The accuracy of Influenza A (H1N1) "swine flu" laboratory testing: A systematic review of diagnostic test accuracy

Dr Sarahlouise White BSc (Hons) PhD

Student number a1184560

The Joanna Briggs Institute, Faculty of Health Sciences, the University of Adelaide

Sarahlouise.white@adelaide.edu.au

Supervisors:

Dr Tim Schultz BSc, PhD

Australian Patient Safety Foundation, Adelaide, SA 5005

Tim.schultz@apsf.net

Catalin Tufanaru MD, MPH

The Joanna Briggs Institute, Faculty of Health Sciences, the University of Adelaide

Catalin.tufanaru@adelaide.edu.au

Table of Contents

Table of Figures and Tables

Abstract

Background

Influenza A (H1N1) recently became pandemic, highlighting the need for a cheap and accurate diagnostic test to diagnose this virus in a clinically relevant timeframe. The current reference standard (viral culture) requires a significant degree of technical expertise, laboratory time, resources and can take up to 10 days to obtain a result, during which time there could be a significant spread of infection. The objective of this systematic review was to obtain summary estimates of the diagnostic accuracy of currently available laboratory tests compared to viral culture for the diagnosis of Influenza A (H1N1) from respiratory samples.

Search Strategy

Diagnostic tests are still poorly indexed by major databases; therefore the search strategy was deliberately very broad and was conducted during May 2010. A range of databases of both published and Grey Literature were searched, using both Medical Subject Headings and text words. The reference lists of included studies and review articles were also searched for additional studies.

Selection Criteria

Studies that compared the diagnostic accuracy of any laboratory test (index test) compared to viral culture as the reference test were considered for inclusion. The inclusion criteria required each patient to undergo both the index and reference test, and for the tests to both be specific for influenza A (H1N1). Methodological quality was determined using the QUADAS checklist, a validated critical appraisal tool. No studies were excluded on the basis of poor methodological quality.

Search Results

The search identified 3843 potentially relevant studies. Of these 56 full text articles were retrieved for further analysis. Twenty nine relevant articles were assessed with the QUADAS checklist, with 24 being excluded on the basis of incongruence with the review objective or for containing insufficient detail. The remaining five studies examined the diagnostic accuracy of polymerase chain reaction (PCR), and were included in this systematic review.

Methodological Quality

The methodological quality of studies was assessed using the 14 item QUADAS checklist.

Data Extraction

Data was extracted from the included studies using both the QUADAS and STARD checklists. The checklists allowed an assessment of the quality and completeness of the conduct and reporting. Data was collected to determine the accuracy of the index tests. Patients were identified as being either H1N1 positive or negative on the basis of the viral culture results. The index test results were then compared for sensitivity and specificity.

Main Results

Five studies were included, containing a total of 1581 patients. Overall the methodological quality of the studies was moderate, however all suffered from incomplete reporting. Two notable areas of deficit were the reporting of reference test details and specific details related to the specificity of the index test. In addition to insufficient detail, the level of heterogeneity between the reference tests was unknown. As a result of these two factors, combining the test results in meta-analysis of data was not appropriate. Two studies reported accuracy measures and when reported, the sensitivity and specificity of PCR was high. Although unable to generate a summary estimate of the overall accuracy of PCR, the results of this systematic review suggest that PCR using respiratory specimens appears to be highly sensitive and specific at identifying patients with H1N1.

Conclusions

There is limited evidence to suggest that polymerase chain reaction diagnostic tests are accurate in diagnosing H1N1. This systematic review highlights the need for better reporting and indexing of studies that examine the diagnostic accuracy of laboratory tests in general, and H1N1 specifically. This rapidly expanding area of research needs structure and consistency in order for its findings to be useful to health care clinicians and decision makers.

Student declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Sarahlouise White and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed:

Dated: 31 July 2011

Chapter 1   Introduction

1   Introduction to systematic reviews of diagnostic test accuracy

1.1   An introduction to diagnostic tests

Clinicians have long relied upon diagnostic tests for 'evidence' of the presence or absence of a disease or a condition. Similarly, policy makers must evaluate the value of a particular diagnostic test, compare it to any others, and decide which test should be made available or funded.[1] Both clinical and policy decision-makers require a thorough evaluation of the test and its ability to accurately determine who has, and who doesn't have, the disease or condition of interest. Additionally, information on the impact on healthcare providers, therapeutic impact, patient outcome and costs and benefits of the technology should be systematically assessed.[2] Diagnostic tests should be used only if they are cheaper, safer than and at least as accurate as existing methods, if they eliminate the need for further investigations without reduction of accuracy, or if they lead to the institution of effective therapy.[2]

Recent advances in technology and better understanding of the aetiology of disease have led to a huge expansion in the availability of new and exciting diagnostic tests that have the potential to significantly enhance patient outcomes across the world. For example, a fully automated molecular test for tuberculosis case detection and drug resistance that is suitable for developing countries was recently developed.[3] In addition to being more accurate than the standard 19th Century technique of smear microscopy, the new test reduces the time required for diagnosis from six-eight weeks to two hours, allowing treatment to commence much earlier and thereby reduce mortality, secondary resistance and ongoing transmission that can occur in such immuno-compromised patients/subjects.[3,4] Advances in the development of new diagnostic tests have been accompanied by an increased research effort to test the accuracy of new and existing techniques and technologies. For example, a Medline search revealed approximately 2000 diagnostic test accuracy studies published in the period 1966-1970, which increased over eight times to about 17000 in the period 1996-2000.[5]

In general, assessing diagnostic test accuracy is achieved by comparing the results of using the novel (or index) test with those obtained using a reference (or standard) test on the same population of patients or subjects. In many cases, the reference test is referred to as a 'gold standard', however, for the purposes of this text we will use the terminology 'reference test', to prevent potential misconceptions around the infallibility of such tests.

Diagnostic accuracy is routinely reported using the following calculations:

sensitivity – the percentage of patients with the disease that test positive with the index test and

specificity – the percentage of patients without the disease that test negative with the index test. Other more detailed summary statistics from studies of diagnostic test accuracy will be discussed in section 1.2

The term meta-analysis was first credited to Gene Glass in 1976[6] and refers to systematic reviews where data is combined according to pre-specified statistical rules and where relative weight or importance of a study occurs based on mathematical criteria.[7] A meta-analysis allows calculation of a summary estimate or summary of the findings from all of the included studies. This summary provides a more precise estimate of the benefits or harms of the intervention for a given population.

Meta-analysis of diagnostic test accuracy studies is required for a number of reasons. Firstly, results of diagnostic tests, such as laboratory, pathology or imaging studies, often vary greatly between different centres, patient populations and countries. There are many potential explanations for this variation – for example, sample sizes of the studies may be small, or the population sampled is not representative.[1] Additionally, diagnostic test results can differ along the spectrum of disease, or with test interpreters or the results of previous testing.[8] Therefore, calculation of a valid summary estimate based on all available data ,and the factors that affect the summary estimate, to inform clinical and policy decision makers is a key aim of meta-analysis[1, 9] and a properly conducted systematic review of valid diagnostic studies sits atop the hierarchy of diagnostic evidence.[10] Meta-analysis also allows for: examination of whether summary estimates are affected by study design characteristics (i.e. study quality) or characteristics of the patients/subjects or test (i.e. sub-group analysis), and identification of areas for future research.[1, 11]

Methods to synthesise evidence from diagnostic test accuracy studies are now emerging and this text examines the methodological basis to the synthesis of diagnostic test accuracy data and describes the processes involved in the conduct of a diagnostic test accuracy systematic review. Although screening studies share some similarities with diagnostic studies and may report similar statistics, screening is typically applied to uncover very early signs of disease or the risk of disease, whereas diagnostic tests are generally applied to individuals with signs or symptoms of disease. Issues of meta-analysis of screening studies are discussed elsewhere.[12, 13]

Generally, systematic reviews and meta-analyses of diagnostic test accuracy studies have lagged behind reviews of effectiveness in terms of volume, methodological development, quality and uptake by clinicians.[1, 9, 14-17] As an illustration of the low volume of meta-analyses conducted, Chalmers and Lau (1993) found that nearly 9 times as many meta-analyses of randomised controlled trials (RCTs) (n=435) than diagnostic studies (n=50) were published in the medical literature over the time period 1950-1992.[3] Similarly, Irwig et al (1994) had found that in the years 1990 and 1991 only 11 meta-analyses of studies of diagnostic test accuracy were identified following an extensive search of the literature.[1] Over the period 1994-2000, a further 90 systematic reviews of diagnostic test accuracy were identified, although only two-thirds of these used meta-analysis.[18]

The experience of the Cochrane Collaboration and the Joanna Briggs Institute is particularly insightful when considering the low profile of systematic reviews of diagnostic test accuracy. In 2003, Cochrane decided to make provisions for including systematic reviews of diagnostic test accuracy in the Cochrane Library.[8] At the time of writing, there are over 4,500 Cochrane reviews in the Cochrane Library. However, this includes only three systematic reviews of diagnostic test accuracy, with the first – examining galactomannan detection for invasive aspergillosis in immunocompromised patients – published in 2008.[19-21] Although this number is set to grow with some 31 systematic reviews in progress as registered protocols, it is still dwarfed by the 2000 or so protocols of effectiveness reviews currently in progress.[22] The same picture is reflected in the Joanna Briggs Institute Library of Systematic Reviews: with over 200 published systematic reviews, of which two are systematic reviews of diagnostic test accuracy.[17, 23]

Leeflang et al (2008) identified the slow pace of methodological development as one of two main reasons for the underuse of meta-analyses of diagnostic studies.[8] Early work on aggregating the results of diagnostic studies lacked appropriate methodology to combine the results of different studies, and resulted in narrative summaries of results including editorial opinion pieces (see e.g.[24, 25]). Detrano et al's (1988) study sought to evaluate factors that affect the accuracy of exercise thallium scintigraphy for predicting angiographic coronary artery disease by reviewing relevant publications from 1977-1986.[26] Although they presented the sensitivity and specificity for 56 individual studies and used the term 'meta-analysis' to describe their method, they calculated 'total' sensitivity and specificity by pooling the total numbers of patients across all studies, however such separate pooling of sensitivity and specificity is not recommended in meta-analysis.[27]

The slow pace of methodological development and the dissemination of this method to systematic reviewers are again illustrated through examining the experience of the Cochrane Collaboration and the Joanna Briggs Institute. Despite developing initial guidance for systematic reviews of diagnostic test accuracy in the late 1990s, and forming a working group to develop methodology, software, and a handbook; it was not until 2007 that the implementation of systematic reviews of diagnostic test accuracy was officially launched.[28] While the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy does include an important completed chapter on meta-analysis and presenting results[29] at the time of writing many other chapters including interpreting results and selecting studies are still incomplete. Currently, the Joanna Briggs Institute guidance is currently underway and preliminary work has begun on developing computer software to the conduct of DTA systematic reviews.

Additionally, the value of many systematic reviews of diagnostic test accuracy studies is tarnished by the low quality of the included studies.[1, 5, 9, 14] In many cases, few high quality studies have been conducted, or it is not possible to properly assess the quality of included studies due to a lack of appropriate reporting.[30] A number of reviews have concluded that the conduct and reporting of primary diagnostic studies is poor and requires improvement to increase the validity of systematic reviews.[1, 11, 19] To this end, there have been initiatives have been developed which will be discussed in detail later in this text.

Finally, the complexity of methods required to conduct and interpret systematic reviews of diagnostic test accuracy studies (including critical appraisal and meta-analysis) is likely another reason why their uptake has been slow. Reviews may be difficult to understand for clinical researchers or may not provide enough appropriate information.[31] It is hoped that this text can simplify the language and methods of systematic reviews of diagnostic studies to enhance their clinical utility and appeal to potential systematic reviewers.

1.2   Diagnostic test accuracy research

A diagnostic test can broadly be defined as any test that helps a clinician to identify the presence or absence of a particular condition in a presenting patient and thereby provides information with which to develop an appropriate treatment plan. The term "test" refers to any method used to obtain information on the health of a patient, ranging from history and physical examination, laboratory tests, functional tests, imaging tests and histopathology. The condition of interest or target condition can refer to a particular disease or to any other identifiable condition that may prompt further clinical action; such as further diagnostic testing, or the initiation, modification or cessation of treatment.[32]

Diagnostic tests are critical components of effective health care as they can allow distinction between similar conditions; however, for a diagnostic test to be useful there must be some beneficial treatment that results from detecting the problem or condition. The basic underlying assumption for utilising diagnostic tests is that early detection leads to better care and ultimately better health outcomes for the patient.[33]

The world of diagnostic tests is dynamic, with new tests being developed, as well as the technology of existing tests being continuously improved. Improvements in the diagnostic test industry are driven by the demand for faster, more cost effective, less labour intensive tests– all of which without compromising patient safety or test accuracy. The ideal diagnostic test is one that aims to accurately identify patients with the target condition and exclude those without the condition within a timeframe and at a cost that allows effective decision making.[7, 29, 34] Often there are several tests for a particular condition, and with competing manufacturers claiming that their test is the best – how does a laboratory choose which test to use? The decision as to which diagnostic tests a particular laboratory uses will depend of several factors, often best judged by the senior medical and/or clinical scientists with expertise in the field. Such factors will include: time taken to provide results, cost of the test, technical requirements needed to conduct the test, in addition to its overall diagnostic test accuracy.

1.3   Diagnostic test accuracy

Primary studies that examine the test performance (as determined by sensitivity and specificity) are referred to as diagnostic test accuracy (DTA) studies and these studies compare a "new" test (or tests) to the best test (or method) that is currently available. The new test(s) is known as the index test and may be marketed as being more accurate, cheaper, faster or less invasive than currently available tests. The index test is compared to a test (or method) that is regarded as being the best in terms of accurately identifying the condition of interest. This test is called the reference test. The term "gold

standard is often used, however this term is misleading as it incorrectly implies that the reference test is perfect, which is unlikely to be the case.[35] The concept that no test is perfect is an important one as there are sure to be situations when an index test outperforms and subsequently replaces a reference test as diagnostic test methods and/or technology develop.

In studies of diagnostic test accuracy, the outcomes from one or more tests under evaluation are compared with outcomes from the reference test, both of which are measured in subjects who are suspected of having the condition of interest.[32] In this framework, the reference test is considered to be the best available method for establishing the presence or absence of the condition of interest and may consist of a single test or a combination of tests and can include laboratory tests, functional tests, imaging tests and pathology.

In the context of a DTA study, the term accuracy refers to the amount of agreement between data from the index test and the reference test. As the reference test is considered to correctly identify only condition positive patients, a positive result using this test, is considered as being a "true positive" and to occur for patients that truly have the condition of interest. The same applies with respect to a negative reference test result and patients who receive a negative reference test result are considered to be "true negatives" and to be condition-free.

The accuracy of the index test is reported relative to the reference test in terms of test sensitivity – how well does the index test correctly identify those as having the condition?, and test specificity – how well does the index test exclude those who do not have the condition of interest? Patients who receive a positive index test result but a negative reference test result are classified as being "false positive" and those who receive a negative index result and a positive reference test result are considered to be "false negative" due to the disagreement between the test results. Patient/subject classification is summarised in Table 1 below.

Table 1   Description of patient classification for diagnostic test accuracy studies

Patient classification	Description of test results
True Positive	Positive Index Test Result Positive Reference Test Result
True Negative	Negative Index Test Result Negative Reference Test Result
False Positive	Positive Index Test Result Negative Reference Test Result
False Negative	Negative Index Test Result Positive Reference Test Result

There is consensus [7, 29, 34] that test accuracy results should be presented as a 2x2 table, such as Table 2 which is adapted from Macaskill et al.[29] A patient (or subject) should appear in only one cell of the table. Once the table has been populated, the figures can then be used to assess the performance of the index test, using measures such as: sensitivity, specificity, likelihood ratios, and predictive values. In order to generate a 2 x 2 table, the data should be dichotomised into positive and negative using a predetermined cut-off value, for each test. The value at which a test result becomes positive will have obvious implications for the number of patients in each category (the sensitivity and specificity for that test) so should it be chosen with caution. This is an important concept - cut-off value that is inappropriately high will wrongly exclude some patients that have the disease/condition and therefore falsely decrease test sensitivity. A cut-off value that is inappropriately low will wrongly classify some patients that do not have the disease/condition and therefore falsely decrease test specificity. In situations where there could be several logical cut-off points, it may be sensible to calculate the sensitivity and specificity at each of those thresholds and present them graphically as a receiver operator curve (ROC) to allow examination of how changing the threshold alters the findings.[7, 34]

Table 2   A typical 2x2 table to classify patient test results and disease status

Test Outcome (Index test results)	Disease/condition status (Reference test results)		
	Disease/Condition positive	Disease/Condition negative	Total
Index test positive	True positives (a)	False positives (b)	Test positives (a + b)
Index test negative	False negatives (c)	True Negatives (d)	Test negatives (c + d)
Total	Disease/condition positives (a + c)	Disease/condition negatives (b + d)	N (a + b + c + d)

The sensitivity and specificity of a diagnostic test can be calculated, using the patient/subject classifications from Table 2, using in the equations:[29]

$$Sensitivity = \frac{a}{a + c}$$

$$Specificity = \frac{d}{b + d}$$

Once the test sensitivity and specificity have been calculated, these measures of test accuracy can be reported in a number of ways that convey relevant and practical information to the clinician about what the test result is likely to mean for the patient. Frequently used measures are predictive values (positive and negative) and likelihood ratios (positive and negative).

