy. Main concerns: The main outcome of the study is the accuracy of detection of influenza infection based on interpretation of the home tests, when compared to the gold standard results provided by RT-PCR testing. It is not clear whether the authors intend to measure the accuracy of the test itself, or the accuracy of the test-at-home setup, which covers several processes such as 1) being provided with a functional home test, 2) being provided with a test that will or will not produce false positives, 3) several participant actions including the use of a smartphone app and self-swabbing their nose, 4) interpretation of the readings, and 5) entering information in an app, among others. As currently stated, measured accuracy will be a composition of several types of inaccuracies along the different study processes. Having details about the accuracy of the flu@home test in a laboratory setting, for example, may be helpful. In any case, being more explicit about the type (or types) of accuracy being measured is mostly needed. Additional, minor concerns:

	- Page 9: “There has not been any prior testing of the app in Australia”. Has there be any prior testing of the flu@home test in Australia? All references used to calculate the sample size seem to come from US contexts, suggesting that the sample size may be too optimistic for a new environment. - The flu@home test will very likely produce false positive and false negative results. For the benefit of participants, there seems to be a need to embed a new study process when contradictory results will arise from the flu@home when compared to RT-PCR results. All positive results at home will be handled in the same way: participants will be given links to usual care recommendations. If this comes with a positive RT-PCR result, perfect, the GP will get in touch with the participant. But what if a flu@home test is positive, but RT-PCR is negative? Will the patient be made aware of this? Or what if the home test is negative, but RT-PCR is positive? The patient will receive a call from the GP, but the test@home will have previously indicated a negative result for influenza. - Related to the previous point, the patient survey does not seem to include questions about the general feeling of testing positive for influenza at home. This reviewer believes it may cause some type of distress among patients, and it may be helpful to evaluate. - The patient survey seems to include several questions that are not directly linked to the study objective. For example, how will questions “How many bedrooms are in your home?” or “In the past week, have you been in contact with any children under five years old for over an hour?” be used? A predictive model using responses to these questions may not be relevant. A revision seems necessary to curate the list of questions in the participant survey. - Data will be encrypted when transmitted to online services for storage. Is data deleted after transmission? It is not clear whether the app will store information locally. If yes (even if it’s just in cache memory), will this data be encrypted too? How? This is important because the app collects sensitive information about participant’s health and risk factors that could be visible to unauthorized people. - The protocol background could be enriched by including references to previous studies that have evaluated the effectiveness of smartphone-based data collection of ILI symptoms both in high and low-income settings.

REVIEWER	Landelle Caroline Grenoble Alpes University Hospital France
REVIEW RETURNED	06-Mar-2020

GENERAL COMMENTS	I have reviewed this protocol with great interest and I thank you for this opportunity. The study is interesting and promising. A self-administered test could indeed be very useful in primary care considering the very high incidence of influenza infection. The app helps to interpret the result and have also an educative impact. However there are some points I would like to discuss. The overall quality of the protocol is good but there is one major modification and a few minor modifications that should be made before considering this article for publication. Major modification:

	About the main outcome, you say you want to assess the "accuracy of detection of influenza infection". I find this formulation too vague. You should be more precise about your objective and the results you expect. Do you expect your test to be as effective as gold standard? If not, what is the tolerance threshold? Is the test objective to rule out flu or rule in flu (positive predictive value, negative predictive value, sensibility, specificity)? Minor modifications: In your introduction, you say that one of the advantages of this test is that it would allow patients to self-diagnose themselves so they can be treated more quickly. Are antivirals like oseltamivir available without a prescription from a physician in the Australian health system? Because if it is not the case, that would mean that patients would still need to see their GP to get a prescription. Please, develop this point. Also in your introduction, you state that influenza is responsible for 378 deaths per year. I don't find this figure in your references. In reference 2, authors say "87 deaths were recorded in which influenza was signified as the direct cause of death". They also say "the true number of deaths attributable to influenza is known to be considerably higher". In reference 3, authors say that influenza causes more than 3000 deaths per year in Australia. Could you check the number of "378 deaths"? About the flu@home app, I read in the study that this software was developed in the USA and previously tested in this country. In spite of this, I notice no reference about it in your article. Should it be understood that no formal study was conducted in the United States? If such is the case, could you explain the reasons for performing the study in Australia rather than the United States? About the self-test kit, you write that it is a customized version of a commercial test. Perhaps it would be good to give more details about the modifications performed. About the ASPREN network, it would be nice to have more information about it, such as the number of GP involved in this network, and the representativeness of these GP and their patients in Australia. About exclusion criteria, why did you not add physical disabilities, such as blindness and other impairments that make the app non-usable? About the recruitment method, you say you include the first three adult patients presenting with ILI symptoms, and all ILI patients aged 65 or more. But when an elder aged 65 or more is one of the first three patients, do you count this person among the three or not? I think it would be good to clarify this. About the sample size calculation, it appears that you included too many patients. You say that FDA requires 120 flu positive specimens for the evaluation of rapid flu diagnostic test; however, you tailor your study to expect 276 positive tests. This part needs to be re written, unnecessary patient inclusions are unethical.

	About the test results, if a patient self-test is negative, while the PCR is positive for flu, how do you manage the discordance? You wrote something about it in the part “Patient discontinuation”, but you do not specify if this apply to the situation described above. I recommend to add a part about “test discordance”, where you specify how you handle such things. In discussion, you might also add information about the cost of this public health measure. Typing errors - Page 9 line 39-10 “and and” - Page 11 line 40 “need to need to”

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Lyon and colleagues present a protocol for a human subjects research study that aims to determine the accuracy of an influenza self-test whose instructions are presented through a smartphone application. The authors describe with eloquence the significance of the study, the objective, the methods, the study instruments, and the potential implications of this work. The study will start in July 2019. This reviewer believes that additional clarifications are needed, particularly about the main measurement, accuracy.

Author response:
Thank you for these comments, we have clarified the language regarding the main outcome which is comparative accuracy of the home-based flu test compared to a reference test for flu obtained by participants’ GPs.

Main concerns:
The main outcome of the study is the accuracy of detection of influenza infection based on interpretation of the home tests, when compared to the gold standard results provided by RT-PCR testing. It is not clear whether the authors intend to measure the accuracy of the test itself, or the accuracy of the test-at-home setup, which covers several processes such as 1) being provided with a functional home test, 2) being provided with a test that will or will not produce false positives, 3) several participant actions including the use of a smartphone app and self-swabbing their nose, 4) interpretation of the readings, and 5) entering information in an app, among others. As currently stated, measured accuracy will be a composition of several types of inaccuracies along the different study processes. Having details about the accuracy of the flu@home test in a laboratory setting, for example, may be helpful. In any case, being more explicit about the type (or types) of accuracy being measured is mostly needed.

Author response:
The reviewer is correct that we are comparing the accuracy of a test for influenza conducted entirely by untrained people at home, compared to a reference standard of an influenza test conducted by their GP. We acknowledge that there are several areas where the accuracy of a home test could indeed be lower than that of a reference test, including the 5 factors noted above. However, we note that the test provided to participants is one that has proven accuracy in clinical settings (and is widely used in clinical settings) as cited now in the manuscript discussion, mitigating points #1 and #2 above. We also note that there is now robust evidence that individuals can obtain nasal swabs for influenza detection with comparable performance to that of health care professionals (Seaman et al - now cited in the discussion), mitigating point 3. We agree that users could misinterpret the test strip reading,

which is why we ask them to take a photo of the test strip using their smartphone, allowing the research team to later make the comparison that the reviewer suggests (ie mitigating interpretation error, point 4 above). The final point about entering information in the app incorrectly is also something that we are able to monitor to some extent – we will measure the proportions of individuals who start but then abandon their test, as well as missingness of variables collected in the app.
We acknowledge many of these potential issues could affect the accuracy of this type of test in home use settings, which is why we are conducting this study. We refer to these more clearly now in the Discussion section.

