Check for updates

RESEARCH ARTICLE

Better than we thought? The diagnostic performance of an influenza point-of-care test in children, a Bayesian re-analysis [version 1; peer review: 1 approved with reservations, 1 not approved]

Joseph Lee

Nuffield Department of Primary Care Health Sciences, Radcliffe Observatory Quarter, Oxford, UK

V1 First published: 18 Jan 2017, 6:53
<https://doi.org/10.12688/f1000research.10068.1>
 Latest published: 18 Jan 2017, 6:53
<https://doi.org/10.12688/f1000research.10068.1>

Abstract

Background: Point-of-care tests (POCTs) for influenza have been criticised for their diagnostic accuracy, with clinical use limited by low sensitivity. These criticisms are based on diagnostic-accuracy studies that often use the questionable assumption of an infallible gold standard. Bayesian latent class modelling can estimate diagnostic performance without this assumption. **Methods:** Data extracted from published diagnostic-accuracy studies comparing the QuickVue® influenza A+B influenza POCT to reverse-transcriptase polymerase chain reaction (RT-PCR) in two different populations were re-analysed. Classical and Bayesian latent class methods were applied using the Modelling for Infectious diseases CEntre (MICE) web-based application. **Results:** Under classical analyses the estimated sensitivity and specificity of the QuickVue® were 66.9% (95% confidence interval (CI) 61.4-71.9) and 97.8% (95% CI 95.7-98.9), respectively. Bayesian latent class models estimated sensitivity of 97.8% (95% credible interval (CrI) 82.1-100) and specificity of 98.5% (95% CrI 96.5-100). **Conclusions:** Data from studies comparing the QuickVue® point-of-care test to RT-PCR are compatible with better diagnostic performance than previously reported.

Keywords

Bayesian latent class models , influenza , diagnostic accuracy , point-of-care test , near-patient test , primary care , paediatrics

Open Peer Review

Approval Status

	1	2
version 1
18 Jan 2017	view	view
1. Nicolas Tremblay , Centre de recherche du Centre hospitalier de l'Université de Montréal, Montreal, Canada 2. Benjamin J. Cowling , The University of Hong Kong (HKU), Hong Kong, China		
Any reports and responses or comments on the article can be found at the end of the article.		

Corresponding author: Joseph Lee (joseph.lee@phc.ox.ac.uk)

Competing interests: No competing interests were disclosed.

Grant information: JLL is a Career Progression Fellow funded by the UK National Institute for Health Research's School for Primary Care Research (<https://www.spcr.nihr.ac.uk/>).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2017 Lee J. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Lee J. **Better than we thought? The diagnostic performance of an influenza point-of-care test in children, a Bayesian re-analysis [version 1; peer review: 1 approved with reservations, 1 not approved]** F1000Research 2017, 6:53

<https://doi.org/10.12688/f1000research.10068.1>

First published: 18 Jan 2017, 6:53 <https://doi.org/10.12688/f1000research.10068.1>

Introduction

Influenza is an infectious disease of global importance and is a target of many near-patient tests^{1,2}. These tests have been criticized for reported low sensitivity. This relatively poor ability to 'rule out' infection has been given as a reason to avoid their use in clinical practice, and instead develop better tests³. There are reasons to suspect some diagnostic-accuracy studies of point-of-care tests (POCTs) may have systematically underestimated sensitivity. If this is the case, the diagnostic accuracy of existing tests may be better than previously thought, with implications for clinical practice and test development.

Classic diagnostic-accuracy studies compare the performance of the index (new) test, with another reference (pre-existing) test, on samples from the same patients. Although rarely explicitly stated, the reference test is assumed to be an infallible 'gold standard'. Under this assumption, whenever the index test and the reference test results differ, the index test is assumed to be wrong. This prevents the index test outperforming the reference, and may systematically underestimate test performance. Many diagnostic-accuracy studies of point-of-care tests for influenza have used these classical methods, raising the possibility that their diagnostic performance have been artificially suppressed⁴.

Established techniques for when a 'gold standard' is not available include: constructing a reference standard by multiple panels of tests, re-testing discrepant results, and statistical modelling⁵. Bayesian latent class models are one such statistical technique^{6,7}. Unlike many other methods, they offer an opportunity to retrospectively analyse existing data, providing a test has been compared to the same reference standard in more than one population⁶. As far as I can tell, this study is the first attempt at Bayesian re-analysis of point-of-care tests for influenza.

