Accuracy of rapid influenza diagnostic tests: a meta-analysis
Chartrand C, Leeflang MM, Minion J, Brewer T, Pai M

CRD summary
The authors concluded that influenza can be ruled in but not out by rapid influenza diagnostic tests. Sensitivity was higher in children than in adults and higher for influenza A than influenza B. Unclear sample sizes, potential for missing studies and potential for error and bias in the review process make the reliability of the findings unclear.

Authors' objectives
To evaluate the accuracy of rapid influenza diagnostic tests (RIDTs) in adults and children with influenza-like illness.

Searching
PubMed, EMBASE, BIOSIS and Web of Science were searched for published papers written in English or French. Search dates spanned 1950 to December 2011. Search terms were reported. Bibliographies of included studies and reviews/guidelines were scanned and manufacturers were contacted for further studies and unpublished data.

Study selection
Studies that compared the diagnostic accuracy of RIDTs against reference standards of viral culture or reverse transcriptase polymerase chain reaction (RT-PCR) in all age groups were eligible for inclusion. RIDT was defined as any commercially available assay that identified influenza viral antigens or neuraminidase activity in respiratory symptoms through simple immunochromatographic formats. Case-control studies and conference abstracts were excluded. Specific attempts were made to exclude studies that might have introduced incorporation and partial verification biases (details were reported in the paper).

The included studies used 26 commercially available RIDTs (Binax, Directigen and QuickVue tests were included). The reference standards were generally used with equal frequency across the studies. Just over half of the studies comprised both adults and children. Approximately one third of studies were conducted during the H1N1 2009 pandemic. Only one third of studies gave a definition of influenza-like illness. Where reported, specimens were throat, nasal or nasopharyngeal swabs, washes or aspirates.

One reviewer selected the studies for inclusion. Where there was uncertainty, a second reviewer was consulted and a consensus reached.

Assessment of study quality
Study quality was assessed with Quality Assessment of Diagnostic Accuracy Studies (QUADAS) criteria.
One reviewer carried out the quality assessment. A second reviewer checked a random sample of studies.

Data extraction
Data were extracted into 2x2 tables to enable calculation of sensitivity and specificity and computation of positive and negative likelihood ratios. Data using RT-PCR were extracted preferentially, where both reference standards were used. Tests were classed as point of care only if stated specifically. Authors were contacted for missing data, where necessary.

One reviewer extracted data. A second reviewer checked a random sample of studies.

Methods of synthesis
Sensitivity and specificity were pooled in a bivariate random-effects regression model with 95% confidence intervals (CI). A hierarchical summary receiver-operating characteristic (HSROC) curve was presented. Subgroup analyses were conducted to explore the influences of age and virus type, reference standard, RIDT commercial brand, specimen type, duration of symptoms prior to test, point of care versus laboratory testing and methodological quality. Subgroups that contained at least five studies were included in the bivariate model. Multivariate meta-regression was carried out to explore potentially interrelated variables.
Results of the review

The review included 159 datasets in 119 studies. The number of included participants was unclear. The inclusion criteria largely ruled out partial verification, differential verification and incorporation bias. Reference standards were considered to be appropriate. Approximately 15% of studies reported blinding of interpreters to reference standard and 20% reported blinding to other clinical data. Approximately 15% to 20% of studies reported on withdrawals and/or uninterpretable results. Approximately one third of studies were clear about the basis for patient or specimen inclusion.

Pooled sensitivity was 62.3% (95% CI 57.9% to 66.6%). Pooled specificity was 98.2% (95% CI 97.5% to 98.7%). Positive likelihood ratio was 34.5 (95% CI 23.8% to 45.2%). Negative likelihood ratio was 0.38 (95% CI 0.34 to 0.43). The HSROC demonstrated variation in sensitivity (range 4.4% to 100%), but only 11% of studies showed less than 85% specificity.

Subgroup analyses showed that RIDTs had a significantly higher pooled sensitivity in children (66%, 95% CI 61.6% to 71.7%; 60 datasets) than in adults (53.9%, 95% CI 47.9% to 59.8%; 33 datasets); specificities were similar between the age groups. RIDTs were associated with significantly higher sensitivity for detecting influenza A (64.6%, 95% CI 59.0% to 70.1%; 72 datasets) than for detecting influenza B (52.2%, 95% CI 45% to 59.3%; 27 datasets). Results were unchanged when RIDT brand, specimen type and reference standard were taken into account. Comparing reference standards, RIDTs had a significantly higher pooled sensitivity when compared with viral culture (72.3%, 95% CI 66.8% to 77.9%; 48 datasets) than RT-PCR (53.9%, 95% CI 48.2% to 59.6%; 67 datasets). Further results are reported in the paper.

Authors’ conclusions

Influenza can be ruled in but not out through use of RIDTs. Sensitivity varied across populations and was higher in children than in adults and for influenza A than for influenza B.

CRD commentary

The review question was clear. Inclusion criteria seemed potentially reproducible. Various sources were used to identify relevant studies. Steps were taken to minimise publication bias. Language bias was a possibility. The review process was carried out with minimal attempts to address possible error and bias. Efforts were made to minimise methodological biases in the development of inclusion criteria; an appropriate quality assessment tool was applied and the results of this process were presented.

Study details were tabulated clearly. The chosen method of synthesis appeared to be appropriate in the presence of observed clinical heterogeneity. It was unclear why studies that provided direct comparisons of RIDTs were not explored separately.

The authors’ conclusion and implications for practice reflect the evidence presented. A large number of studies were included, but lack of reporting of sample sizes, potential for missed studies (acknowledged by the authors) in languages other than English and French and potential for error and bias in the review process make the reliability of the findings unclear.

Implications of the review for practice and research

Practice: The authors stated that RIDTs might be a useful prompt intervention at the point of care, provided that clinicians understood its limitations and recognised the need to confirm negative results with other tests.

Research: The authors stated a need for further studies to examine whether and when RIDTs decreased use of ancillary tests and empirical antibiotic treatment and increased appropriate use of antiviral therapy. Cost-effectiveness studies were warranted.

Funding

Canadian Institutes of Health Research.

Bibliographic details

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.