Additional, minor concerns:
Reviewer comment: - Page 9: “There has not been any prior testing of the app in Australia”. Has there be any prior testing of the flu@home test in Australia? All references used to calculate the sample size seem to come from US contexts, suggesting that the sample size may be too optimistic for a new environment.

Author response:
As noted above, the app that was designed to guide users in conducting this research study was evaluated by the app developers with user feedback from individuals in the USA, it was then submitted and approved by Apple and Android app stores, meeting all such requirements in Australia. The app was used in a study of home-based flu testing in the USA, prior to the current study in Australia, providing further opportunities for improving the app and user experience (manuscript in preparation). The research team did work very closely with our Australian based collaborators who operate the ASPREN GP network, to ensure further feedback on the language and content of the app, before it was approved by the Australian Apple and Android app stores. We do not believe that further testing of the app is necessary for this research study.

Reviewer comment: - The flu@home test will very likely produce false positive and false negative results. For the benefit of participants, there seems to be a need to embed a new study process when contradictory results will arise from the flu@home when compared to RT-PCR results. All positive results at home will be handled in the same way: participants will be given links to usual care recommendations. If this comes with a positive RT-PCR result, perfect, the GP will get in touch with the participant. But what if a flu@home test is positive, but RT-PCR is negative? Will the patient be made aware of this? Or what if the home test is negative, but RT-PCR is positive? The patient will receive a call from the GP, but the test@home will have previously indicated a negative result for influenza.
Author response:
There is clearly a risk for discordant results. We have clarified in the manuscript that a participant’s clinical care was entirely under their own GP, and the flu@home test is a research test, and that any discordant results or concerns should be brought to their GP to provide appropriate medical advice.

Reviewer comment: - Related to the previous point, the patient survey does not seem to include questions about the general feeling of testing positive for influenza at home. This reviewer believes it may cause some type of distress among patients, and it may be helpful to evaluate.
Author response:
We appreciate this concern. The study design mitigates patient distress by providing GP care prior to initiating home-testing. Participants received standard clinical care from their GP prior to completing the home test, and received a message in the app that says, “The interpretation of your result may differ from a medical test conducted in a clinical lab environment. In no circumstances should the results of this test be relied upon without independent consideration and confirmation by a qualified medical practitioner.” Patients will be notified of the results of the reference test by their GP, who will

provide standard care based on the RT-PCR results. Understanding general feelings of testing positive for influenza at home was out of scope for this accuracy study. We believe that the user feedback survey is robust and asks many questions around the impact or utility of the home test among users. We have used this same survey in a US evaluation of the flu@home test, and intend to compare US vs Australian responses to these surveys.

Reviewer comment: - The patient survey seems to include several questions that are not directly linked to the study objective. For example, how will questions “How many bedrooms are in your home?” or “In the past week, have you been in contact with any children under five years old for over an hour?” be used? A predictive model using responses to these questions may not be relevant. A revision seems necessary to curate the list of questions in the participant survey.
Author response:
The questions in the patient survey are designed to determine exposure risk of influenza from household contacts, contact with children (among other factors) which are well known risk factors for influenza infection in the community. This information will facilitate interpretation of our results in terms of participant characteristics, including age, exposure risk, vaccination etc. We have added this explanation to the “Other variables” section on page 5).

Reviewer comment: - Data will be encrypted when transmitted to online services for storage. Is data deleted after transmission? It is not clear whether the app will store information locally. If yes (even if it’s just in cache memory), will this data be encrypted too? How? This is important because the app collects sensitive information about participant’s health and risk factors that could be visible to unauthorized people.
Author response:
The flu@home Australia app is available for personal devices which are expected to be under control of an individual who uses a passcode to access the device. All supported devices use encryption to protect app data resident on the device. This encryption is afforded by the device itself, not a specific application. In the event that a device is stolen, the device's onboard locking feature is the front-line defense against access to data on the device. The flu@home application does not collect the user's name, email, or other key identifiable information in the app. It focuses on data collection of symptoms, disease presentation, and some demographics. The level of data protection offered by the flu@home app is the same level of protection afforded to most other health applications, email, messaging, etc. available on a mobile device. This information has been clarified in the manuscript and Appendix A.