Positive and negative predictive values (PPV and NPV) are used to assess the usefulness of a result, once the test result is known. For example, a PPV of 95% indicates that 95% of patients who received a positive test result actually had the condition or disease of interest. A NPV of 60% indicates that 60% of patients that received a negative test result were truly disease/condition free. The equations used to calculate PPV and NPV are below.

Positive Predictive Value

$$PPV = \frac{TP}{TP + FP}$$

Or, by using the patient classifications from Table 2 this can be written as:

$$PPV = \frac{a}{a + b}$$

Negative Predictive Value

$$NPV = \frac{TN}{TN + FN}$$

Or, by using the patient classifications from Table 2 this can be written as:

$$NPV = \frac{d}{c + d}$$

A ratio of test sensitivity and specificity is often reported as a likelihood ratio (LR). The likelihood ratio describes the probability or likelihood that the test result would be expected in a person with the condition, compared with the probability or likelihood that the same result would be expected in a person without the condition.[11] Likelihood ratios can be both positive and negative values. A positive likelihood ratio (+ve LR) and can be expressed as; people with the condition are X times more likely to receive a positive test result that those who are well.[36]

A negative likelihood ratio (-ve LR) can be expressed as; people with the condition are X times more likely to receive a negative test result that those who are well.[36] Mathematically this can be expressed by the equations:

Positive Likelihood Ratio

$$+ve\ LR = \frac{Sensitivity}{1 - specificity}$$

Or, by using the patient classifications from Table 2 this can be estimated as:

$$+ve\ LR = \frac{a/(a + c)}{b/(b + d)}$$

Negative Likelihood Ratio

$$-ve\ LR = \frac{1 - sensitivity}{specificity}$$

Or, by using the patient classifications from Table 2 this can be estimated as:

$$-ve = \frac{c/(a + c)}{d/(b + d)}$$

1.4   Diagnostic test accuracy evidence and healthcare

The evidence base for diagnostic test accuracy is growing increasingly being used in healthcare to compare the performance of different tests in the diagnosis of specific conditions, where several tests exist for a condition. Over the past decade, estimates of publications related to diagnostic test accuracy have greatly increased. A cursory search of PubMed (June 2011) for "diagnostic test accuracy" revealed that between 2001 and 2011 over 5000 (5612) publications. Less than half of this total (1718) was available between 2001 and 2005, implying that most of the publications occurred in the later part of the decade.

Such increases in diagnostic tests accuracy studies portray the quest to make available to the clinician the appropriate tools to serve the patient/subject, who in turn will be assured of receiving prompt quality treatment.

Direct comparison of diagnostic tests provides crucial information about the accuracy of the test in terms of correctly identifying "condition positive" patients/subjects (sensitivity) and correctly identifying "condition negative" patients/subjects (specificity). Correctly determining how well the test performs at identifying whether the patient has the condition of interest or not is important for clinicians to know as exaggerated and/or biased results from poorly designed or reported DTA studies could lead to a premature uptake of a test and lead to clinicians making incorrect diagnoses, followed by subsequent mistreatment.[32] A rigorous evaluation of diagnostic tests before their introduction into clinical practice would not only reduce the number of unwanted clinical consequences (such as incorrect diagnoses and subsequent mistreatment), related to misleading estimates of test accuracy, but also reduce healthcare costs by preventing unnecessary testing. Studies to determine the diagnostic accuracy of a test are a vital part of this evaluation process.[32]

The drivers for improvements in the diagnostic test industry include;

Faster Results
The need for a test to be able to provide results in a short timeframe to enable the clinician to confer a diagnosis so that a patient may begin a course of treatment more quickly, increasing the likelihood of positive patient outcomes. A short time between conduct of a diagnostic test and availability of the results is also important in situations where there may also be public health concerns, such as pandemic conditions.

Cheaper Tests

Finding less expensive methods to enable a clinician to diagnose a patient is beneficial for healthcare budgets in general – as long as test accuracy is not compromised. This becomes especially important in regions that are resource poor (such as developing countries) or in situations where many patients require testing (such as during a pandemic).

Easier Tests

There are less likely to be operational/technical errors with tests that are simple and easy to use. Ease of use is of benefit not only to those conducting and interpreting results of a diagnostic test, but also increases patient safety and confidence.

All of these drives seek to improve diagnostic tests without decreasing the accuracy or reliability of the results. Most diagnostic accuracy studies are, in principle, cross-sectional studies. Delayed- cross-sectional studies are those in which results of indicator tests are decided sometime after conclusions are made on the condition of the subject.

1.5   Systematic reviews in historical context

Summarising the findings of several studies on the same topic, using similar measures in a knowledge synthesis or systematic review is not a new concept. In the 1960s, reviews became popular in many areas of research such as psychology, education and the social sciences.[37] With the recognition of the importance of evidence based decisions in healthcare, such work has grown in popularity with clinicians and researchers alike as a way of summarising salient information. There are many formats that a review of currently available literature may take and some of those most often used in healthcare are discussed in a recent paper.[38]

Traditionally, summarising a field of research to address whether a treatment was effective or not was the domain of experts in that field who would read studies that addressed a particular question, summarise the findings of those studies and then arrive at a conclusion regarding effectiveness.[8, 37] This type of review is often called a narrative or a literature review and often provides little detail on how the information was gathered, assessed and summarised.[38] Systematic reviews represented a substantial improvement in the transparency of the review process.

Since the 1990's this type of review has become more popular (as determined by the increased numbers indexed by commercial databases) in healthcare research for its systematic approach toward literature searching, critical appraisal of studies to be included in the review, data extracted from included studies and methods of data synthesis,[7, 37] together with its overall transparency. The systematic review differs from a narrative review in several key ways - from deciding on the review question, searching for studies, to extraction and synthesis of data, as discussed recently.[15] The main points of this paper are summarised in .[39]

Table 3. Due to the rigorous way in which systematic reviews are conducted and reported, they provide a solid evidence base on which to make healthcare decisions – attempting to bridge the gap between research and decision making.[39]

Table 3   The main differences between a narrative and a systematic review.

Title	The title of a systematic review generally contains explicit information regarding the PICO (Participants, Intervention or phenomenon of interest, Comparator(s) and Outcomes), and has a clear question to be answered or hypothesis to be tested. The narrative review may also start with a clear question; however there is more often a general discussion about what is known on the subject with no hypothesis.
Search Strategy	A systematic review aims to locate sources of information, relevant to the review question. In addition to those published in commercial journals, attempts are made to limit publication bias by searching sources of grey literature such as theses, conference proceedings and reports. In comparison, there is generally no clearly defined search strategy presented in a narrative review.
Inclusion Criteria	A systematic review will generally define the criteria on which studies will be considered for inclusion – often in terms of participants, intervention (or phenomena of interest), Comparator and Outcome(s) of interest – abbreviated to PICO. There are generally no explicit inclusion criteria for a narrative review.
Critical Appraisal	A systematic review clearly defines how studies will be included (and excluded) in the review. Standardised critical appraisal tools/checklists are often used to assess how well the study has addressed areas of concern or interest (such as different types of bias) in an effort to assess the methodological quality. How methodological quality is determined is not usually explicit in a narrative review.
Data Extraction	The data to be extracted is predefined and is congruent with the outcome measures for a systematic review. Standardised tools/checklists are often used; however, this is not generally the case for a narrative review.
Data Synthesis	A systematic review should predefine how data will be summarised, on the basis of the type of data anticipated (e.g. dichotomous or continuous). Details are provided on how the data will be combined statistically if appropriate or narratively where there is sufficient heterogeneity. This is not generally the case for a narrative review.

The systematic and transparent approach to knowledge synthesis and systematic reviews offers advantages over the traditional approach, such as decreased subjectivity and transparency of how the conclusions were reached.[7] In narrative reviews for example, review authors may use different criteria about which studies to include in the review. Once included studies have been decided upon, authors may place different credence or value upon certain studies (based on items such sample size or method used). The process of how data is combined and how conclusions are drawn is most often not clearly explained in narrative reviews.

As summarised in Table 3, with the systematic review approach, the criteria for which studies are to be included and excluded are clearly defined in advance. The data to be extracted and how that data will be analysed are also clearly detailed in advance. These are clearly major advantages over a traditional literature, providing greater transparency and allowing a logical progression from the development of the review question to the synthesis of results. A systematic review may or may not include statistical combination (meta-analysis) of data from primary studies, depending on the data.

For data to be combined in meta-analysis, certain criteria need to be fulfilled. The first is that the data is appropriate to combine in terms of where it came from (e.g. similar participant population, settings and study design) and how it was collected (e.g. comparable methods and outcome measures). The second criterion is that is it useful to do so. Even if the study designs and patients are comparable, there may be no benefit of combining data from primary studies where the data gives results in opposing directions.[40]

To satisfy whether it is appropriate to combine data statistically, clinical, methodological and statistical homogeneity should be demonstrated as defined below.

• Clinical Homogeneity – Are the participants of the included studies similar in terms of age, co-morbidities, disease state and medications? If so, then it may be appropriate to make a summary estimate and generalisations for this population.

• Methodological Homogeneity – Do the included studies utilise similar study designs and methods? Only studies that use similar methodologies can be combined as how the study is conducted determines how, when and what data is collected.

• Statistical Homogeneity – Do the included studies measure similar outcome measures, using similar scales? Statistical tests such as Chi square and I^2 can be used to determine the degree to which numerical data is homogenous.

1.6 Systematic review of diagnostic study data

The main objective of a DTA systematic review is to summarise the evidence on the accuracy of a test in comparison to an appropriate reference standard. Accuracy in this context is measured by sensitivity, specificity and may also include predictive values and/or likelihood ratios. Other objectives of DTA systematic reviews may include areas of scope, such as those suggested by Zhuo et al.[34] Identify the number, quality and scope of primary research studies that examine diagnostic test accuracy

- Summarise the diagnostic test accuracy across reported studies
- Determine whether there is heterogeneity among the accuracy values across studies
- Examine the relationship of diagnostic test accuracy to study quality, patient and/or test characteristics
- Compare the diagnostic test accuracy of related tests, increasing the statistical power to detect differences in accuracy over individual comparative studies
- Examine the relationship between test comparisons and study quality, patient characteristics and test characteristics
- Identify directions for future research

A synthesis of data from well reported, high quality studies is an appropriate way of comparing the performance of individual diagnostic tests and can provide a useful way of drawing together studies that utilise the same test on a similar patient population. Aside from the actual diagnostic test results, it is also important to extract sufficient information to determine whether or not there is sufficient homogeneity (e.g. in methods, cut-off points and measurements made) to combine the data statistically

in meta-analysis.[34] This can be hampered by poor reporting of details such as patient characteristics, test methods and what constituted a positive result (cut-off value).

When index test sensitivity and specificity are calculated relative to the reference test, the assumption is being made that the reference test identifies only those with the disease/condition; therefore it is crucial that an appropriate reference standard is used and that sufficient detail is provided to ensure that the tests are similar enough that combination of the results makes sense and is appropriate. As discussed previously, assessing the diagnostic accuracy of a test is complex and problematic. As a result, comparing and combining such studies in meta-analysis presents its own series of challenges, which have been topics of previous discussion[8, 41-44] and although there has been progress in some areas, several issues remain as discussed in the following sections.

1.7 Challenges of undertaking systematic reviews of diagnostic test accuracy

Some challenges encountered in undertaking systematic reviews of diagnostic test accuracy data are discussed in the following sections:

1.7.1 Searching for and identifying relevant primary research

In areas of research where they are consistently applied, easily identifiable terms – such as randomised controlled trial, specific search filers for electronic databases have proved to be useful at reducing the number of papers to screen, whilst not losing relevant papers.[8, 45] Usefulness of such filters relies on the correct indexing of terms related to methodology and text words used in reporting the results. An extensive literature is available on designing search strategies to identify therapeutic intervention studies, however the corresponding body of literature on diagnostic test search strategies, is currently relatively small, therefore the recommendation is not to use search filters that are based on single search terms alone method.[12, 45, 46]

Identifying primary research that reports test performance is one challenge that authors face when assessing diagnostic test accuracy.[47-49] There is no unequivocal keyword or indexing term for an accuracy study in the literature databases. Use of the Medical Subject Heading (MeSH) "sensitivity and specificity" may look suitable but it is inconsistently applied by the major electronic databases.[8, 47-49]

Recent work on search filters has shown that although indexing of diagnostic accuracy studies has improved in PubMed, there are still no consistently applied terms that will identify as many studies as a comparable unfiltered search of major databases.[48] Until indexing systems consistently index studies of diagnostic test accuracy, searching such databases will remain challenging and may require additional manual searches, such as screening reference lists.[8, 44] An additional factor that may also hamper study identification is that data on the diagnostic accuracy of a test may be hidden within studies (such as analytical sensitivity or clinical trials) that did not have test accuracy as their primary objective.[8, 45]

Manual (or hand)searching through publications, reference checking, and searching for unpublished reports, also increases the likelihood of identification of relevant studies, especially to assess the extent of publication bias. Finally, it is important to document search activity and the outcome of each search, in order to maintain transparency – a crucial feature in the credibility of the systematic review process.[12, 13, 45]

The overall consensus is that a search for diagnostic studies must be comprehensive, objective, and reproducible.[50] Where possible, the search should not only be restricted to commercially operated databases and journals but include other publications, such as conference proceedings, theses and dissertations (Grey literature), to ensure the maximum breadth and odds of identifying relevant studies.

1.7.2   Variations in study populations

Diagnostic tests performed in study populations located in different settings and with different characteristics could provide results that differ from each other by virtue of the differences inherent in those populations.[41] For example, the clinical manifestations of malaria (fever, lassitude/lethargy) among non-immune persons (for example children or persons not resident in a malaria endemic region) may be far more severe (and therefore perhaps more readily detectable) in comparison to those with immunity, such as adults residing in a malaria endemic region.[51] As such when studying manifestations of clinical malaria, it would be appropriate to compare outcomes among similar populations of immune and less immune subjects in their defined settings, enhancing the homogeneity of outcomes whilst reducing the level and influence of heterogeneity. This would ultimately facilitate comparison and interpretation of studies.[41]

1.7.3 Publication bias

It has been well documented that studies with significant results are more likely to be reported than those with non-significant findings, this is known as publication bias.[52] One way of assessing the size of publication bias within a research area is to compare the total number of studies undertaken (as determined by the number of studies approved by ethics review committees and research boards), with the number of studies that publish results. A second is to search for studies that are published in non-commercially published journals, conference proceedings, monographs and other forms of Grey Literature.

Diagnostic studies with poor test performance results that are not published may lead to exaggerated estimates of the true sensitivity and specificity of a test in a systematic review.[41] Attempts at addressing this publication bias has been made for randomised controlled trials, and several visual and statistical methods have been proposed to detect and correct for unpublished studies.[41] One solution to the problem of publication bias is the mandatory registration of all clinical trials before patient enrolment and for therapeutic trials; considerable progress has already been made in this area. A clinical trials registry has been suggested for clinical outcomes of diagnostic tests[41] however this would prove to be problematic for diagnostic studies, as often there is no ethical review or study registration; therefore tracking of studies from registration to publication status is not currently possible.[8] DTA studies are often carried out on patient specimens that have served their purpose and are no longer required; therefore an ethical review is unnecessary.

Approaches to identify and minimise publication and retrieval bias could include the following:[53, 54]

1.   Pre-empting publication bias before a systematic review (e.g. prospective registration of trials)

2.   Reducing retrieval bias in the course of a systematic review (e.g. comprehensive search for primary studies from diverse sources in diverse languages, hand (manual) searching of journals, contacting topic experts and checking reference lists of relevant publications. The non-commercially published body of literature such as government reports, theses, and University websites (so called Grey literature) may also be a useful resource.