This paper aims to examine the extent to which published estimates of influenza point-of-care test accuracy are constrained by the infallible gold standard assumption, with a view to informing clinical practice, and future diagnostic-accuracy studies.

Methods

Published data were re-analysed using Bayesian latent class modelling and classical analysis. Data were extracted from two studies^{8,9} comparing the same reference and index tests (reverse-transcriptase polymerase chain reaction (RT-PCR) vs. QuickVue® influenza A+B influenza), in two separate primary care populations.

Analyses were performed using the free online application Modelling for Infectious diseases CENTre (MICE; <http://mice.tropmedres.ac/home.aspx>), which has been described elsewhere, and runs parallel analyses of Bayesian latent class models and classical frequentist statistics for diagnostic test accuracy⁷. Data are input into MICE via a simple online portal, and results are stored online or emailed to the user.

MICE employs Markov Chain Monte Carlo (MCMC) simulations. These use the data provided to estimate all unknown parameters: the specificity and sensitivity of both reference and index tests, and

the prevalence in the study population(s). The predicted combinations of test results are compared to the actual observed data, and the process is iterated, ideally until the estimates converge on the best fitting values for specificity, sensitivity and prevalence. MICE presents these results in the form of a table, with further graphs of the iterated estimates to allow the user to check convergence of MCMC chains, and Bayesian P values to allow the fit of the final model to the observed data to be assessed.

For this study the 'two tests in one population' model was selected and default values were used. Under the default settings non-informative priors (beta distribution 0.5, 0.5) are used to initiate the analysis, with the specificity of both tests constrained to above 40%. MCMC simulation used default initial values for diagnostic accuracy: (90% and 30% for prevalence, 90% and 70% for sensitivity, 90% and 99% for specificity). The analysis ran for 5000 iterations of pre-analysis adjustment (burn in), and 20,000 iterations.

Results

Data were extracted from two studies comparing a QuickVue® point-of-care test to Reverse-Transcriptase PCR in children with influenza-like-illness. Gordon *et al*⁸ studied 989 children in Nicaragua, Harnden *et al*⁹ included 157 children in England. Patient characteristics and study procedures were similar, with a low risk of bias (Table 1).

MCMC chains converged for all estimates. Model fit, assessed by Bayesian p value, was close between observed and expected values, with the exception of cases positive by RT-PCR, but negative by point-of-care test in the English study: 26 were predicted and 34 observed (Bayesian p value 0.081 – close to 0.5 indicates a good fit).

In both populations estimated prevalence was lower with Bayesian analysis: 34.0% (95% credible interval (CrI) 29.6-40.3) Vs 45.3% (95% CI 41.2-49.5) in Nicaragua and 18.6% (95% CrI 12.7-27.6) Vs 38.9% (95% confidence interval (CI) 31.3-47.0) in England (Table 2).

RT-PCR performance was assumed to be 100% under the classical model, and estimated by Bayesian modelling. Bayesian sensitivity and negative predictive values were close to the assumed values at 98.8% (95% CrI 94.3-100) and 99.3% (95% CrI 96.7-100), but specificity 80.1% (95% CrI 75.9-87.0) and positive predictive value 68.4% (95% CrI 62.0-81.0) were reduced (Table 2).

The performance estimates for the Quick-View® point-of-care test were markedly different under Bayesian assumptions. Sensitivity increased from 66.9% (95% CI 61.4-71.9) to 97.8% (95% CrI 82.1-100). Accordingly, the estimates for negative predictive value also increased, from 79.0% (95% CI 75.2-82.4) to 99.0% (95% CrI 90.7-100). Specificity was more similar between models (Classical 97.8%; 95% CI 95.7-98.9 Vs. Bayesian 98.5%; 95% CrI 96.5-100), as was estimated positive predictive value (Classical 96.0%; 95% CI 92.3-98.0 Vs. Bayesian 96.6%; 95% CrI 92.0-100) (Table 2).

Table 1. Characteristics of the two data sets used for reanalysis.