Reviewer comment: The protocol background could be enriched by including references to previous studies that have evaluated the effectiveness of smartphone-based data collection of ILI symptoms both in high and low-income settings.
Author response:
We have added several citations of studies that have used smartphone apps to collect data on ILI symptoms. However, it is worth distinguishing that previous studies have collected ILI symptoms for public health surveillance of influenza, whereas the flu@home app serves as a tool to facilitate individual patient’s diagnosis and management of an influenza illness.

Reviewer: 2
I have reviewed this protocol with great interest and I thank you for this opportunity.
The study is interesting and promising. A self-administered test could indeed be very useful in primary care considering the very high incidence of influenza infection. The app helps to interpret the result and have also an educative impact. However there are some points I would like to discuss. The overall quality of the protocol is good but there is one major modification and a few minor modifications that should be made before considering this article for publication.

Author response:
Thank you for these comments, we have clarified the language regarding the main outcome measure of comparative accuracy.

Major modification:
About the main outcome, you say you want to assess the “accuracy of detection of influenza infection”. I find this formulation too vague. You should be more precise about your objective and the results you expect. Do you expect your test to be as effective as gold standard? If not, what is the tolerance threshold? Is the test objective to rule out flu or rule in flu (positive predictive value, negative predictive value, sensibility, specificity)?

Author response:
We have restated this more clearly in the introduction, namely that this is a comparative accuracy study of home based influenza test (index test) compared to GP obtained influenza test (reference test). We do not state a priori that we expect the index test to be more or less effective than the reference test, nor do we state its ‘tolerance threshold’. We also do not feel it is necessary to specify which of the test accuracy metrics are of more or less interest here, all have value depending on the pretest probability of flu, expected use of the test results and multiple other factors.

Reviewer comment
Minor modifications:
In your introduction, you say that one of the advantages of this test is that it would allow patients to self-diagnose themselves so they can be treated more quickly. Are antivirals like oseltamivir available without a prescription from a physician in the Australian health system? Because if it is not the case, that would mean that patients would still need to see their GP to get a prescription.
Please, develop this point.

Also in your introduction, you state that influenza is responsible for 378 deaths per year. I don’t find this figure in your references. In reference 2, authors say “87 deaths were recorded in which influenza was signified as the direct cause of death”. They also say “the true number of deaths attributable to influenza is known to be considerably higher”. In reference 3, authors say that influenza causes more than 3000 deaths per year in Australia. Could you check the number of “378 deaths”?

Author response:
Thank you for pointing this out. The 378 deaths per year was mistakenly pulled from a research study that did not encompass all of Australia. We have updated the statistic to 1,500- 3000 deaths per year from influenza, as reported by the Influenza Specialist Group (citation added). We have also amended the discussion to note that a) antivirals are only currently available with prescription, but this does not necessarily require a face to face consultation, b) corrected the references in the introduction to ensure we have cited these correctly.

Reviewer comment:
About the flu@home app, I read in the study that this software was developed in the USA and previously tested in this country. In spite of this, I notice no reference about it in your article. Should it be understood that no formal study was conducted in the United States?
If such is the case, could you explain the reasons for performing the study in Australia rather than the United States?
Author response:
We have responded to this point above.

One reason we conducted this study in Australia is to take advantage of the southern hemisphere flu season, having previously conducted a study in the US flu season immediately prior. The northern hemisphere/southern hemisphere winter seasons and flu seasons facilitates research on this seasonal infection. Furthermore our Australian collaborators were similarly interested in the potential value for this type of self-test in Australian primary care, given the burden of influenza.

Reviewer comment:
About the self-test kit, you write that it is a customized version of a commercial test. Perhaps it would be good to give more details about the modifications performed.
Author response:
We have clarified that we did not modify the Quidel Quickvue test in any way. The packaging of the test was rebranded to include the flu@home logo, but all of the test components inside the box remain exactly the same at the Quidel Quickvue test.