3.   Detecting and estimating publication bias in the course of a systematic review (e.g. funnel plots, sensitivity analysis modelling)

4.   Minimising the impact of publication bias after a systematic review (confirmatory large scale trials and subsequent update of the systematic review)

Decisions on publication bias in DTAs should not be based on funnel plot based tests in view of the fact that they may provide misleading results and alternatives created for DTAs end up being poorly powered.[8, 11]

1.7.4 Assessing methodological quality of diagnostic studies

The quality of a study is determined by its design, methods by which the study sample is recruited, the conduct of tests involved, blinding in the process of interpreting tests and the completeness of the study report. The reliability of a systematic review is greatly enhanced by the use of studies of the highest possible quality.[14, 55] The terms "Assessing methodological quality", "assessing study validity" and "assessment of risk of bias" are synonymous and describe techniques used to evaluate the extent to which the results of a study should be believed or to be deemed valid after rigorous assessment.[55]

The quote "the extent to which all aspects of a study's design and conduct can be shown to protect against systematic bias, non-systematic bias that may arise in poorly performed studies, and inferential error"[41, 56] aptly describes factors contributing to the methodological quality of a study. Biases in study outcomes may be in the form of small study sample sizes that do not appropriately represent the population being studied, non-random sampling of the study sample and observers not blinded to the outcomes of the study.[41] The standards for reporting of diagnostic accuracy (STARD) 25-item checklist and the quality assessment of diagnostic accuracy studies (QUADAS) 14-item checklist have been devised to improve on the quality of reporting of primary diagnostic studies and the assessment of primary DTA studies included in diagnostic systematic reviews respectively.[57, 58]

Diagnostic test accuracy studies with design deficiencies can produce biased results, leading to misinformation regarding the accuracy of the test under evaluation. Other factors resulting in bias within diagnostic test accuracy studies include differences in study populations and small sample sizes.[8] Table 4 is modified and expanded from[8] and presents the major types of bias that can occur in studies and result from flawed or incomplete DTA reporting.

Table 4   Major types of bias that result from incomplete reporting in DTA studies

Type of bias	When does it occur?	How does bias affect the diagnostic test accuracy assessment?	Steps that can be taken to address or prevent type of bias
Patients/Subjects Spectrum bias	When included patients do not represent the intended spectrum of severity for the target condition or alternative conditions	Depends on which end of the disease spectrum the included patients represent	Ensure that the included patients represent a broad sample of those that the test is intended for use with in clinical practice.
Selection bias	When eligible patients are not enrolled consecutively or randomly	Usually leads to overestimation of accuracy	Consider all eligible patients and enroll either consecutively or randomly
Index test Information bias	When the index results are interpreted with knowledge of the reference test results, or with more (or less) information than in practice.	Usually leads to overestimation of accuracy, unless less clinical information is provided than in practice, which may result in an under estimation of accuracy	Index test results should be interpreted without knowledge of the reference test results, or with more (or less) information than in practice.
Reference test Misclassification bias	When the reference test does not correctly classify patients with the target condition	Depends on whether both the reference and index test make the same mistakes	Ensure that the reference correctly classifies patients within the target condition.
Partial verification bias	When a non-random set of patients does not undergo the reference test	Usually leads to overestimation of sensitivity, effect on specificity varies	Ensure that all patients undergo both the reference and index tests.
Differential verification bias	When a non-random set of patients is verified with a second or third reference test, especially when this selection depends on the index test result	Usually leads to overestimation of accuracy	Ensure that all patients undergo both the reference and index tests.
Incorporation bias	When the index test is incorporated in a (composite) reference test	Usually leads to overestimation of accuracy	Ensure that the reference and test are performed separately
Disease/Condition progression bias	When the patients' condition changes between administering the index and reference test	Under – or over estimation of accuracy, depending on the change in the patients' condition	Perform the reference and index with minimal delay. Ideally at the same time where practical
Information bias	When the reference test data is interpreted with the knowledge of the index test results	Usually leads to overestimation of accuracy	Interpret the reference and index data independently
Data analysis Excluded data	When uninterpretable or intermediate test results and withdrawals are not included in the analysis	Usually leads to overestimation of accuracy	Ensure that all patients who entered the study are accounted for and that all uninterpretable or intermediate test results are explained

Attempts have been made to improve methodological quality and diagnostic test study reporting as mentioned previously. The development of tools to aid primary researchers address and avoid sources of bias such as those listed in the table, as it is only with complete and accurate reporting that research becomes transparent and its internal and external validity can be assessed. Current approaches to quality assessment of diagnostic studies include using a validated checklist instrument. This approach allows an author to work through a pre-defined list of items, with the overall aim of minimising bias and improving the quality of their report. The checklist published by the STARD initiative [32, 59] is one such instrument aimed at primary researchers conducting and reporting DTA studies.

For those interested in comparing findings from individual DTA studies (such as systematic review authors), assessing the quality of a diagnostic study requires consideration of several features of both study design and conduct, including factors such as definition of the research question and clinical context, specification of appropriate patient population, description of the diagnostic techniques under study and their interpretation, detailed accounting of how the reference standard information was defined and obtained, and any other factors that can affect the integrity of the study and the generalisability of the results.[12, 13, 45]

A checklist approach can prove convenient for those comparing DTA studies, as the included studies are compared against a checklist of predefined items, noting their presence or absence, as a way of assessing key qualities. A significant development in this area was the quality assessment tool for diagnostic accuracy (QUADAS).[57, 60] The QUADAS tool (Appendix III) is a rigorously constructed checklist of 14 items that can be used by investigators undertaking reviews of diagnostic test accuracy.

1.7.5 Assessing study quality using the QUADAS tool

Information regarding the methodological quality of a study may be collected in 5 stages:[55]

1. Choosing the quality items to be assessed and developing guidance for the assessment based on the requirements of the review

2. Development of an assessment form

3. Testing and refinement the assessment form

4. Collection of data from the identified studies

5. Resolution of any disagreements or to seek clarification for data that is ambiguous or missing

Assessing the methodological quality of a study is usually performed by two independent reviewers and the process of resolving disagreements should be clearly defined in the protocol of the review. The response options to each QUADAS checklist item should be "Yes", "No" or "Unclear". "Yes" implying that the methodological feature is optimal; "No" meaning that the methodological feature is less than optimal with the potential of introducing bias or limiting its applicability. Other checklist items may be added to assess methodological quality, examples are shown in (Appendix VI) depending on the nature of the review in question.[55]

The use of a checklist allows an overall view of study quality to be gained and a quality score to be generated. This assessment can then be used to rank or exclude the study from the review. It was however observed that the assessment process of generating scores produced different rankings and different conclusions by different groupings. As such it is advised to use quality assessment as a means of investigating associations between the individual quality items of the QUADAS tool and estimates of diagnostic accuracy instead of arriving at quality decisions based on a summary quality score.[50]

1.7.6   Extracting and combining data from diagnostic test accuracy studies

A further challenge that faces authors of DTA systematic reviews is how to handle and sensibly combine the data from studies included in the review.[8, 12, 42] Systematic reviews of diagnostic test accuracy are concerned with test results that can be presented in different formats[7, 29, 34] as summarised in Table 5   Typical formats for data in DTA studies.

Table 5   Typical formats for data in DTA studies

Dichotomous	test results are reported as yes/no or positive/negative
Continuous	test results are reported on a continuous scale or as a count (e.g. a concentration of a substance, number of features, mg, kg etc.)
Ordinal	test results are reported as a set of numbered categories (e.g. 1= definitely normal, 2= probably normal, 3=equivocal, 4= probably abnormal, 5=definitely abnormal

However the test results are presented, it is important for their interpretation, that what constitutes test positive and test negative results is clearly defined.

1.7.7 Calculation of summary estimates of diagnostic test accuracy

The outcome measures reported in DTA systematic reviews are summary measures of test accuracy: sensitivity, specificity, likelihood ratios and receiver operator curve information. When combining diagnostic test accuracy data in meta-analysis, there are several factors to consider.

Table 6 is derived from recent Cochrane Collaboration guidance[29] and summarises the major challenges faced when undertaking systematic reviews of diagnostic test accuracy, as compared with systematic reviews that examine effectiveness of therapeutic interventions.

Table 6   Considerations when undertaking meta-analysis of DTA studies

Consideration	Implications
Different data is extracted for DTA reviews than for systematic reviews of intervention effectiveness.	Evaluating DTA requires knowing both the sensitivity and specificity of a test, whereas data from intervention studies generally only have a single measure to consider (e.g. difference in means or risk ratio). This requires reporting of True Positive, True Negative, False Positive and False Negative categories.
Greater heterogeneity in data is expected in DTA reviews than for reviews of intervention effectiveness	Heterogeneity is to be expected in the results of a DTA meta-analysis, therefore a random effects model should be used to describe the variability across the included studies
As two measures need to be considered when determining test accuracy, DTA reviews require more sophisticated statistical methods than effectiveness systematic reviews	A meta-analysis has to allow for the trade-off that occurs between sensitivity and specificity in studies that vary in the threshold used to define test positive and negatives Statistical methods that consider sensitivity, specificity, the relationship between them and the heterogeneity in test accuracy across studies are complex and therefore require specialised statistical knowledge and sophisticated computer software

Briefly, the need to consider both sensitivity and specificity when assessing the accuracy of a diagnostic test makes data extraction and analysis more complex than for systematic reviews that examine therapeutic effectiveness.

Chapter 2   Introduction

2   Introduction to Influenza A H1N1 "Swine flu"

2.1   Influenza Viruses

The Influenza virus affects mainly the nose, throat, bronchi and, occasionally the lungs. Infection usually lasts for about a week and is characterised by sudden onset of high fever, aching muscles, headache and severe malaise, non-productive cough, sore throat and rhinitis.[61] The virus is transmitted easily from person to person via aerosol droplets and small particles produced when infected people cough or sneeze. Influenza tends to spread rapidly in seasonal epidemics. Most infected people recover within one to two weeks without requiring medical treatment. However, in the very young, the elderly, and those with other serious medical conditions, infection can lead to severe complications of the underlying condition, pneumonia and death.[61, 62]

In their simplest form, viruses are composed of an outer protein coat and an inner core of either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), however, unlike bacteria, no virus has been shown to contain both[63] indicating they are unable to transcribe DNA to RNA, thus requiring host cells. A virus particle is present as either in either a replicating or defective form within the infected cell. The extracellular virus is defined as a virion.

Influenza viruses belong to the family of Orthmyxoviridae. This family of viruses have a single stranded, negative sense, eight-segmented RNA genome [64] The eight segments of RNA encode 11 proteins, of which nine are assembled into the infectious virion.[65] Three proteins are found on the surface of the virion, haemagglutinin (HA) and neuraminidase (NA), as well as the transmembrane ion channel protein (M2).

The Orthmyxoviridae family is comprised of five genera:
- Influenza virus A
- Influenza virus B
- Influenza virus C
- Isavirus
- Thogovirus

This family of viruses are enveloped by a lipid envelope that is derived from proteins synthesised by the host. The lipid envelope confers some protection to the virus, not only from host immune cells but also from environmental factors. Influenza A, B and C viruses differ in the host species used and their pathogenicity[66] and can be distinguished on the basis of antigenic differences between proteins on their outer protein coat.[67] Of the influenza viruses, only A and B cause frequent (and occasionally severe) diseases in humans.[64] There is only one type of Influenza B, whereas Influenza A has multiple subtypes, based on a combination of genes that encode for surface proteins,[64] discussed further below.

A single extracellular Influenza A virus particle (virion) has a characteristic spiky appearance when viewed under electron microscopy.[68] The spikes are due to glycoproteins that protrude beneath the lipid envelope. The major glycoproteins are Haemagglutinin and Neuraminidase. A minor component of the lipid envelope is the M2 protein which acts as an ion channel. The matrix protein (M1) lies beneath the lipid envelope and covers the ribonucleoprotein (RNP) complexes. The RNP complexes are comprised of the viral RNA, covered with nucleoprotein and associated with other protein complexes. The viral genome consists of eight separate RNA segments and the segments 1, 2 and 3 code for viral polymerase i.e., PB2, PB1, and PA respectively, segment 4 for haemagglutinin (HA), segment 5 for Nucleoprotein (NP), segment 6 for neuraminidase (NA), segment 7 for matrix protein (M1), and membrane channel proteins (M2), segment 8 for non-structural proteins (NS1 and NS2).[69] The sketch in Figure 1 is a sketch showing the main features of a typical influenza A virus.

NOTE:
This figure is included on page 33
of the print copy of the thesis held in
the University of Adelaide Library.

Figure 1   Schematic influenza A virion[69]

The category of Influenza A viruses is further subdivided on the basis of haemagglutinin (H) and neuraminidase (N) proteins expressed on the outer surface of the virus[66] Both of these proteins are important for the virus to be able to get into the host cell, reproduce and be released to infect other cells – making them potential targets for both diagnosis and therapy. Haemagglutinin allows the virus to enter host cells, such as epithelial cells lining the respiratory tract. This molecule mediates viral attachment to the host cell and its subsequent entry into cells by causing fusion of the cell membranes (endocytosis). Variation in haemagglutinin molecules can result in altered affinity for the different types of the cellular receptor (sialic acid) for binding either the human or the avian variant forms. Because of this, receptor specificity of the haemagglutinin molecule is thought to be able to determine the transmissibility of an influenza virus strain in a given species.[64]

Once the virus is inside the host cell (such as epithelial cells), it uses the cell to synthesise copies of itself. Neuraminidase hydrolyses the bonds between haemagglutinin and sialic acid after the virus has reproduced, allowing it to leave the cell to continue infecting other cells. Neuraminidase activity is required for the release of the viral particles/virions and the activity of neuraminidase is the target of the current antiviral drugs, Oseltamivir and Zanamivir.[64] Treatment prevents the action of neuraminidase, resulting in a failure of the release the new virions being released and therefore prevents further spread of the virus.

Influenza A viruses are given the denotation HxNy, where x can be 1-16, as there 16 known subtypes of haemagglutinin and y can be 1-9 as there are 9 known neuraminidase subunits, however only H 1, 2, 3 and N 1 and 2 are commonly found in humans.[64] The 2009 pandemic influenza H1N1 consists of genetic material from viruses from swine, human and avian origins.[70]

2.2   Influenza A H1N1in historical context

Influenza outbreaks are generally epidemics that spread rapidly among individuals in an area or population at the same time, however, sometimes influenza can spread much faster and wider and become pandemic. Two conditions must be satisfied for an outbreak of influenza to be classed as a pandemic. Firstly, the infection, arising in a specific geographical area, spreads throughout the world. A high percentage of individuals are infected, resulting in increased mortality rates. Secondly, a pandemic is caused by a new influenza virus A subtype, the haemagglutinin of which is not related to that of influenza viruses circulating immediately before the outbreak, and could not have arisen from those viruses by mutation.[71]

Globally, at any one time there are likely to be several influenza viruses circulating in seasonal patterns. Influenza viruses gradually change over time by a process called antigenic drift. This process is characterised by sporadic point mutations in either the haemagglutinin or neuraminidase protein.[67] Occasionally, a major antigenic change in either the haemagglutinin or neuraminidase protein occurs by a process called antigenic shift, resulting in a virus that mutates much faster and with the ability to spread much more widely, quickly and replace the prevailing seasonal influenza strain e.g. the 2009 pandemic influenza H1N1. This may result in a new strain of virus.