Study characteristics	Harnden <i>et al.</i> 2003 ^a	Gordon <i>et al.</i> 2009 ^b
Patients		
Criteria for inclusion:	GPs identified "children with cough and fever who they thought had more than a simple cold"	"...fever, or history of fever or feverishness, and cough and/or sore throat within five days of symptom onset..."
Dates:	January to March 2001 and October to March 2002	2008: January 1st to December 31st
Number and setting:	157 children routinely attending primary care in Oxfordshire	989 Children in a cohort study in Nicaragua
Age:	Six months to 12 years, median 3 years	Two to 13 years, mode 3 years
Gender:	100 (63.7%) male	576 (49.8%) male
Tests and procedures		
Index tests:	QuickVue Influenza A+B	QuickVue Influenza A+B
Target condition and reference standard:	Influenza A&B, RT-PCR	Influenza A&B, RT-PCR
Flow and timing:	Nasal swab sample by a research nurse, who undertook POCT immediately, Nasopharyngeal aspirate from other nostril for RT-PCR sent to laboratory within four hours.	Nasal swab for POCT immediately performed, followed by nasal and throat swabs for RT-PCR in central laboratory after storage at 4C for up to 48 hrs.
Risk of bias assessment		
Was a consecutive or random sample of patients enrolled?	Consecutive, from routine clinical practice	Random selection from 3,935 medical visits (of 13,666 in cohort) that met the criteria.
Was a case-control design avoided?	Yes	Yes
Did the study avoid inappropriate exclusions?	Yes	Yes
Could the selection of patients have introduced bias?	Low risk	Low risk
Concerns regarding applicability	Low risk	Low risk
Are there concerns that the included patients and setting do not match the review question?	Low concern	Low concern
Overall risk of bias:	Low risk	Low risk
Citation	Harnden A, Brueggemann A, Shepperd S, <i>et al.</i> Near patient testing for influenza in children in primary care: comparison with laboratory test. <i>BMJ</i> 2003; 326 : 480	Gordon A, Videia E, Saborio S, <i>et al.</i> Performance of an influenza rapid test in children in a primary healthcare setting in Nicaragua. <i>PLoS One</i> 2009; 4 : e7907

Table 2. Estimates of diagnostic accuracy and influenza prevalence using classical and Bayesian latent class models for the QuickVue® influenza A+B Point-of-care test versus the reverse-transcriptase polymerase chain reaction technique for influenza.

Estimates	Classical model (95% confidence interval)	Bayesian latent class model (95% credible interval)
Prevalence		
Nicaragua	45.3% (41.2-49.5)	34.0% (29.6-40.3)
Oxfordshire	38.9% (31.3-47.0)	18.6% (12.7-27.6)
Reverse-transcriptase polymerase chain reaction reference test		
Sensitivity	Assumed 100% (100-100)	98.8% (94.3-100)
Specificity	Assumed 100% (100-100)	80.1% (75.9-87.0)
Positive predictive value	Assumed 100% (100-100)	68.4% (62.0-81.0)
Negative predictive value	Assumed 100% (100-100)	99.3% (96.7-100)
QuickVue® influenza A+B Point-of-care test		
Sensitivity	66.9% (61.4-71.9)	97.8% (82.1-100)
Specificity	97.8% (95.7-98.9)	98.5% (96.5-100)
Positive predictive value	96.0% (92.3-98.0)	96.6% (92.0-100)
Negative predictive value	79.0% (75.2-82.4)	99.0% (90.7-100)

Discussion

The classical results for QuickVue® presented here are typical. A systematic review of all point-of-care tests for influenza reported overall specificity of 98.2% (CI, 97.5% to 98.7%), but sensitivity only 62.3% (95% CI, 57.9% to 66.6%)². In contrast, Bayesian analysis estimated sensitivity of 97.8% (95% CrI 82.1-100), with a negative predictive value of 99.0% (95% CrI 90.7-100), suggesting a test of clinical importance, where there is little room for improvement in the ability to ‘rule out’ infection, apparently answering one of the major criticisms of point-of-care tests³.

The findings suggest false positives by the ‘infallible’ RT-PCR reference test. RT-PCR multiplies nucleic acids exponentially, making it both highly sensitive and vulnerable to false positives. Even the smallest amount of contamination can lead to a false-positive result. This is well recognised, so laboratories often use multiple negative controls¹⁰. Gordon *et al* did not mention negative controls; Harnden *et al* used water.