Reviewer comment:
About the ASPREN network, it would be nice to have more information about it, such as the number of GP involved in this network, and the representativeness of these GP and their patients in Australia.
Author response:
We have provided some more information about the ASPREN network and added citations to previous work from this group (Study Population section, page 4).

Reviewer comment:
About exclusion criteria, why did you not add physical disabilities, such as blindness and other impairments that make the app non-usable?
Author response:
Our exclusion criteria are “non-English speakers, people who are incarcerated, people highly dependent on medical care who may be unable to give consent, and people with a cognitive impairment, an intellectual disability or mental illness”. We did not specifically mention physical disabilities or impaired vision but left the decision to recruit a patient up to their own GP at the time of their visit which is consistent with most research conducted in GP settings.

Reviewer comment:
About the recruitment method, you say you include the first three adult patients presenting with ILI symptoms, and all ILI patients aged 65 or more. But when an elder aged 65 or more is one of the first three patients, do you count this person among the three or not? I think it would be good to clarify this.
Author response:
In the recruitment method described in the paper, we will recruit three adults with ILI symptoms under age 65 per week, plus all individuals over age 65 who present ILI in clinic. In the scenario posed by the reviewer, the elder patient would not count towards the first three adult patients presenting ILI in clinic. However, after the submission of this protocol we expanded the enrollment to any patient presenting with an ILI who is 18 years and over, has a smartphone and agrees to participate in the study. Therefore, this issue is no longer included in the manuscript.

Reviewer comment:
About the sample size calculation, it appears that you included too many patients. You say that FDA requires 120 flu positive specimens for the evaluation of rapid flu diagnostic test; however, you tailor your study to expect 276 positive tests. This part needs to be re written, unnecessary patient inclusions are unethical.
Author response:
We provide the FDA recommendation for regulatory approval of influenza tests that are designed for use in clinical settings, however they do not provide guidance for sample size for studies of influenza that are conducted in home settings. The sample size recommended by the FDA does not have any

particular scientific basis. We expect that the confidence intervals around test accuracy measures are likely to be far broader for the flu@home test, given the multiple potential reasons why accuracy may be lower, hence a larger sample size is necessary.

Reviewer comment:
About the test results, if a patient self-test is negative, while the PCR is positive for flu, how do you manage the discordance? You wrote something about it in the part “Patient discontinuation”, but you do not specify if this apply to the situation described above.
I recommend to add a part about “test discordance”, where you specify how you handle such things.
Author response:
As noted above, clinical care for participants will use the GP’s reference test and will entirely be based on the patient’s own GP. Within the materials given to patients in the study, they are provided with guidance on what to do with test results that are discordant with their GP test results and/or recommendations. Participants are clearly provided information that this is a research study, and that their GP’s care is the standard of care.

Reviewer comment:
In discussion, you might also add information about the cost of this public health measure.
Author response:
We have noted that the comparative costs of home testing (vs usual GP care) would need to be considered in deployment of home based testing. Obviously in this study we are not intending to perform an economic analysis.

Reviewer comment:
Typing errors
- Page 9 line 39-10 “and and”
- Page 11 line 40 “need to need to”
Author response:
We have made these edits.

VERSION 2 – REVIEW

REVIEWER	Jose Tomas PRIETO Private sector, France
REVIEW RETURNED	22-Jun-2020

GENERAL COMMENTS	Re - main outcome of the study. The authors have not satisfactorily changed the description of the main outcome in introduction and abstract, and this reviewers believes that the problem has not been resolved yet. As described, the main outcome seems to continue to be the accuracy of the home test and not the accuracy of results in this particular, home context. This is a problem at least for two main reasons: 1) because there seems to be an implicit parallel between the test conditions in the GP context and the test conditions in the home context, and 2) interpretation - e.g., what would a 90% accuracy result mean? Would this be explained by sensitivity/specificity of tests, or rather, by participant dexterity with apps or swabs? More transparency is needed to avoid misinterpretation of the work done. Re - prior testing of the app in Australia. The authors have not responded to this reviewer's earlier comment. Authors state on page 9 that "The expected completion rate of the home testing procedure is based on a USA-based pilot study that found that