Antigenic shift occurs under rare conditions where a host is infected with a combination of two or more influenza A viruses of different origins or strains that are able to become mixed (or ressorted), leading to the generation of a new viral strain. The resulting virus has the ability to spread rapidly within a population (and also to other populations and subsequently globally), as people are likely to be immunologically naive to this new virus. The 2009 H1N1 virus was a triple ressortment influenza virus with swine human and avian genetic material that rapidly became pandemic.[64, 70, 72] However, although the 2009 H1N1 strain was novel, it shared features with other H1N1 strains which aided in its identification.[73]

Pandemic influenza viruses are thought to arise when there is close and frequent contact between humans and other animal species that can be infected by Influenza viruses. The virus develops the ability to jump the species barrier to be able to infect humans. This "crossing" is made possible by certain genetic mutations that permit the binding of the animal viruses to surface proteins in the human respiratory tract.[64] Prior to the 2009 H1N1 pandemic, there have been three well-documented Influenza A pandemics in recent history.[74]

Table 7 summaries some of the major details of recent Influenza A pandemics, with the aim of framing the 2009 H1N1 pandemic in a historical context.[65, 66, 74]

Table 7   Recent Influenza A pandemics

Pandemic	Year	Causative agent	Estimated global effect
Spanish flu	1918	H1N1	500 million cases, 50 million deaths
Asian flu	1957	H2N2	1-1.5 million fatalities
Hong Kong flu	1968	H3N2	0.75 – 1 million fatalities
Swine flu	2009	H1N1	~17,000 (till March 2010)

The first of these pandemics was caused by a H1N1 virus and coincided with the outbreak of World War I. It is estimated that roughly a third of the world's population was infected (approximately 500 million people) and resulted in approximately 50 million deaths.[64] The 1957 and 1968 pandemics resulted in a lower number of cases and fatalities; however, they still resulted in a significant global effect. The 2009 pandemic was caused by a H1N1 virus and although the number of fatalities was much lower than any of the previous pandemics, concern was raised by how quickly the virus spread. The outbreak began in the state of Veracruz, Mexico in March/April 2009 and spread rapidly across the world. Only a matter of weeks later (in June 2009), the World Health Organization (WHO) and US Centers for Disease Control (CDC) stopped counting cases (30,000 confirmed cases) and the WHO declared the outbreak a pandemic.[62] Subsequent estimates have reported millions of cases and at least 16,813 deaths.[66]

Due to the availability of epidemiological records and preserved archival material, the viruses responsible for the recent pandemics have been identified, allowing patterns of infectivity to be modelled.[73] The haemagglutinin gene of 2009 H1N1 is derived from "classical swine H1N1" virus, which likely shares a common ancestor with the human H1N1 virus that caused the pandemic in 1918. By using homology modelling, Igarashi[73] showed that the two viral strains exhibit such a high degree of structural homology that prior exposure to the 1918 strain is likely to confer specific immunity against H1N1. Techniques that model the structure and homology of haemagglutinin and neuraminidase proteins are useful in identifying potential therapeutic targets for drug and/or vaccination development,[75, 76] as well as potential targets for diagnostic tests. H1N1 was first described in the 1918 pandemic and made a resurgence in April 2009 in the form of a triple-reassortant influenza A virus, which is composed of a combination of human, swine, and Eurasian avian strains.[77] What was so unusual about this Influenza A strain was that it appeared to be comprised of several separate types of known flu strain and was described as " ..an unusually mongrelised mix of genetic sequences"[72] In June 2009, the World Health Organisation declared an H1N1 pandemic – the first global pandemic since the 1968 Hong Kong flu. Infections with influenza H1N1 viruses are usually less severe in their impact than those with Influenza B viruses; however they appear to have more pandemic potential, possibly as a result of the type of haemagglutinin protein present.[64]

Currently, there is no known cure for the influenza virus, therefore strategies have focussed on prevention (such as the use of vaccination programs) and reduction of symptoms (such as with antiviral drugs such as Oseltamivir and Zanamivir). The lack of a cure relates to the success of the influenza virus to adapt in order to evade the immune system. Distinguishing an influenza virus from other circulating pathogens on the basis of signs and symptoms alone is unreliable due to the large degree of similarity in presentation. For example, *Mycoplasma pneumonia* – causative agent of primary atypical pneumonia (also known as pleuro-pneumonia like organism or PPLO) causes similar signs and symptoms of those of respiratory tract infection.[78] Laboratory testing of respiratory specimens offers a much improved method of identifying causative agents and therefore improving effectiveness of management.

2.3   Diagnostic testing for Swine flu (H1N1)

Given that there may be different strains of influenza virus circulating at any one time, as well as other organisms eliciting very similar symptoms, identifying a viral strain responsible for an outbreak can be difficult. Often the diagnosis of influenza is based on presenting symptoms[79] instead of the more accurate laboratory testing.[61] One reason for this could be the amount of time taken to obtain laboratory test results. The reference or definitive test to identify a virus is viral culture which can take up to 10 days – by which time the patient may have potentially infected many others. Considering the importance of a making correct diagnosis and the time taken to make this diagnosis, there is a need for a diagnostic test that can be used to correctly identify influenza in a clinically relevant timeframe of hours and not days.

Therefore, diagnostic tests that can accurately identify a causative agent in a clinically useful timeframe are extremely useful in clinical practice. In laboratory testing for influenza, respiratory tract samples (e.g. fluid, swabs or sputum) are collected from patients in order to characterise the agent causing the infection. Definitive diagnosis of swine flu H1N1 largely requires a test that is able to distinguish between influenza A subtypes.[77] Currently there are several methods available, each of which relies on particular features of the virus. Broadly, the types of test available to influenza viruses are:

- Infectivity based assays
- Protein binding based assays
- Nucleic acid based assays

2.4 Infectivity Assays

Viruses require host cells to infect and to be able to replicate. Once a cell has been infected with a virus, the cell responds is a predictable way. Infectivity assays are in vitro systems that aim to identify a virus on the basis of the cell type it infects and characteristics of the infection, such as changes in the cell and how the infection progresses. The major infectivity assay used in hospital laboratories in the diagnosis of viruses is viral culture.

Viral culture is often considered the most sensitive method for diagnosis of type A influenza virus infection is isolation of the virus in cell culture.[80-82] and is traditionally used as the reference standard for comparison test.[83] Viruses require host cells to survive and reproduce, therefore cell lines are inoculated with a sample from the patient and grown in culture medium. Cells infected with the virus are visually identified under standard microscopy. Growth, detection and identification may require several days (often 3 – 10 days) to complete. Inevitably, this results in a delay in diagnosis, administration of antiviral agents and increases the risk of the virus spreading. Major disadvantages with this method include the requirement for specialised laboratory conditions, considerable technical expertise and is time consuming. Furthermore, only viable viruses are able to be detected in such assays, in contrast to techniques that establish the presence of a virus based on particular proteins or nucleic acids. This could potentially result in lower levels of detection of a virus and therefore potentially a false negative result for a patient.

There is a need, therefore for a method accuracy that is comparable to viral culture but is much less time, resource and skill intensive. For the purposes of this review and due to its high level of accuracy and reliability, viral culture will be used as the reference test, to which more recently developed tests will be compared to.

2.5 Protein binding based assays

The aim of a protein binding assay is to identify a virus by picking out specific proteins (antigens) that are likely to be expressed by that virus alone. One such technique is immunofluorescence whereby a labelled antibody raised against all or part of the viral protein is used. An antibody can be labelled in several ways, either directly with a fluorescent reporter dye or indirectly – first with an intermediate antibody, which in turn is then detected with a labelled antibody - depending on laboratory requirements and preferences.

The amount of the labelled antibody is used as a measure of the amount of viral protein present in the patient specimen, but is unable to give an indication of whether or not the virus is viable and therefore able to cause infection.

Antigen detection using immunofluorescence techniques was pioneered in the 1970s, and commercial reagents are now widely used for the detection of influenza viruses.[84] These assays require less technical expertise than viral culture techniques and have the advantage of allowing direct evaluation of specimen quality. This technique aims to identify the virus by exploiting antigenic properties of the proteins it expresses on its outer protein coat. Antibodies raised against virus-specific protein sequences are raised in an animal, then are fluorescently labelled (or tagged). The patient sample is incubated with the antibodies, the antibodies bind to the proteins on the virus and the protein of interest becomes visible under a fluorescent microscope. Synthesis of specific antibodies can take a substantial amount of time, depending on the animal species. Immunofluorescence results are generally available within hours but this type of technique can be less sensitive than viral culture as it relies on the specificity of the detection antibodies.

2.5.1   Rapid diagnostic tests for viral proteins

Rapid diagnostic test (RDT) kits signal the presence of a protein of interest with a chromatic change of membrane bound reagents. Such kits have become increasingly more commercially available and claim to identify viral proteins with a good degree of accuracy.[80, 82] These diagnostic test kits can be produced as dipsticks, cassettes, or cards and they contain internal positive and negative controls. Many of these tests can be automated and can be carried out in a matter of minutes (5 – 40).[84] The sensitivity and/or specificity may be greatly reduced and is thought to vary with virus type or subtype, timing of specimen collection, specimen type, patient age, and the test comparator. This reduced accuracy (compared to viral culture) may, however, still be useful in situations where a result is needed urgently – such as a pandemic situation or in resource poor situations. With the emergence of the pandemic influenza A (H1N1) 2009 virus, RDTs have been widely used for patient triaging, although there are limited data available on their clinical accuracy.[84] To correctly interpret results of diagnostic tests such as RDTs that have relatively low sensitivity and/or specificity, the prevalence of influenza in a community must be considered.[85] During peak disease activity, positive predictive values are highest, but false-negative results more likely. The opposite is true during times of low disease activity.[61] When the disease prevalence is low or unknown, RDT results become difficult to interpret and of limited use.[61]

2.6   Nucleic acid based assays

Nucleic acid techniques aim to identify a virus on the basis of short, specific sequences of genetic material (nucleic acids). For influenza A viruses, targets for nucleic acid based assays are conserved regions of the haemagglutinin, neuraminidase and matrix genes. Polymerase chain reaction (PCR) is one such group of techniques.

Nucleic acid techniques possess several advantages over infectivity- or antigen-based techniques for the detection of influenza. These tests have improved sensitivity for detecting organisms that are fastidious, no longer viable, or only present in small amounts.[84] They are also able to provide rapid genetic information regarding sequence evolution, geographic variation, or the presence of virulence factors or antibiotic resistance.[84] Their rapid turnaround times (relative to viral culture) allow them a more prominent role in patient management, and the ability to be able to test for multiple pathogens simultaneously has aided in the diagnosis of nonspecific respiratory syndromes, such as in outbreak settings.[84]

2.6.1   PCR – polymerase chain reaction techniques

PCR aims to reveal the presence of viral genetic material using small fragments of complimentary sequences (primers). A small sequence of DNA is chosen to be specific to a gene of interest (computer software is available to help identify target sequences), so that when it binds to the viral DNA it can be amplified and identified. There are three main types of PCR: traditional, real time and multiplex. Regardless of the type of PCR used, all PCR assays require good primer design taking into consideration gene target, gene number, mobility of genes between species, stability of gene, and the presence of mutations.[84] In addition to optimal primer sequences, PCR requires high purity of target nucleic acid fragments in order to have minimal interference with this test.

2.6.2   Traditional PCR

PCR utilises a pair of primers, which are complementary to a defined sequence on each of the two strands of the cDNA. The cDNA is used as this is the DNA sequence that would be directly translated into protein after transcription. The primers are then extended by a DNA polymerase and a copy of the strand is made after each cycle, leading to exponential amplification. The exponential amplification via reverse transcription polymerase chain reaction provides for a highly sensitive technique in which a very low copy number of RNA molecules can be detected.

RT-PCR is widely used in the diagnosis of genetic diseases and, semi-quantitatively, in the determination of the abundance of specific different RNA molecules within a cell or tissue as a measure of gene expression.[86] Reverse transcription polymerase chain reaction (RT-PCR) is a variation that has an additional step of generating (by reverse transcribing) the DNA sequence prior to its amplification. RT-PCR is commonly used in studying the genomes of viruses whose genomes are composed of RNA, such as Influenza virus A and retroviruses like HIV.[86] RT-PCR includes three major steps. The first step is reverse transcription (RT), in which RNA is reverse transcribed to complementary DNA, or cDNA, using reverse transcriptase. This step is very important in order to perform PCR since DNA polymerase can act only on DNA templates. The RT step can be performed either in the same tube with PCR (one-step PCR) or in a separate one (two-step PCR). The next step involves separation of the double stranded DNA (dsDNA), so that the primers can bind to the separate strands. This is done by using high temperature (such as 95°C) so that the bonds between the DNA strands denature. Then, the temperature is decreased until it reaches the annealing temperature which can vary depending on the set of primers used, their concentration, the probe and its concentration (if used), and the cation concentration.[86] The final step of PCR amplification is DNA extension from the primers. The length of the incubation at each temperature, the temperature alterations, and the number of cycles are controlled by a programmable thermal cycler. In conventional PCR, the products are detected using agarose gel electrophoresis and ethidium bromide (or other dyes that bind to nucleic acids). Real time PCR can also be used to visualise products of PCR using a fluorescent reporter dye as the reaction occurs and the fluorescence increases as the amplification progresses and the instrument.

2.6.3   Real time PCR

Real time PCR (rt RT PCR) is another PCR technique and allows visualisation of the RNA of interest as the binding occurs. Fluorescently labelled primers can be detected using a spectrophotometer and fluorescent signal intensity increases proportionally as the number of RNA copies is amplified. Real time RT PCR has several features that are considered improvement over the traditional agarose gels. Firstly, the two steps of amplification and detection are combined in one reaction, increasing the speed and efficiency of testing and reducing the risks of operator error and cross-contamination of samples. Secondly, as the fluorescent signal is proportional to the amount of RNA amplification, this technique is semi-quantitative and can allow estimates of the amount of starting nucleic acid material. Reliability of PCR depends upon the specificity of the primer sequence and that the labelled RNA probes only bind to regions of viral RNA specific to that individual virus strain.

As with protein based assays, nucleic acid assays are unable to distinguish between viable and non-viable viruses within a sample.[86]

2.6.4 Multiplex PCR

Multiplex PCR systems are used to search for multiple PCR targets at the same time in one reaction. This has the advantage of increasing the number of pathogens tested for, without increasing the amount of reagents, technician time or specimen material required.[84] Multiplex assays have broadened the scope of respiratory surveillance studies, and have also led to the increasing recognition of dual or triple infections in the same individual.[84] There are some technical considerations when attempting to identify multiple PCR products in the same reaction tube. In the amplification step, all multiplex platforms must balance the competing optimal PCR conditions for each individual target, and must overcome problems of competition and inhibition among the various primers and probes. The detection step is also complex, as the multiple PCR products need to be distinguished from one another. This can be done on several bases, such as the use of different colour reporter dyes, or by differences in their size (molecular weight), using resolution techniques such as agarose gel electrophoresis to differentiate by weight, and capillary-based auto-sequencers that identify targets by length and sequence.[84]

2.7   Summary of techniques used to diagnose Influenza A

Viral culture is often used as the reference standard for laboratory diagnosis; however this requires significant time, resources, expertise and cost. The impetus has been to find tests that are as least as accurate (i.e. have equal or superior sensitivity and specificity), quicker, less expensive and/or can be conducted with less expertise. Table 8 has been constructed from data from[84, 87] to compare the major features of techniques commonly used to identify Influenza A viruses. Stars are used in this table to denote the relative degree of the technical expertise and laboratory resources required for each of the techniques. The number of stars is proportional to the greater degree of expertise and resources required.

Table 8   Main features of the major techniques used to identify Influenza A viruses

Type of test	Basis of virus identification	Time to results	Degree of technical expertise	Laboratory resources required
Traditional viral culture	Infectivity assay	Days	****	****
Immunofluorescence	Protein based assay	Minutes	***	***
Rapid diagnostic tests	Protein based assay	Minutes	**	**
PCR techniques	Nucleic acid sequence based assay	Hours	***	***

Chapter 3   Systematic review protocol

3   The systematic review protocol

The following chapter is the approved a priori protocol [17] for this systematic review. The format is based on that recommended by the Joanna Briggs Institute and consists of standardised sections and includes some material from the previous chapters (Chapter 1 and 2) as background to the review.

3.1   Background

3.1.1   Diagnostic tests

Diagnostic tests aid clinicians in making a diagnosis and a subsequent treatment plan for the patient based on the test results. [34] Diagnostic test interpretation involves comparing features of the patient against known features of the suspected condition and generating a likelihood of the presence or absence of the condition. In order to determine how accurate a diagnostic test is, the test should be compared to a reference test/standard that has been proven as being reliable for the particular condition of interest and patient population. Systematic reviews of diagnostic tests are a recent development and aim to inform the clinician by critically appraising and synthesising the best available evidence concerning the accuracy of diagnostic tests, ideally in comparison with alternate techniques. [11]However, this type of review can be challenging as determining what is meant by accuracy in this context is more complex than with other study designs, such as studies of therapeutic interventions. [12]

The accuracy of a test is determined by posing two questions:

- How often does the test give positive results for patients that have the condition? (sensitivity) and
- How often does the test give negative results for patients that do not have the condition? (specificity).

As with any procedure, there are benefits and limitations. One of the major benefits is that (generally) diagnostic tests are relatively straightforward to conduct and provide the clinician with evidence on which to base a diagnosis within a useful timeframe. When conducting a diagnostic test, a patient sample (e.g. blood, urine, scan etc.) is compared against a reference or test standard, therefore test accuracy can be compromised if the wrong type of patient sample or reference test/material is used.