A weakness is that the original data were not collected specifically for this analysis; there may therefore be differences between in the study conduct by Gordon *et al.* and Harnden *et al.* other than the populations. Despite this, the studies appear to be remarkably similar. The imperfect fit of the data to an element of the Bayesian model should be balanced against classical modelling, where the fit of the data to the ‘perfect’ reference standard is rarely acknowledged, let alone assessed.

Overall, the findings are consistent with higher sensitivity than previously reported, and this underestimation can be attributed to the use of RT-PCR as a ‘gold standard’. These findings have implications for clinical practice, test development, and diagnostic-accuracy studies.

Data availability

Data used in this analysis are from the articles ‘Performance of an influenza rapid test in children in a primary healthcare setting in Nicaragua’⁸ by Gordon *et al.* (available at <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007907>) and ‘Near patient testing for influenza in children in primary care: comparison with laboratory test’⁹ by Harnden *et al.* (available at <http://www.bmj.com/content/326/7387/480>).

Competing interests

No competing interests were disclosed.

Grant information

JJL is a Career Progression Fellow funded by the UK National Institute for Health Research’s School for Primary Care Research (<https://www.spcr.nihr.ac.uk/>).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Hayward AC, Fragaszy EB, Bermingham A, *et al.*: **Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study.** *Lancet Respir Med.* 2014; **2**(6): 445–54.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Chartrand C, Leeftang MM, Minion J, *et al.*: **Accuracy of rapid influenza diagnostic tests: a meta-analysis.** *Ann Intern Med.* 2012; **156**(7): 500–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. World Health Organization: **WHO Public Health Research Agenda for Influenza.** *Public Health.* 2009; 1–18.
[Reference Source](#)
4. Petrozzino JJ, Smith C, Atkinson MJ: **Rapid diagnostic testing for seasonal influenza: an evidence-based review and comparison with unaided clinical diagnosis.** *J Emerg Med.* 2010; **39**(4): 476–490.e1.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Rutjes AW, Reitsma JB, Coomarasamy A, *et al.*: **Evaluation of diagnostic tests when there is no gold standard. A review of methods.** *Health Technol Assess.* 2007; **11**(50): iii, ix–51.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Limmathurotsakul D, Turner EL, Wuthiekanun V, *et al.*: **Fool's gold: Why imperfect reference tests are undermining the evaluation of novel diagnostics: a reevaluation of 5 diagnostic tests for leptospirosis.** *Clin Infect Dis.* 2012; **55**(3): 322–31.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Lim C, Wannapinij P, White L, *et al.*: **Using a web-based application to define the accuracy of diagnostic tests when the gold standard is imperfect.** *PLoS One.* 2013; **8**(11): e79489.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Gordon A, Videa E, Saborio S, *et al.*: **Performance of an influenza rapid test in children in a primary healthcare setting in Nicaragua.** *PLoS One.* 2009; **4**(11): e7907.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Harnden A, Brueggemann A, Shepperd S, *et al.*: **Near patient testing for influenza in children in primary care: comparison with laboratory test.** *BMJ.* 2003; **326**(7387): 480.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Lion T: **Current recommendations for positive controls in RT-PCR assays.** *Leukemia.* 2001; **15**(7): 1033–7.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 06 March 2017

<https://doi.org/10.5256/f1000research.10848.r20688>

© 2017 Cowling B. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Benjamin J. Cowling

WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong (HKU), Hong Kong, China

This is an interesting re-analysis of published data on rapid test sensitivity, making the case that the rapid test might be more sensitive than previously thought, because of inaccuracy of the gold standard.

Major comment

I rated the article "approved with reservations" because I believe the current description of methodology is inadequate. It is not sufficient to refer to an online application for further details. The methods section of this article should provide technical details, for example the likelihood function, and any other information that would be needed for independent replication of the results using a different software package. This information should be sufficient to reveal assumptions that have been made in the model. The basic idea of a Bayesian model for two tests is not complex and can easily be programmed, my point is that the author of this article is responsible for explaining what he has done, and it is not satisfactory to say that the data were plugged into an online tool.

Other comments

I do not oppose the idea of questioning the sensitivity of PCR. There are other studies using serology and PCR which show that many influenza virus infections occur (based on observed rises in antibody titers) but are not detectable by RT-PCR. However, I am not convinced by this work alone that the reason for apparent low rapid test sensitivity is because PCR is an imperfect gold standard. I think the author should propose additional studies that could be done to confirm whether PCR is indeed an imperfect gold standard for the rapid test. I would suggest the author to be more cautious in his enthusiasm for his alternative explanation, in the absence of stronger evidence for that explanation.