	60% of individuals completed the flu@home test kit when it was mailed to them." This was used to calculate the sample size in the Australian case. Is there any evidence that the USA case can be transposed to the Australian case? For example, are participation rates in similar contexts comparable between the countries? Are smartphone or app usage similar? Authors responded that "any discordant results or concerns should be brought to their GP to provide appropriate medical advice." Has this been discussed with participating GPs, and are they aware of the probable additional burden of work? Re - "Understanding general feelings of testing positive for influenza at home was out of scope for this accuracy study." If the goal is to scale-up, this reviewer believes that the user experience is key when dealing with a health app. The fact that this question was not asked in a previous survey is not reasonable justification.

VERSION 2 – AUTHOR RESPONSE

Author response: We have added this as a limitation. See also response below

Reviewer 1

Re - main outcome of the study. The authors have not satisfactorily changed the description of the main outcome in introduction and abstract, and this reviewer believes that the problem has not been resolved yet. As described, the main outcome seems to continue to be the accuracy of the home test and not the accuracy of results in this particular, home context. This is a problem at least for two main reasons: 1) because there seems to be an implicit parallel between the test conditions in the GP context and the test conditions in the home context, and 2) interpretation - e.g., what would a 90% accuracy result mean? Would this be explained by sensitivity/specificity of tests, or rather, by participant dexterity with apps or swabs? More transparency is needed to avoid misinterpretation of the work done.

Author response: We are clear in the revised manuscript that the main outcome of the study is indeed comparative accuracy. In addition, we also will evaluate multiple aspects that could impact the accuracy of the index test. This includes asking participants about their experience in completing a low-nasal swab, as well as determining the mobile app usability.

Re - prior testing of the app in Australia. The authors have not responded to this reviewer's earlier comment. Authors state on page 9 that "The expected completion rate of the home testing procedure is based on a USA-based pilot study that found that 60% of individuals completed the flu@home test kit when it was mailed to them." This was used to calculate the sample size in the Australian case. Is there any evidence that the USA case can be transposed to the Australian case? For example, are participation rates in similar contexts comparable between the countries? Are smartphone or app usage similar?

Author response: We do not have any additional information on which to base the expected response rate, there are no published trials of this type of technology in Australia. The most comparable studies we identified in the Australian context were self-tests for malaria that required a blood sample (such as Jelinek et al. 1999 and Whitty et al. 2000) and self-tests for HIV (Chan et al. 2015). None of the studies of home-based diagnostics conducted in Australia have used a mobile app plus a diagnostic test, which may affect completion rates. For these reasons, we felt it was inappropriate to base our sample size for this influenza study based on these studies. Therefore, it seems reasonable to base our expected completion rate on data from a study in the USA using this smartphone-enabled influenza self-test.

Re: Authors responded that "any discordant results or concerns should be brought to their GP to provide appropriate medical advice." Has this been discussed with participating GPs, and are they aware of the probable additional burden of work?

Author response: The Australian co-investigators on this study have very close relationships with the GPs in the practice network that will be used for this trial and have approved all study procedures, as has the local ethics review committee. All participating GPs were made aware of the probably additional burden of work before opting into the study. We do not feel that further clarification is needed to address this point

Re - "Understanding general feelings of testing positive for influenza at home was out of scope for this accuracy study." If the goal is to scale-up, this reviewer believes that the user experience is key when dealing with a health app. The fact that this question was not asked in a previous survey is not reasonable justification.

Author response: It is correct, we did not ask this exact question “understand general feelings of testing positive”. In our survey, we asked about multiple elements of participant experience with using the influenza test at home, including their experience with the test, ease of use of the test, skills needed to use the test, their understanding of the results of the test, their opinion of their GP’s support for doing this test, their use of shared decision making with their GP, self-efficacy, health information seeking, etc. We have added as a limitation that we did not also ask about ‘understanding of general feelings of testing positive’.

We hope that our revisions are satisfactory.