3.1.2 Meta analysis of diagnostic accuracy

Evidence from systematic reviews of diagnostic test accuracy is important in determining whether a test is being used appropriately and can be used to guide health care policy and funding decisions. Combining diagnostic tests in meta-analysis pose many challenges – not least because what is meant by accuracy in this context is more complex than with other study designs such as randomised controlled trials, as both measures of sensitivity and specificity need to be considered. [12] Interpretation of study results can also be complex and may suffer from bias due to poor reporting or under reporting of no-effect results. [41]

Initiatives aimed at authors of diagnostic test research, such as STARD [59] and QUADAS [57, 60] are an attempt at improving the quality of this area of research by standardising reporting. In addition, the Cochrane Collaboration has published guidance [29, 88] to aid authors of diagnostic test accuracy systematic reviews.

3.1.3 Influenza A H1N1 "Swine flu"

Influenza is a viral infection that affects mainly the nose, throat, bronchi and, occasionally, lungs. Infection usually lasts for about a week, and is characterised by sudden onset of high fever, aching muscles, headache and severe malaise, non-productive cough, sore throat and rhinitis. [61] The virus is transmitted easily from person to person via droplets and small particles produced when infected people cough or sneeze. Influenza tends to spread rapidly in seasonal epidemics. Most infected people recover within one to two weeks without requiring medical treatment. However, in the very young, the elderly, and those with other serious medical conditions, infection can lead to severe complications including pneumonia and death.

In their simplest form, viruses are composed of an outer protein coat and an inner core of either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), however, unlike bacteria, no virus has been shown to contain both [63]indicating they are unable to transcribe RNA to DNA, one reason why they require host cells. The viruses that cause Influenza can be categorised on the basis of proteins on the outer protein coat into either Influenza A or B. The category of Influenza A viruses is further subdivided on the basis of two specific proteins on the viral surface: Haemagglutinin (H) and neuraminidase (N). Both of these proteins are important for the virus to be able to enter the host cell, reproduce and be released to infect other cells – making them potential targets for detection and therapies. Haemagglutinin allows the virus to enter host cells, such as epithelial cells lining the respiratory tract. The haemagglutinin subunits bind with sialic acid residues on epithelial cells of the lungs and throat.

Once the virus is bound, the virus enters the cell by endocytosis and replicates. Neuraminidase hydrolyses the bonds between haemagglutinin and sialic acid after the virus has reproduced, allowing it to leave the cell to continue infecting other cells.

There are 16 known subtypes of haemagglutinin and 9 neuraminidase subunits, however only H 1, 2, 3 and N 1 and 2 are commonly found in humans. The outer protein coat of the Swine flu virus is comprised of H1 and N1 subunits and therefore is given the denotation H1N1. The haemagglutinin gene of 2009 H1N1 is derived from "classical swine H1N1" virus, which likely shares a common ancestor with the human H1N1 virus that caused the pandemic in 1918. By using homology modelling, [73]show that the two viral strains show such a high degree of structural homology that prior exposure to the 1918 strain is likely to confer specific immunity against H1N1. Techniques that model the structure and homology of haemagglutinin and neuraminidase proteins are useful in identifying potential therapeutic targets for drug and/or vaccination development [75, 76]as well as targets for diagnostic tests.

Influenza A H1N1 was responsible for the 1918 pandemic and subsequently made a resurgence in April 2009 in the form of a triple-reassortant influenza A virus, which is composed of a combination of human, swine, and Eurasian avian strains. [77] What was so unusual about this Influenza A strain was that it appeared to be comprised of several separate types of known flu strain and was described as " ..an unusually mongrelised mix of genetic sequences" [72] In June 2009, the World Health Organisation declared an H1N1 pandemic – the first global pandemic since the 1968 Hong Kong flu. [72] Infections with influenza H1N1 viruses are usually less severe in their impact than those with influenza B viruses; however, they appear to have more pandemic potential, possibly as a result of the type of haemagglutinin protein present.

3.2   Diagnostic testing for Swine flu (H1N1)

Given that there are often several strains of influenza virus circulating at any one time, as well as bacteria eliciting the same symptoms, identifying the strain responsible for an outbreak can be difficult and often a diagnosis in primary care is based on presenting symptoms [79] instead of the more accurate laboratory testing.[61] One reason for this could be the amount of time taken to obtain laboratory test results. The reference or definitive test to identify a virus is viral culture which can take up to 10 days. Hence the drive for an accurate diagnostic test that generates reliable results in a clinically relevant timeframe. The drive for a clinically relevant timeframe to obtain results was highlighted by the recent 2009 H1N1 pandemic.

Public health recommendations (CDC and WHO) were to promote diagnostic testing for H1N1 early in the pandemic while public health measures such as isolation and antiviral therapy could be effective at reducing (or at least slowing down) the spread of infection.

Respiratory tract samples (e.g. fluid, swabs or sputum) are collected from patients to characterise the agent causing the infection. Definitive diagnosis of swine flu H1N1 largely requires a test that is able to distinguish between influenza A subtypes.[77] Currently there are several methods available, each of which relies on particular features of the virus. Broadly, the types of test available to influenza viruses are:

3.2.1 Viral culture

Isolation of the virus infection in cell culture is often considered the most sensitive method for diagnosis of type A influenza virus infection. [80-82] Viruses require host cells to survive and reproduce, therefore cell lines are inoculated with a sample from the patient and grown in culture medium. Cells infected with the virus are visually identified under standard microscopy. Growth, detection and identification may require several days (often up to 10 days) to complete. Inevitably, this results in a delay in diagnosis, the prophylactic administration of antiviral agents and increases the risk of the virus spreading. Major disadvantages with this diagnostic method include the requirement for specialised laboratory conditions, considerable technical expertise and the lengthy timeframes. The search is on therefore for a quicker method with comparable accuracy that is less resource and skill intensive. Due to its high level of accuracy and reliability, viral culture will be used as the reference test to which more recently developed tests will be compared in this review.

3.2.2 Immunofluorescence

This type of technique aims to identify the virus by exploiting antigenic properties of the proteins it expresses on its outer protein coat. Antibodies raised against virus-specific protein sequences are raised in an animal, then are fluorescently labelled (or tagged). The patient sample is incubated with the antibodies, the antibodies bind to the proteins on the virus and the protein of interest becomes visible under a fluorescent microscope. Synthesis of specific antibodies can take a substantial amount of time, depending on the animal species. Immunofluorescence results are generally available within hours but this type of technique can be less sensitive than viral culture as it relies on the specificity of the detection antibodies.

3.2.3   PCR – polymerase chain reaction

The PCR technique aims to reveal the presence of viral genetic material (RNA ribonucleic acids) using small fragments of complimentary sequences (primer). A small sequence of RNA bases is chosen and labelled so that when it binds to the viral RNA, it can be identified. Real time PCR allows visualisation of the presence of the RNA of interest as the binding occurs. Primers are fluorescently labelled which can be detected using a type of spectrophotometer and the fluorescent signal intensity increases as the number of RNA copies is amplified). PCR products can also be detected on agarose gels using electrophoresis. The reliability of PCR depends upon the specificity of the primer sequence and that the labelled RNA probes only bind to regions of viral RNA specific to that individual virus strain.

3.2.4   Rapid diagnostic tests for viral proteins

More recently, rapid diagnostic kit tests have become commercially available which claim to identify viral proteins with a good degree of accuracy.[80] Generally they aim to identify either amino acid sequences or protein(s) specific to a virus. Many of these tests can be automated and can be carried out in a matter of minutes. The sensitivity and/or specificity may be greatly reduced but there may be enough information on which to base preliminary diagnosis. [84]

3.3   Aims of the review

Before embarking on this systematic review, a detailed search of all the major databases, as well as The Joanna Briggs Library of Systematic Reviews and the Cochrane Collaboration library, was conducted to establish that there were no existing or underway systematic reviews on this topic.

The aim of this systematic review is to comprehensively search the available literature and to synthesise the best available evidence to determine the diagnostic accuracy of currently available laboratory tests for swine flu (H1N1), using viral culture as a reference test.

3.4   Inclusion Criteria

3.4.1   Types of participants

Studies will be considered for inclusion in the review if the participants are human patients exhibiting influenza-like symptoms who have been tested for swine flu (H1N1) using both a diagnostic and reference test. There will be no exclusion based on age, gender or co-morbidities, but subgroup analysis will be conducted if there is sufficient data and if appropriate.

3.4.2   Phenomenon of interest

The phenomenon of interest is the correct identification of H1N1 infection. This review will consider studies that compare the accuracy of laboratory tests aiming to diagnose swine flu (H1N1) in patients presenting with influenza-like symptoms, in comparison to viral culture.

3.4.3   Outcome Measures

The accuracy of diagnostic tests will be determined by how well the test correctly gives positive results for patients with confirmed swine flu (H1N1) and negative results for patients without, as determined by viral culture. It is anticipated that data will be presented as raw test results or sensitivities and specificities of the test compared with viral culture will be presented. Other statistics to elaborate on diagnostic test accuracy may include likelihood ratios and/or predictive values.

3.4.4   Types of studies

This review will consider any quantitative study that examines the diagnostic accuracy of laboratory tests utilised for swine flu (H1N1) where the participant undertakes both the index and reference test.

3.4.5   Exclusion criteria

Studies that utilise non-human participants, focus on the analytical sensitivity of a test, do not distinguish between influenza A subtypes, or do not use viral culture as the reference test will be excluded.

3.5   Search strategy

The search strategy aims to find both published and unpublished literature. Initial search terms and databases were chosen in discussion with a research librarian, with the aim of identifying the maximum number of articles. Databases will be searched from their inception, to 31st May 2010. A three-step search will be used. Initially, a limited search of Pubmed and Cinahl will be undertaken in order to identify appropriate keywords. These keywords are presented in Appendix I. Analysis of the text words and Mesh heading identified by the search to describe relevant articles will then be used to identify additional search terms which will then be used to search across all included databases. The databases to be searched are listed in Appendix II. Thirdly, the reference list of identified papers will be searched for additional studies. The search will not be limited by year but due to a lack of translational resources, included studies will be limited those published in the English language.

3.6   Methods of the Review

3.6.1   Critical appraisal

The methodological quality of each paper will be critically appraised by two reviewers independently, in order to limit potential reviewer bias. Several quality criteria need to be considered when evaluating studies for potential inclusion. These include:

•   the clinical spectrum of included patients

•   blinded interpretation of test and reference standard results

•   potential for verification bias

•   patient sampling, prospective design

•   adequate description of the index test, reference standard, and study population.

In order to assess the methodological quality of papers included in the review, the checklist developed by the QUADAS initiative checklist [57, 60] will be used as a critical appraisal instrument. Studies will not be excluded on the basis of quality and issues relating to study quality will be explored. The checklist is presented in Appendix III. In situations where study features may not be reported in the primary studies, the reviewers might need to contact authors or seek additional information for clarification.

3.6.2   Data extraction

Data will be extracted from included studies using the STARD 25 item checklist [59] (Appendix IV) and will include details pertinent to: population, setting, test details and the sensitivity and specificity for the index test, as compared with viral culture.

3.6.3   Data analysis and synthesis

Where appropriate, the sensitivities and specificities from individual studies will be combined to generate a summary estimate of the accuracy of PCR techniques, as compared with viral culture methods. Revman 5 (Cochrane Collaboration) and Microsoft Excel computer software will be used for data management and analysis. If meta-analysis is not possible, the data will be discussed in narrative summary and the reasons why the data was unable to be combined will be explored.

3.6.4   Conflicts of interest

None

3.6.5   Acknowledgments

As this systematic review will form part of a Masters of Clinical Sciences thesis, a secondary reviewer (TS) will be used for critical appraisal only.

Chapter 4   Results

4   Results

4.1   Results of the search

The search for relevant studies was conducted during May-June 2010. Using the search strategy detailed in Appendix I and the 15 databases listed in Appendix II, 3843 potentially relevant titles were identified. Even though database filters such as "human only" and "English language" were applied to the search, a number of studies examining swine flu in pigs and studies published in languages other than English were identified in the search. All of 3843 of the potentially relevant titles were screened, and on the basis of title and keywords, abstracts of 360 studies were retrieved. Appendix IV is an example search strategy tailored to the PubMed database.

Following careful comparison of those 360 abstracts to the inclusion criteria, the full text articles of 56 studies were retrieved for consideration. A further 22 studies were excluded following full text retrieval. The remaining 29 studies underwent critical appraisal by two independent reviewers and seven were deemed to be relevant. The tool used for critical appraisal can be found in Appendix III. After careful investigation two studies were later excluded due to incomplete data sets. The authors were contacted, however one did not respond and the second was unable to provide the necessary data.

The discrepancy between the number of identified titles (3843) and the number of retrieved full text articles (56), illustrates the difficulty in locating relevant diagnostic studies, which is consistent with previous reports, as discussed in section 1.6. The flowchart (Figure 2) details the number of studies removed at each stage of the search process.

Details of the five included studies can be found in Appendix VI and details of the excluded full text articles together with reasons for exclusion can be found in Appendix VII.
Of the full text articles to be excluded, the most common reasons for exclusion were:
- The study focussed on the analytical sensitivity (not diagnostic accuracy) of a test, for example research questions focused on the minimum concentration of virus that a test was able to detect (n=11).
- The study (and/or the test) did not distinguish between Influenza A subtypes and therefore the data on H1N1 could not be separately identified (n=22).

- The study did not use viral culture as a reference standard (n=12)

- There was insufficient data (n=7).

Of the five included studies, three included studies were conducted in the USA, [89-91] one study in Iran [92] and one study in China. [93] A total of 1581 patients were included across the five studies, with sample sizes ranging from 100 [89] to 689. [92] There was however no stratification by either age or gender.

```
┌─────────────────────────────┐
│ Potentially relevant papers │
│ identified by literature    │
│ search                      │
│ n=3843                      │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Abstracts retrieved for     │        │ Studies excluded after      │
│ examination on the basis of │───────▶│ evaluation of abstract      │
│ title and keywords          │        │ n= 304                      │
│ n=360                       │        └─────────────────────────────┘
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Full text papers retrieved  │        │ Studies identified from the │
│ for detailed examination    │◀───────│ reference lists of retrieved│
│ n=56                        │        │ studies                     │
│                             │        │ n=3                         │
└─────────────────────────────┘        └─────────────────────────────┘
              │
              │                         ┌─────────────────────────────┐
              │───────────────────────▶│ Studies excluded after      │
              │                         │ detailed examination of     │
              │                         │ full text paper             │
              │                         │ n=27                        │
              ▼                         └─────────────────────────────┘
┌─────────────────────────────┐
│ Papers assessed for         │        ┌─────────────────────────────┐
│ methodological quality      │        │ Studies excluded following  │
│ n=29                        │───────▶│ critical appraisal of full  │
│                             │        │ text paper                  │
└─────────────────────────────┘        │ n=24                        │
              │                         └─────────────────────────────┘
              ▼
┌─────────────────────────────┐
│ Papers included in the      │
│ systematic review           │
│ n=5                         │
└─────────────────────────────┘
```

Figure 2   Flowchart detailing study identification

4.2   Methodological Quality of the Included Papers

4.2.1   Critical appraisal using the QUADAS checklist items

The QUADAS checklist [57, 60] was used to assess the quality of the conduct and reporting of eligible studies. The checklist items relate to known areas of bias and are recognised indicators of methodological quality. [57, 60] No studies were excluded from the review on the basis of poor methodological quality and performance of the included studies performed against the assessment criteria will be further discussed in the following section. Table 9 shows how each of the studies compared for each of the QUADAS checklist items (Q1-Q14) which are listed in full in Appendix III.

Table 9   Methodological quality assessment using items from the QUADAS checklist

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
Beck et al [89]	✓	x	✓	x	✓	✓	✓	✓	x	x	x	u	✓	NA
Bose et al[90]	✓	x	✓	x	✓	✓	✓	✓	x	x	x	u	✓	NA
Ginnochio et al [91]	✓	x	✓	x	✓	✓	✓	✓	x	x	x	u	✓	NA
Mokhtar-Azad et al[92]	✓	x	✓	✓	✓	✓	✓	✓	✓	x	x	u	✓	NA
Wu et al[93]	✓	x	✓	✓	✓	✓	✓	✓	✓	x	x	u	✓	NA

Key:

✓ = yes or positive response, x = no or negative response, u = unclear, NA = not applicable

4.2.2 QUADAS Checklist items

1.   Was the spectrum of patients representative of the patients who will receive the test in practice?

All five of the included studies used patients that presented with influenza-like symptoms. Therefore, the population does represent the spectrum of patients who would receive the test in practice. None of the included studies detail the spectrum of disease for the study patients.

2.   Were selection criteria clearly described?

None of the included studies detailed how the patients were selected for inclusion in the study.

3.   Is the reference standard likely to correctly classify the target condition?

All of the included studies used viral culture to identify Influenza A virus and then used subtyping methods to identify H1N1, which is likely to correctly identify the target condition.