I am not sure if the author is aware that other studies of rapid test sensitivity have examined the

viral load in respiratory swabs, and found that rapid test sensitivity tends to be lower in specimens with lower viral load, and specifically the detection threshold seems to be around 1-2 logs higher for rapid tests than for PCR¹. I would view this observation as reasonably good evidence for poorer sensitivity of rapid tests compared to PCR. Wouldn't it be useful to do further analysis of specimens that are positive by the rapid test but negative by PCR? What kind of analysis is suggested?

References

1. Cheng CK, Cowling BJ, Chan KH, Fang VJ, et al.: Factors affecting QuickVue Influenza A + B rapid test performance in the community setting. *Diagn Microbiol Infect Dis*. 2009; **65** (1): 35-41 [PubMed Abstract](#) | [Publisher Full Text](#)

Competing Interests: I have published work on the sensitivity of rapid diagnostic tests including QuickVue.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 10 February 2017

<https://doi.org/10.5256/f1000research.10848.r20109>

© 2017 Tremblay N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nicolas Tremblay

Unité d'immunovirologie moléculaire, Centre de recherche du Centre hospitalier de l'Université de Montréal, Montreal, QC, Canada

In this report, Lee uses a Bayesian latent class model to estimate the specificity and sensitivity of QuickVue Point-Of-Care-Test for Influenza A & B. The author concludes that, based on his retrospective analysis of two studies (Gordon et al, 2009; Harnden et al, 2003), the sensitivity of the POCT test is much higher than expected, in major part due to the gold standard used for comparison.

Major Issues

1. The introduction is an oversimplification of the actual knowledge and match the style of an editorial much closer than an actual overview of the field. Many key points are not addressed:
 - The introduction does not discuss of the most likely cause of heterogeneity of the sensitivity of Influenza POCT (patient age, duration of symptoms, type of specimen, season of

sampling, etc.)

- The introduction does not highlight that a positive test is able to "rule in" an influenza infection which is of clinical significance for therapy initiation, infection control, reduction of ancillary tests, etc.
- The introduction does not report any information on the gold standard test for influenza diagnostic in regards to advantage, limitation and turnaround.
- The usefulness of the approach is not well anchored and raises concerns about the clinical or biological significance or usefulness of the anticipated results
- The rationale for the study is now well established by the information provided and, as such, it is difficult for the reader to understand the logical flow of ideas.
- Many key references are missing

2. The methods section is missing key elements

- There is no information about the data that were extracted from the two selected studies
- There is no information about the methods of inclusion of the two selected studies
- There is no information or reference towards the method used to assess the risk of bias of the two selected studies such as QUADAS-2 or else.

3. The discussion is missing several key elements

- The discussion is missing references for many of the statements made in this section. As an example, "The classical results for QuickVue presented here are typical". The statement should be backed up by a reference and a comparison such as a meta-analysis of POCT.
- There is no discussion about other gold standard test and how POCT performs against them.
- The first paragraph makes a cookie-cutter conclusion that is well outside the scope of the present study.
- The second paragraph does not discuss the limitation of a clinical RT-PCR appropriately. The contamination of a diagnostic sample, as the only example used by the author, is far from reflecting why RT-PCR might be too sensitive, in regards to clinical significance and benchmarking, as a gold standard test for influenza testing.
- The authors does no report or discussion of the limitation of the Gordon et al. and Harnden et al. studies and conclude, without substance, that both study are remarkably similar.
- The last paragraph is an oversimplification of the results presented and does not account for various limitations of the study.

Minor point

- The 2x2 tables of the data extracted from the two selected studies are not provided. It makes it difficult to replicate the results. As an example, Gordon has 2 stratified data sets available based on clinical presentation (Table 2. n=1157 and n=578). It is unclear which data were used.

Conclusions and Recommendations

It is the opinion of the referee that, while the paper raise an interesting idea and novel approach, the paper is of limited significance because it presents major drawbacks and pitfalls. We suggest that the author review the comments made thereof to improve the study design and the overall presentation and content of the paper.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research