4.   Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?

Three studies [89-91] conducted the index tests study using patient samples that had previously been frozen. The reference test had been conducted on the samples before the samples were frozen. No details were given regarding the time period between the reference and the index test; therefore we are unable to ascertain whether the period is short enough for changes to occur in the target condition. The other two studies conducted the reference and the index tests simultaneously on fresh patient samples. [92, 93]

5.   Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?

All of the five included studies processed patient samples for both the reference and index tests.

6.   Did patients receive the same reference standard regardless of the index test result?

Patient samples were processed for the same reference test regardless of the index test result. Two studies conducted the reference and the index tests at the same time [92, 93] and the remaining three where the reference test results were known prior to conduct of the index test, used de-identified patient samples. [89-91]

7.   Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?

All of the included studies used independent reference and index tests.

8.   Was the execution of the index test described in sufficient detail to permit replication of the test?

All of the included studies either reported the index test in detail [89, 90, 92, 93] or referred to the manufacturers' instructions [91]

9.   Was the execution of the reference standard described in sufficient detail to permit its replication?

The reporting of viral culture methods was poor by all of the included studies. In particular, details such as: culture conditions, cells used to grown viruses from patient samples and time in culture were omitted. One paper gave reference for the viral methods used [93] and another provided limited details. [92] Three studies simply stated that patient samples were previously identified as being viral culture positive. [89-91]

10.   Were the index test results interpreted without knowledge of the results of the reference standard?

Three of the included studies [89-91] were performed using specimens for which a diagnosis had already been made on the basis of viral culture results. The specimens where however de-identified to prevent this knowledge influencing the index test. The other two studies [92, 93] processed samples in parallel for both tests.

11.   Were the reference standard results interpreted without knowledge of the results of the index test?

None of the included studies explicitly address this criterion but it is reasonable to assume that at least for the three studies conducted on previously frozen patient samples [89-91] that the reference test results were interpreted without the knowledge of the index test.

12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?

It is unclear whether the same clinical data would be available for those patient samples that had been frozen. [89-91] For the remaining two studies, [92, 93] it is reasonable to assume that as the tests are conducted simultaneously, the same clinical data was available when the test results were interpreted, as would be available in practice.

13. Were uninterpretable/ intermediate test results reported?

Indeterminate results were reported in two studies, [89, 90] however, in each case, they were excluded from subsequent accuracy calculations, leading to higher estimates of sensitivity and specificity than would otherwise have been reported.

14. Were withdrawals from the study explained?

This question was not applicable to the studies included in this review as the reference and index tests were carried out on patient samples that had already been collected for routine diagnosis, therefore, patient consent was not required and so withdrawal was not an option.

4.2.3   Assessing methodological quality using the STARD checklist items

Although the primary objective of the STARD checklist was to help authors of DTA research papers, this checklist also provided a convenient and detailed data extraction tool. In this review the STARD checklist was used to extract methodological quality data in addition to those of the QUADAS checklist.

Table 10 provides a summary of how each of the included studies compared against each of the STARD checklist points. Responses to individual checklist items are discussed in greater detail (Q1-Q25) below. The checklist questions are listed in full in Appendix V.

Table 10   STARD Methodological quality checklist items

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25
Beck et al [89]	✔	✔	x	u	x	P	x	✔	x	x	x	x	✔	✔	x	x	x	x	✔	NA	x	✔	x	✔	✔
Bose et al [90]	✔	✔	x	u	x	P	x	✔	x	x	x	x	✔	✔	x	x	x	x	✔	NA	✔	✔	x	✔	✔
Ginnochio et al [91]	✔	✔	x	u	x	P	x	✔	x	x	x	x	✔	✔	x	x	x	x	✔	NA	x	x	x	✔	✔
Mokhtar-Azad et al [92]	✔	✔	x	u	x	P	x	✔	x	x	x	x	✔	✔	x	x	✔	x	✔	NA	x	x	x	✔	✔
Wu et al [93]	✔	✔	x	u	x	P	x	✔	x	x	x	x	✔	✔	x	x	✔	x	✔	NA	x	x	x	✔	✔

Key:

✔ = yes or positive response, x = no or negative response, NA = not applicable P = prospective, U= unclear

4.2.4 STARD Checklist items

1   Was the study identified as being a diagnostic accuracy study?

All of the included studies contained the terms sensitivity and specificity in the abstract. None of the included studies contained the phrase "diagnostic test accuracy" in the title, keywords or abstract sections of the paper.

2   Were research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups, detailed?

The study aims of the included studies included a comparison of the accuracy of the index test against viral culture.

3   Does the study describe the study population, inclusion and exclusion criteria, setting and locations where the data were collected?

All of the patients were described as presenting with influenza-like illness to either a GP or a hospital. However, none of the included studies provided any detail on the spectrum of patient illness or any demographic information. There was also no detail on any exclusion criteria for any of the included studies.

4   Does the study describe participant recruitment?: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?

None of the included studies described recruitment procedures, other than that the samples were taken from patients presenting with influenza-like illness. All of the patients received both the reference and the index test.

5   Describe participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected

How the participants were sampled was not addressed by any of the included studies.

6   Describe data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?

All of the included studies used a prospective study design.

7   Did the study describe the reference standard and its rationale?

Reporting of viral culture methods was poor by all of the included studies. In particular, details such as: culture conditions, cells used to grown viruses from patient samples, and time in culture were omitted. One paper referenced the viral methods used [93] and another provided limited details. [92] Three studies simply stated that patient samples were previously identified as being viral culture positive. [89, 91] None of the included studies described the rationale for using viral culture as a reference standard. Similarly, none of the included studies described what constituted a positive result using this technique.

8   Describe technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.

All of the included studies provided a good level of detail about the index test; however the reference test (viral culture) was poorly described and referenced.

9   Describe definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.

None of the included studies provided detail of what constituted a positive viral culture; however the units used and cut off points for all of the index tests were well described.

10   Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard

None of the included studies addressed training or expertise of laboratory personnel conducting the study at any stage.

11   Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers

None of the included studies addressed blinding. No other clinical data was presented other than that the patients presented with influenza-like symptoms.

12   Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals)

The included studies reported accuracy measures with varying degrees of completeness; however, no study described methods for calculating these measures.

13   Describe methods for calculating test reproducibility, if done.

All of the included studies described test reproducibility for the index test, but not for the index test. A test was deemed to be reproducible when similar results were obtained for a number of replicates.

14   Report when study was done, including beginning and ending dates of recruitment

Only two studies report details of when and where the samples were collected and processed.[92, 93] The other three studies [89-91] were conducted in the USA and were probably conducted early in 2009, although this was not explicitly stated.

15   Does the study report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centres?

The patients for each included study are described as presenting with influenza-like symptoms but no further details are provided. One study [91] reported the age range of the included patients.

16   Does the study report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended)

Not detailed in any of the included studies. The studies report that all of the patients underwent both the reference and the index test.

17   Does the study report time interval from the index tests to the reference standard, and any treatment administered between?

Three studies [89-91] conducted the index test using previously frozen patient samples, following processing for the reference test. No details were given regarding the time period between the reference and the index test. The other two studies conducted the reference and the index tests simultaneously on fresh patient samples. [92, 93]

18   Does the study report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition?

Not detailed in any of the included studies. The patients are simply defined as presented with influenza-like illnesses.

19   Does the study report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard?

None of the included provided a cross-tabulation of the results for analysis of H1N1 on its own. All of the included studies presented the total number of patient samples that tested positive for the reference and the index test. Cross classification for these positive results into true and/or false positive results was not done for any of the included studies. None of the included studies provided the number of patient samples testing negative for each test, preventing calculation of test specificity.

20   Does the study report any adverse events from performing the index tests or the reference standard?

Not applicable as both the reference and the index test are performed on patient samples that had been collected during routine treatment.

21   Does the study report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)?

This was done with varying degrees of completeness; one study [89] reported a confidence interval (CI) but did not report the level (i.e. 90% or 95%). Two studies reported negative predictive values and positive predictive values (NPV and PPV), but no estimate of prevalence. [89, 90] Two studies did not report a CI [92, 94] One study reported NPV and PPV with an estimate of prevalence [91] One study reported a 95% CI [90] and one study gave no measure of statistical uncertainty. [93]

22   Does the study report how indeterminate results, missing responses and outliers of the index tests were handled?

Two studies reported indeterminate results, [89, 90] They were however excluded from the accuracy calculation in both instances which may have led to overestimates of test accuracy.

23   Does the study report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centres?

Not addressed by any of the included studies.

24   Does the study report estimates of test reproducibility?

All of the included studies provided data on test reproducibility of the index test.

25   Does the study discuss the clinical applicability of the study findings?

All of the included studies discussed the clinical implications of the diagnostic test accuracy findings, with regard to identifying patients with H1N1.

4.3 Summary estimates of DTA of Influenza A H1N1 "swine flu" PCR tests

The aim of this systematic review was to extract accuracy measures - such as sensitivities, specificities, and likelihood ratios from the included studies, with the aim of generating an overall summary estimate of diagnostic test accuracy for laboratory tests used to diagnose influenza A H1N1. Although the search was not restricted by type of index test, all of the included studies examined the diagnostic test accuracy of a polymerase chain reaction (PCR) technique against a viral culture reference test.

When sensitivity and specificity are not reported, these measures can be calculated and in order to do this suitable data is necessary– either as a 2 x 2 cross classification table or as raw counts that permit the generation of such a table. A typical cross classification table was presented and discussed in as Table 2 in Chapter 1. The same table is presented again below (Table 11), however with the difference that four cells are highlighted. The highlighted cells are those required to calculate sensitivity and specificity and those that require data from the included studies.

These cells contain patients who are classified as being:

- True Positives
- False Positives
- True Negatives
- False Negatives

Table 11   Cross classification table used to estimate test sensitivity and specificity

Test Outcome (Index test results)	Disease status (Reference test results)		
	Disease positive	Disease negative	
Index test positive	True positives (a)	False positives (b)	Total index test positives (a + b)
Index test negative	False negatives (c)	True negatives (d)	Total index test negatives (c + d)
	Reference test positives (a + c)	Reference test negatives (b + d)	Total number of samples N (a + b + c + d)

The remainder of this chapter will summarise the reported data from each included study with a view to populating a 2x2 cross classification table for each study.

4.4   Data from Beck et al [89]

This cross sectional study was conducted in the USA using nasopharyngeal swab samples that had been collected from patients presenting with influenza-like symptoms for routine diagnosis by viral culture and were then stored frozen until conduct of this study. In an attempt to limit bias, the samples were de-identified and assayed blindly. The table of included studies (Appendix V) provides a summary of the important study details, as reported by the authors.

The study compared several automated reverse transcriptase polymerase chain reaction assays (RT-PCR) for both influenza A and B viruses. H1N1 positive samples were identified as influenza A positive using a semi-automated PCR assay called the Jaguar assay and were subsequently subtyped for H1N1. Details of how subtyping of samples was conducted were not provided. The primers used were designed for highly conserved regions of the Influenza matrix gene. The precision and reproducibility of the primers/Jaguar combination were tested against known strains of H1N1, as well as against 23 other known respiratory pathogens. While stock strains of H1N1 were identified, no cross reactivity was reported with any other respiratory virus. Samples that were positive for the amplified PCR product of the matrix sequence primers were classified as being H1N1 positive if the MS2 internal control was positive and the FAM count was >40, with the appropriate amplification curve and a melt temperature within expected values (63°C +/- 2°C).

No details were given on the viral culture methods used or for the criteria used to determine a positive viral culture result.

Patient samples were not categorised using the 2 x 2 table designations and insufficient data was reported to allow one to be constructed, as only the total number of positive results for each test were reported. The authors did however report sensitivity and specificity values and these together with the raw data is summarised in Table 12. Confidence intervals (95%) were reported for sensitivity and specificity, however the value (i.e. percentage such as 90 or 95%) of these limits was not reported. Table cells containing a dash "–" denotes fields for which there was no published data.

Table 12   Data extracted from Beck et al [89]

Number of patient samples (N)	100
Patients identified as H1N1 positive by viral culture	16
Patients identified as H1N1 negative by viral culture	-
Patients identified as H1N1 positive by PCR	15
Patients identified as H1N1 negative by PCR	-
Indeterminate/inconclusive	7
Sensitivity	1.00 (CI[#] 0.74, 1.00)
Specificity	1.00 (CI[#] 0.96, 1.00)

[#]Confidence intervals were given, however the percentage (e.g. 90% or 95%) was not detailed.

4.5   Data from Bose et al [90]

This cross sectional study was conducted in the USA using nasopharyngeal swab samples collected from 315 patients presenting with influenza-like symptoms. The table of included studies (Appendix V) provides a summary of the important study details, as reported by the authors. Viral culture was performed on the freshly collected swabs and then swabs were then frozen. The PCR test was subsequently conducted at a later (undefined) date. Viral culture samples were identified (typed) as being influenza A positive and were then subsequently subtyped for H1N1 using a type-specific serum, however subtyping details were not provided.

Samples were classified as being H1N1 positive by PCR if the MS2 internal control was positive and the FAM count was >40, with the appropriate amplification curve and a melt temperature within expected values (60^oC +/- 2^oC). Samples were considered to be negative if they did not fulfil those criteria.
A diagnosis had been made for the patients prior to the study based on viral culture results, however no details were given on what constituted a positive viral culture result. In an attempt to limit bias, the samples were de-identified and assayed blindly. The study compared a semi-automated reverse transcriptase polymerase chain reaction assays that allowed reaction products to be visualised in real time (rt RT-PCR) for both influenza A and B viruses. The primers for the assay were directed against highly conserved regions of the haemagglutinin gene.

Patient samples were not categorised using the 2 x 2 table designations and insufficient data was reported to allow one to be constructed, as only the total number of positive results for each test were reported.

The authors did however report sensitivity and specificity values and these together with the raw data is summarised below in Table 13. The study reported an indeterminate result that was excluded from accuracy calculations. This may have resulted in an overestimate in reported specificities and sensitivities.

Table cells containing a dash – denotes fields for which there was no published data. Confidence intervals (95%) were reported for sensitivity and specificity.

Table 13   Data extracted from Bose et al. [90]

```
NOTE:
This figure is included on page 66
of the print copy of the thesis held in
the University of Adelaide Library.
```

4.6   Data from Ginocchio et al [91]

This cross sectional study was conducted in the USA and was conducted using nasopharyngeal swabs from 288 patients presenting with influenza-like illness. The table of included studies (Appendix V) provides a summary of the important study details, as reported by the authors. The aim of the study was to examine the sensitivity and specificity of four different methods of determining H1N1 status. However, three of these tests were types of rapid diagnostic test that were not based on PCR techniques. Details of these three tests are given in Appendix V and only data from the PCR based Respiratory Virus Panel (RVP) assay method are presented in this section. The results of the other tests can be found in the table of included studies (Appendix V). Viral culture was used as the reference standard, however the criteria that defined a positive viral culture result was not included. The RVP assay utilised primers directed against highly conserved regions the matrix gene. Specific primer sequences details were not reported in this study. Explicit details of what constituted a positive result by RVP were not detailed in the report.

Patient samples were not categorised using the 2 x 2 table designations and insufficient data was reported to allow one to be constructed, as only the total number of positive results for each test were reported. No estimates of sensitivity and specificity were reported.

Data extracted from the study is summarised below in Table 14. Table cells containing a dash – denotes fields for which there was no published data.

Table 14   Data extracted from Ginocchio et al [91]

Number of patient samples (N)	288
Patients identified as H1N1 positive by viral culture	133
Patients identified as H1N1 negative by viral culture	-
Patients identified as H1N1 positive by PCR	129
Patients identified as H1N1 negative by PCR	-
Indeterminate/inconclusive	-
Sensitivity	-
Specificity	-

The paper details investigations using various combinations of the four tests - two rapid diagnostic tests, RVP and viral culture. However, only a subset of 288 patient samples underwent all four of the tests. Measures of diagnostic accuracy (sensitivity, specificity, PPV and NPV) are presented for various combinations of the four tests and definitive comparisons were difficult to extract. There was no clear presentation of sensitivity and specificity for RVP relative to viral culture.

4.7   Data from Mokhtari-Azad et al [92]

This cross sectional study was conducted in Iran with the aim of assessing the diagnostic accuracy of a PCR assay that allows the concurrent identification of multiple influenza A subtypes (multiplex assay). This study was conducted on a mixture of fresh nasopharyngeal swabs and washes (161 swabs and 527 washes). On arrival at the laboratory, the sample was divided for viral culture and PCR and both tests were conducted on fresh specimens.

Method details were well described for both viral culture and for PCR. The results from the different types of specimen types were not reported separately for H1N1 and no details were given as to whether or not the specimen type influences detection of H1N1. The table of included studies (Appendix V) provides a summary of the important study details, as reported by the authors.

PCR primers were designed against highly conserved regions of the haemagglutinin gene and the primer sequences and properties are well described in the paper. The products of the PCR reaction were separated on a 1.5% agarose gel according to size using electrophoresis. A sample was considered positive for H1N1 if an ethidium bromide stained band was present at the desired size (944 base pairs). What constituted a positive viral culture was not explicitly stated.

Patient samples were not categorised using the 2 x 2 table designations and insufficient data was reported to allow one to be constructed, as only the total number of positive results for each test were reported. Estimates of sensitivity and specificity were not reported. The raw data taken from Mokhtari-Azad et al summarised below in Table 15. Table cells containing – denotes fields for which there was no published data.

Table 15   Data extracted from Mokhtari et al [92]

Number of patient samples (N)	689
Patients identified as H1N1 by viral culture	17
Patients identified as H1N1 negative by viral culture	-
Patients identified as H1N1 positive by PCR	26
Patients identified as H1N1 negative by PCR	-
Indeterminate/inconclusive	-
Sensitivity	-
Specificity	-

Measures of accuracy (sensitivity, specificity, NPV and PPV) are published for Influenza A in general but were not provided for H1N1. Raw numbers of viral culture positive and PCR positive samples are given (as detailed in the table above), however an additional column to represent samples that were positive for both tests was also given (n=27), however, this value does not appear to be consistent with the published raw data.

4.8   Data from Wu et al [93]

This cross sectional study was conducted in China using throat swabs samples collected from 189 patients presenting with influenza-like symptoms. No details were given regarding sample storage, although viral culture and PCR were undertaken at the same time. The table of included studies (Appendix V) provides a summary of the important study details, as reported by the authors.

The study compared a reverse transcriptase polymerase chain reaction assay that allowed reaction products to be visualised in real time (rtRT-PCR) for both influenza A and B viruses. Several primer sequences were used and the aim to was to develop a multiplex assay capable of identifying the presence of several viruses in the same reaction, at the same time. Products of the rt RT-PCR assay were separated according to their molecular weight using electrophoresis and H1N1 positive samples were identified on a 2% agarose gel. Samples that showed amplified gene products that were dual labelled with the fluorescent markers FAM and HEX, following 40 cycles of amplification were considered to be H1N1 positive by PCR. No details were given as to what constituted a positive viral culture result.

Patient samples were not categorised using the 2 x 2 table designations and insufficient data was reported to allow one to be constructed, as only the total number of positive results for each test were reported. Estimates of sensitivity and specificity were not reported. The raw data taken from Wu et al summarised below in Table 16. Table cells containing – denotes fields for which there was no published data.

Table 16   Data extracted from Wu et al[93]

Number of patient samples (N)	189
Patients identified as H1N1 positive by viral culture	11
Patients identified as H1N1 negative by viral culture	-
Patients identified as H1N1 positive by PCR	5
Patients identified as H1N1 negative by PCR	-
Indeterminate/inconclusive	-
Sensitivity	-
Specificity	-

4.9   Summary of included data

Relevant raw data, as extracted from the included studies is presented in Tables 12-16. The table of included studies (Appendix VI) provides a summary of the important study details for each individual study and salient results are summarised in Table 17.

Table 17   Summary of study data

Study	Gene Target	Sample	Cut off points	Sensitivity	Specificity
Beck et al USA N=100	Matrix (M) gene	Frozen nasopharyngeal swabs	VC – not reported PCR - FAM count was >40,appropriate amplification curve and a melt temperature within expected values (63°C +/- 2°C)	1.00 (CI 0.74, 1.00)	1.00 (CI 0.96, 1.00)
Bose et al USA N=315	Haemagglutinin (H1) gene	Frozen nasopharyngeal swabs	VC – not reported PCR - FAM count was >40,appropriate amplification curve and a melt temperature within expected values (60°C +/- 2°C)	0.95 (CI (0.87, 0.99)	0.95 (CI 0.92,0.98)
Ginocchio et al USA N=288	Matrix (M) gene	Either refrigerated or frozen nasopharyngeal swabs	VC – not reported PCR– not reported	Not reported	Not reported
Mokhtari-Azad et al Iran N=689	Haemagglutinin (H1) gene	Fresh nasopharyngeal swabs or washes	VC – not reported PCR – 944 bp amplification product on a 1.5% agarose gel.	Not reported	Not reported
Wu et al China N=189	Neuraminidase (N1) gene	Fresh throat (?nasopharyngeal)swabs	VC – not reported PCR – amplification products that were dual labelled with the fluorescent markers FAM and HEX, following 40 cycles of amplification	Not reported	Not reported

Likelihood ratios

Likelihood ratios are not presented by either of the two studies that reported sensitivity and specificity, however they are calculated as shown using the formulae below:

Beck et al

Sensitivity = 1.00

Specificity = 1.00

Therefore,

Positive likelihood ratio = $+ve \text{ LR} = \frac{1.00}{1-1.00} \, which = \frac{1.00}{0}$

Negative likelihood ratio = $-ve \text{ LR} = \frac{1-1.00}{1.00} \, which = \frac{0}{1.0}$

Neither likelihood ratio equations yield sensible numbers.

Bose et al

Sensitivity = 0.95

Specificity = 0.95

Therefore,

Positive likelihood ratio = $+ve \text{ LR} = \frac{0.95}{1-0.95} \, which = \frac{0.95}{0.05} = 19$

Negative likelihood ratio = $-ve \text{ LR} = \frac{1-0.95}{0.95} \, which = \frac{0.05}{0.95} = 0.05$

These likelihood ratios can be interpreted to mean that a person infected with influenza A H1N1 is 19 times more likely to get a positive PCR test result than an uninfected person or is very unlikely (0.05 times less likely) to receive a negative PCR test result than an uninfected person

Chapter 5   Discussion

5   Discussion

Viral culture was used as the reference standard for this systematic review as it is considered to be the most accurate method of diagnosing influenza and is the method most commonly used for diagnosis in hospitals globally. [83] Recent CDC guidance appears to suggest a replacement of viral culture with RT-PCR for the diagnosis of influenza A under both routine and pandemic conditions. Exploration of whether this method is at least as good at correctly identifying those with and without H1N1, has not been reported. It was the aim of this systematic review to summarise available research evidence, in support of this recommendation. A recently published literature review [95] with meta-analysis examined the diagnostic test accuracy of available point of care rapid antigen tests for H1N1. In this review RT-PCR was chosen as the reference standard, even though the authors acknowledge viral culture to be the traditional reference test for viruses. Critical appraisal of the included studies was not reported and the review offers little discussion of the methodological quality of the 14 included studies or the decision to statistically combine data from such widely varying tests

In the current review, the QUADAS and STARD checklists were used to assess the methodological quality of the included studies. Broadly, the checklists assess characteristics of the study that can be grouped for discussion into the following sections: patients, conduct and reporting. This chapter will discuss the quality of the included studies based on these three broad study characteristics, and then consider the overall diagnostic test accuracy of PCR methods for H1N1.

5.1   Patient samples

A total of 1581 patients were included across the five studies included in this systematic review. Respiratory samples were collected from patients presenting with influenza-like symptoms in the USA, [89-91] Iran [92] and China. [93] None of the included studies collected patient samples specifically for the study, instead utilising patient samples that had been collected previously for routine diagnosis by viral culture. A consequence of this sample collection method is that the researchers had little control over patient sampling and therefore potential sources of bias associated with patient sampling. Apart from the convenience, a major benefit was that as the samples were already collected, the researchers were not required to obtain patient consent and as a consequence patient withdrawal/attrition was not a concern.

One possible area of concern concerning the samples (and potentially a source of heterogeneity between the studies) is the fact that three of the included studies conducted PCR on previously frozen samples [89-91] and two studies on fresh patient samples. [92, 93] The stability of the H1N1 virus following a freeze-thaw cycle was not discussed in any of these studies.

Only one study gave any description of the patients [91] and that was to comment on the range of patient ages. None of the included studies gave detail on the disease spectrum or range of symptoms of the patients, therefore the potential influence of spectrum and/or selection biases is not able to be assessed. The studies used samples that had been collected from patients presenting with influenza-like symptoms, which is the target population for H1N1.

One aspect of patient sampling that was done well by all of the included studies was that all of the patient samples were tested for both the reference and index tests, regardless of either test result. This aspect ensures a fair comparison of the tests and aims to minimise the presence of selection bias.

5.2   Conduct of the included studies

5.2.1   Reference test methods

It is difficult to comment on the methodological quality of the conduct of the viral culture reference test as details provided by most of the included studies were extremely limited. The reference test was used to confer a diagnosis of H1N1 status for all of the included patient samples within a study; however the details of how the reference tests were conducted were not well reported, simply stating that the samples were H1N1 positive by viral culture. The exceptions were Mokhtari-Azad et al [92] who provided sufficient detail to enable replication of the viral culture and isolation methods used and Wu et al [93] who cited a reference for viral culture methods.

All of the studies reported that the viral culture methods used were likely to correctly identify the target condition of H1N1 and were the same ones used for all of the patient samples within a study. The reference test was conducted independently and did not form part of the index for all of the included studies. In order for the reference test to be replicated, more information is required regarding the cell lines used to grow the H1N1 virus and the culture conditions used (such as culture media, temperature, humidity), as well as the length of time in culture and the methods used to determine infection and to type and subtype the virus(es) present in the patient sample.

In the main (except for [92]), such details were poorly reported and were insufficient to allow replication of viral culture technique used. This lack of detail makes assessment of methodological heterogeneity between studies difficult and prevents the combination of results in meta-analysis.

None of the included studies reported details on what constituted a positive viral culture test result or the threshold/cut-off point that led to patient classification as either infected or not with H1N1. This may have been due to the subjective nature of the reference test (i.e. determining the presence (and extent) of viral infection using judgement by recognising cellular features down a microscope) and the use (and reporting) of an implicit cut off point for classification of positive and negative results. Reporting such as this hinders the potential synthesis of information from similar studies as it remains unknown as to whether the included studies used similar cut-off thresholds.

5.2.2 Index test methods

As the included studies were conducted using samples that had already been used to diagnose patients, the usefulness of clinical data to the interpretation of the index results is unclear. The presence of signs and symptoms in those patients whose samples were frozen for an indeterminate amount of time [89-91] were unlikely to still be evident when the PCR test was conducted, as infection with Influenza A is unlikely to persist much longer than 2 weeks [64]

In the majority of cases, [89-91] the index test was run after an unspecified amount of time after the reference standard which introduces the possibility that the target condition changed between administration of the two tests – a risk of disease progression bias. The fact that the samples were frozen probably preserved the viral nucleic acids prior to extraction so that they did not differ significantly from the nucleic acids present when the reference test was undertaken, however, none of the studies that used frozen samples commented on test performance on frozen versus fresh samples.

None of the studies provided details of the criteria used to determine a positive result by viral culture. This is probably not due to a lack of expertise of the laboratory staffs, as the tests that they conduct are used for clinical diagnosis; instead it may be seen as a failure of reporting.

Except where the index and reference tests were conducted simultaneously,[92, 93] the index test was conducted with the knowledge of the reference test results. For at least three of the included studies, [89-91] a diagnosis had already been made based on the patient samples, prior to conduct of the index test. In an attempt to limit any subjective biases (such as information bias), the patient samples were de-identified prior to conduct of the index tests. No mention was made of whether the samples remained de-identified during the interpretation of the results.

Unlike the reference test, details of how the index test was conducted in each study were well reported for each of the included studies. However, the level of detail provided did vary, with the lowest level of detail being referral to the manufacturer instructions [91] through to full description of amplification and detection steps (Appendix IV). All of the studies provided details of the criteria used to determine a positive result, which was consistent across the studies and was based on the number of amplification cycles to obtain a product of a pre-determined molecular size. Cut-off criteria were given for each index test and are detailed in the table of included studies in Appendix IV.

Test replication in order to determine precision of the index test was reported in all studies. Results that were inconclusive or indeterminate were only reported in two of the five studies. Beck et al [89] reported seven indeterminate patient samples and Bose et al [90] reported one. In both studies the indeterminate results were ignored during sensitivity and specificity calculations, which is likely to have led to an overestimate of test sensitivity. Wu et al [93] reported that 52 patient samples were positive by PCR but not by viral culture but there was no exploration of this by the authors.

5.3 Reporting of included studies

The STARD checklist was published in several journals simultaneously in 2003, with the aim of improving the conduct and reporting of diagnostic test accuracy studies. All of the included studies were published after this date, although only one [90] of the five included studies was published in a journal that had officially promoted the STARD initiative, the remaining four were published in non-STARD journals.

All of the included studies contained index terms associated with diagnostic test accuracy studies such as sensitivity and specificity in the abstract – although none in the title or keywords and none of the five studies included the term "diagnostic test accuracy" in the title, abstract or keywords.

Each of the included studies reported a research aim which was to show equivalence or superiority of a PCR technique to the viral culture method for identifying H1N1 positive patients using respiratory samples (mostly nasopharyngeal swabs). The overarching aim of the studies was to validate the use of a particular PCR technique for diagnosing patients in situations where a rapid result was needed (i.e. pandemic conditions).

As previously discussed in section 4.1, reporting of patient characteristics was poor for all of the included studies. Details on the sampling procedures and the disease spectrum of included patients were equally as poorly reported. The data collection for all of the studies used a prospective design. Description of the reference test was poor and the rationale for its use was not detailed in any study – simply that patients were viral culture positive or diagnosed H1N1 positive using routine viral culture methods. The rationale for using viral culture as the reference test was probably due to it being the laboratory standard for diagnosis within the hospital system under typical (i.e. non-pandemic) conditions [93] and that the patient samples were readily available and the patients diagnosed. The technical specifications of the materials and methods required for the index test were generally well reported, however the same cannot be said for details of the reference test, as discussed in section 4.2. This may be due to familiarity of the authors with viral culture (i.e. the techniques used are used routinely and as such are not novel), which may have resulted in their full description being overlooked or thought of as "uninteresting" by the authors – especially when most journals set word limits for included articles. This approach is unhelpful to those undertaking systematic reviews of diagnostic test accuracy.

Viral culture uses an implicit cut-off level to determine what constitutes a positive result, therefore the technical expertise and the experience of the technician could be an important factor which may vary between and/or within the studies, which is one reason why a summary estimate of sensitivity and specificity were not calculated. None of the included studies reported on the training or expertise of the technician or whether they were blinded at any stage.

The focus of data reporting for all of the included studies was how well the PCR assay identified patients who tested positive for H1N1 by viral culture, in order to determine sensitivity and specificity. In order to be able to calculate these measures of accuracy, either a cross classification table (**Error! Reference source not found.**) or the raw data to generate such a table, need to be reported. None of the studies presented test performance data as a cross-classification table or reported a complete data set to be able to generate such a table. The major omission was the number of true and false negative results, preventing the calculation of specificity. Two of the included studies estimated sensitivity and

specificity [89, 90], however one stated a confidence interval but did not report the level of that confidence interval. [89] Another included study gave no statistical estimate of test accuracy. [93]

5.4   Diagnostic testing accuracy of PCR for Influenza A H1N1

The objective of the review was to generate a summary estimate of diagnostic test accuracy for laboratory tests for H1N1. This was not possible due to the incomplete reporting in the primary studies, as mentioned in the previous section (4.3) the major flaws being; the lack of information as to the proportion of positive results that were true positives and false positives, and the omission of true and false negative results. This may have been at least in part due to the focus some of the included being on influenza A in general and not specifically on the H1N1 subtype.

The ideal diagnostic test should have high sensitivity and high specificity so that a clinician can be sure that the test result accurately reflects the true disease status of the patient. A cross classification of patients who tested positive and negative for both the reference and index test is necessary to calculate sensitivity and specificity measures.

The research aims of the included studies were to establish equivalence or superiority of a PCR index test, compared with a viral culture reference test. Three [89-91] of the five included studies concluded that PCR was a highly accurate way of identifying H1N1 positive patients and the remaining two studies that PCR was superior, [92, 93] as PCR gave more H1N1 positive results than the viral culture method, however only two of the five included studies [89, 90] reported sensitivity and specificity values. The two studies that suggest superiority of PCR [92, 93] did not discuss the implications of their findings in terms of the possibility of false positive results and the subsequent impact on test accuracy measures.

Only two [89, 90] of the five studies reported sensitivity and specificity estimates. If taken at face value, the estimates would suggest that PCR is an accurate method of identifying H1N1 positive patients, compared with viral culture. Using the calculated positive likelihood ratio from Bose et al, a person infected with H1N1 would be 19 times more likely to receive a positive test result that a person not infected with H1N1, which in clinical practice may sound convincing. However, both Beck et al and Bose et al reported indeterminate results that were excluded for accuracy estimates – seven samples for [89] and one sample for [90]This exclusion of results may lead to an overestimate of the true test sensitivity and specificity and perhaps may have been more accurately reflected in a 3 x 2 table, as recommended for use in such situations by the Centre for Disease Control and Prevention [96] which would allow a more transparent examination of how the indeterminate results influence the test accuracy.

The finding that PCR assays were able to identify 9 more (out of a total of 43 positive results) [92] and 6 more (out of a total of 16 positive results) [93] patients as being H1N1 positive than viral culture methods was not explored by either paper and both studies concluded that PCR was superior to viral culture, without further discussion of the possibility (and clinical implications) of false positive results. The possibility that this increase in number is due to cross reactivity with other similar sequences of nucleic acids cannot be ruled out with the available data set and as such the results may represent misclassification of the patients and be considered false positive results.

The possibility that PCR may be superior to viral culture cannot be ruled out either. As discussed in section 1.2, identification of a particular protein of interest (H1N1 in this case) is based on identifying a specific sequence of nucleic acids of the "message" that leads to synthesis of that protein. [84] As a technique, PCR is more sophisticated than viral culture as it can identify the presence of H1N1 (in this case) from a fragment of genetic material, whereas viral culture requires the virus not only to be intact but also to be viable. [87] On this basis, PCR should therefore be able to detect the presence of H1N1 from a much lower presence (titre) of virus than a viral culture technique. One disadvantage though of this increased sophistication is that PCR techniques make no distinction between active and inactive virions, which may lead to overestimates of infectivity if inactive virions (and therefore the risk of subsequent infection) are present.

Chapter 6   Conclusions

6   Summary

All of the included studies reported clinical relevance of their findings in terms of how accurate their particular PCR assay was at identifying H1N1 positive patients. All of the included studies concluded that PCR was at least as accurate (based on sensitivity and specificity estimates) if not more accurate than viral culture. This conclusion should, however, be interpreted with caution, given the low number (five) of studies and the methodological limitations of those studies.

There is limited evidence (two individual studies) to suggest that polymerase chain reaction diagnostic tests are an accurate way to diagnose H1N1. The estimates of sensitivity and specificity were not pooled in meta-analysis for these studies as we were unable to ascertain whether the reference test methods were similar enough for this to be appropriate, due to poor quality reporting.

This systematic review highlights the need for better reporting and indexing of studies that examine the diagnostic accuracy of laboratory tests and or H1N1. Publication of the STARD and QUADAS initiatives have helped but this rapidly expanding area of research needs structure and consistency in order for its findings to be useful to clinical decision makers.

6.1   Recommendations for Practice and Research

This systematic review makes suggestions, in alignment with the QUADAS and STARD initiatives, based on the five included studies.

Patient samples to be used for DTA studies should be specifically collected for the purpose conducting a diagnostic test accuracy study, in order to avoid sampling related errors. This would also provide the opportunity for the researcher to collect basic demographic information about the patient. Authors of DTA studies should include basic descriptions of included patients, to allow better determination of whether studies have drawn from comparable populations.

Additionally, reporting of the reference test should be in sufficient detail to allow replication of the method used. Cut-off criteria used to determine the boundary between a positive and negative result should be explicitly stated.

6.2   Limitations of the review

The objective of this systematic review was to identify the best available evidence to generate a summary estimate of the diagnostic test accuracy of laboratory tests for Influenza A H1N1 compared to a viral culture reference standard. There are however some limitations of this review. The search was restricted to English language studies up to May 2010 which may have resulted in the exclusion of potentially relevant studies published in other languages.

Previous reports suggest that identifying studies that assess diagnostic accuracy can be challenging to locate due to poor indexing or application of keywords, therefore it is a potential limitation of this review that relevant studies were not identified. Attempts were made to combat this by using a deliberately broad search strategy.

Chapter 7   References

1. Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC*, et al.* Guidelines for meta-analyses evaluating diagnostic tests. Annals of Internal Medicine. 1994 April 15, 1994;120(8):667-76.

2. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. CMAJ. 1986 Mar 15;134(6):587-94.

3. Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F*, et al.* Rapid molecular detection of tuberculosis and rifampin resistance. New England Journal of Medicine. 2010;363(11):1005-15.

4. World Health Organization. WHO endorses new rapid tuberculosis test2010: Available from: http://www.who.int/tb/features_archive/new_rapid_test/en/index.html.

5. Knottnerus JA, van Weel C. General introduction: evaluation of diagnostic procedures. In: Knottnerus JA, editor. The Evidence Base of Clinical Diagnosis. London: BMJ Books; 2002. p. 1-18.

6. Glass G. Primary, secondary and meta-analysis of research. Education Research. 1976;10:3-8.

7. Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to meta-analysis. first ed. Chipenham: Wiley; 2009.

8. Leeflang M, Deeks J, Gatsonis C, Bossuyt P. Systematic reviews of diagnostic accuracy. Annals of internal medicine. 2008;149:889-97.

9. Irwig L, Macaskill P, Glasziou PP, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol. 1995;48(1):119-30.

10. Pai M, McCulloch M, Enanoria W, Colford J. Systematic reviews of diagnostic test evaluations: what's behind the scenes? EBM Online [serial on the Internet]. 2004; 9.

11. Deeks J. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Davey Smith G, Altman D, editors. Systematic reviews in healthcare: Meta-analysis in context: BMJ publishing group; 2001. p. 248-81.

12. Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: Methodologic primer. American Journal of Roentgenology. 2006;187:271-81.

13. Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. J Clin Epidemiol. [Meta-Analysis Research Support, Non-U.S. Gov't]. 1999 Oct;52(10):943-51.

14. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. BMJ. 2001 Jul 21;323(7305):157-62.

15. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med. 1987;107(224-33).

16. Vamvakas EC. Meta-analyses of studies of the diagnostic accuracy of laboratory tests: a review of the concepts and methods. Arch Pathol Lab Med. 1998 Aug;122(8):675-86.

17. White S, Schultz T, Tufanaru C. The accuracy of Influenza A H1N1 (Swine Flu) laboratory testing: a systematic review of diagnostic test accuracy. The Joanna Briggs Institute Library of Systematic Reviews. 2011;(protocol).

18. Honest H, Khan KS. Reporting of measures of accuracy in systematic reviews of diagnostic literature. BMC Health Serv Res. 2002;2:4.

19. Brazzelli M, Sandercock PAG, Chappell FM, Celani MG, Righetti E, Arestis N*, et al.* Magnetic resonance imaging versus computed tomography for detection of acute vascular lesions in patients presenting with stroke symptoms. Cochrane Database of Systematic Reviews [serial on the Internet]. 2009; (4): Available from: http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD007424/frame.html.

20. Leeflang M, Debets-Ossenkopp YJ, Visser Caroline E, Scholten RJPM, Hooft L, Bijlmer HA*, et al.* Galactomannan detection for invasive aspergillosis in immunocompromized patients.

Cochrane Database of Systematic Reviews [serial on the Internet]. 2008; (4): Available from: http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD007394/frame.html.

21. van Leeuwen M, Opmeer BC, Yilmaz Y, Limpens J, Serlie MJ, Mol BW. Accuracy of the random glucose test as screening test for gestational diabetes mellitus: a systematic review. Eur J Obstet Gynecol Reprod Biol. 2011 Feb;154(2):130-5.

22. The Cochrane Collaboration. The Cochrane Library. The Cochrane Collaboration; 2011 [cited 2011 June 21]; Available from: http://www.thecochranelibrary.com/view/0/index.html.

23. Raj A, de Verteuil R. Systematic review of the diagnostic accuracy of the single, two and three field digital retinal photography for screening diabetic retinopathy The Joanna Briggs Institute Library of Systematic Reviews. 2011;9 (16):491-537.

24. Johnson RL, Bungo MW. The diagnostic accuracy of exercise electrocardiography--a review. Aviat Space Environ Med. 1983 Feb;54(2):150-7.

25. Thacker SB, Berkelman RL. Assessing the diagnostic accuracy and efficacy of selected antepartum fetal surveillance techniques. Obstet Gynecol Surv. 1986 Mar;41(3):121-41.

26. Detrano R, Janosi A, Lyons KP, Marcondes G, Abbassi N, Froelicher VF. Factors affecting sensitivity and specificity of a diagnostic test: the exercise thallium scintigram. Am J Med. 1988 Apr;84(4):699-710.

27. Shapiro DE. Issues in combining independent estimates of sensitivity and specificity of a diagnostic test. Acad Radiol. 1995;2(s37-s47).

28. Diagnostic Test Accuracy Working Group. Diagnostic Test Accuracy Working Group. Diagnostic Test Accuracy Working Group; 2011 [cited 2011 21 June]; Available from: http://srdta.cochrane.org/more-about-us.

29. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Analysing and presenting resuits. In: Deeks J, Bossuyt P, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy: The Cochrane Collaboration; 2010.

30. White S, Schultz T. The accuracy of lnfluenza A H1N1 (Swine Flu) laboratory testing: A systematic review of diagnostic test accuracy. The Joanna Briggs Institute LIbrary of Systematic Reviews. 2011((In Press)).

31. Deville WL, Buntinx F, Bouter LM, Montori VM, de Vet HC, van der Windt DA*, et al.* Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol. 2002 Jul 3;2:9.

32. Bossuyt P, Reitsma J, Bruns D, Gatsonis C, Gatsonis C, Irwig L*, et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Annals of Internal Medicine. 2003;138:40-4.

33. Courtney M. Evidence for nursing practice: Elsevier Churchill Livingstone; 2004.

34. Zhuo X, Obuchowski N, McClish D. Statistical methods in diagnostic medicine. New York: Wiley; 2002.

35. Virgili G, Conti A, Murro V, Gensini G, Gusinu R. Systematic reviews of diagnostic test accuracy and the Cochrane collaboration. Internal Emerg Medicine. 2009;4:255-8.

36. Grimes D, Schulz K. Refining clinical diagnosis with likelihood ratios. Lancet. 2005;365:1500-5.

37. Chalmers I, Hedges L, Cooper H. A brief history of research synthesis. Evaluation and the Health Professional. 2002;25(1):12-37.

38. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information and Libraries Journal. 2009;26:91-108.

39. Tricco AC, Tetzlaff J, Moher D. The art and science of knowledge synthesis. Journal of Clinical Epidemiology. 2010;In Press, Corrected Proof.

40. Hatala R, Keitz S, Wyer P, Guyatt G. Tips for learners of evidence-based medicine: 4. Assessing heterogeneity of primary studies in systematic reviews and whether to combine their results. CMAJ. 2005;172(5):661-5.

41. Tatsioni A, Zarin D, Aronson N, Samson D, Flamm C, Schmid C*, et al.* Challenges in systematic reviews of diagnostic technologies. Annals of internal medicine. 2005;142:1048-55.

42. Hasselblad V. Meta-analysis of screening and diagnostic tests. Psychological Bulletin. 1995;117(1):167-78.

43. Lijmer J, Bossuyt P, Heisterkamp S. Exploring sources of heterogeniety in systematic reviews of diagnostic tests. Statistics in Medicine. 2002;21:1525-37.

44. Mallett S, Deeks J, Halligan S, Hopewell S, Cornelius V, Altman G. Systematic reviews of diagnostic tests in cancer: review of methods and reporting. BMJ [serial on the Internet]. 2006; 333(413).

45. Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol. 2006 Mar;59(3):234-40.

46. Walter SD, Jadad AR. Meta-analysis of screening data: a survey of the literature. Stat Med. 1999;18:3409-24.

47. Garg A, Iansavichus A, Wilczynski N, Kastner M, Baier L, Shariff S*, et al.* Filtering Medline for a clinical discipline: diagnostic test assessment framework. BMJ [serial on the Internet]. 2009; 39.

48. Haynes R, Wilczynski N. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. BMJ [serial on the Internet]. 2004.

49. Ritchie G, Glanville J, Lefebvre C. Do published search filters to identify diagnostic test accuracy studies perform adequately? Health Information and Libraries Journal. 2007;24:188-92.

50. Whiting P, Harbord R, de Salis I, Egger M, Sterne J. Evidence-based diagnosis. J Health Serv Res Policy. 2008 Oct;13 Suppl 3:57-63.

51. Castelli F, Caligaris S, Gulletta M, el-Hamad I, Scolari C, Chatel G*, et al.* Malaria in migrants. Parassitologia. 1999;41(1-3):261-5.

52. Mower W. Evaluating Bias and Variability in Diagnostic Test Reports. Annals of Emergency Medicine. 1999;33:85-91.

53. Parekh-Bhurke S, Kwok CS, Pang C, Hooper L, Loke YK, Ryder JJ*, et al.* Uptake of methods to deal with publication bias in systematic reviews has increased over time, but there is still much scope for improvement. J Clin Epidemiol. 2011 Apr;64(4):349-57.

54. Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. J Clin Epidemiol. 2008 Apr;61(4):357-64.

55. Reitsma J, Whiting P, Vlassov V, Leeflang M, Deeks J. Chapter 9: Assessing methodological quality. . In: Deeks JJ GC, editor. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. York: The Cochrane Collaboration; 2009.

56. Lohr K. Assessing "best evidence": issues in grading the quality of studies for systematic reviews The joint commission journal on quality improvement 25(9): 470-479. The joint commission journal on quality improvement 1999;25(9):470-9.

57. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology. 2003;3(25):1-13.

58. Harbord RM, Whiting P, Sterne JA, Egger M, Deeks JJ, Shang A*, et al.* An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. J Clin Epidemiol. 2008 Nov;61(11):1095-103.

59. Meyer G. Guidelines for reporting information in studies of diagnostic accuracy: The STARD initiative. Journal of Personality Assessment. 2003;81(3):191-3.

60. Whiting P, Weswood M, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Medical Research Methodology. 2006;6(9):1-8.

61. World Health Organisation. WHO information for laboratory diagnosis of pandemic (H1N1) 2009 virus in humans (revised 23/11/2009). World Health Organisation; 2009 [updated 23/11/2009; cited 2010 27 January 2010]; WHO Evidence-based advice for diagnosing H1N]. Available from: http://www.who.int/csr/resources/publications/swineflu/diagnostic_recommendations/en/index.html.

62. World Health Organisation. World now at the start of 2009 influenza pandemic. WHO; 2009 [cited 2010 Oct 10th 2010]; Available from: http://www.who.int/mediacentre/news/statements/2009/h1n1_pandemic_phase6_20090611/en/index.html.

63. Baker F, Silverton R. Introduction to medical laboratory technology 6th ed. 6 ed. Baker F, editor. Sevenoaks: Butterworths; 1985.

64. Tang J, Shetty N, Lam T, Hon K. Emerging, novel and known Influenza virus infections in humans. Infect Dis Clin N Am. 2010;24:603-17.

65. Potter C. The history of influenza. Journal of Applied Microbiology 2001;91:572-9.

66. Taubenberger J, Morens D. Influenza: The once and future pandemic. Public Health Reports. 2010;125(Supp 125):16-26.

67. Lamb R, Choppin P. The gene structure and replication of influenza virus. Ann Rev Biochem. 1983;52:467-506.

68. Shaw M, Palese P. Orthomyxoviruses: Molecular Biology. Elsevier; 2008. p. 483-9.

69. Rachakonda S. STUDIES ON CONFORMATIONAL STABILITY OF THE ECTODOMAIN OF INFLUENZA VIRUS HEMAGGLUTININ Berlin2005 [cited 2011 May 9th 2011]; Available from: http://edoc.hu-berlin.de/dissertationen/rachakonda-p-sivaramakrishna-2005-02-15/HTML/chapter1.html.

70. Farley M. 2009 H1N1 Influenza: a twenty-first century pandemic with roots in the early twentieth century. The American Journal of the Medical Sciences. 2010;340(3):202-8.

71. Webster RG, Laver WG. The origin of pandemic influenza. . Bulletin of the World Health Organisation. 1972;46:449-52.

72. New Scientist. Deadly new flu virus in US and Mexico may go pandemic. 2009 [cited 2009 26/09/2009]; Available from: http://www.newscientist.com/article/dn17025-deadly-new-flu-virus-in-us-and-mexico-may-go-pandemic.html. .

73. Igarashi M, Ito K, Yoshida R, Tomabechi D, Kida H, Takada A. Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. PLoS One. 2010;5(1):e8553.

74. Lamb R. Influenza. Elsevier; 2008. p. 95-104.

75. Noble S, McGregor MS, Wentworth DE, Hinshaw VS. Antigenic and genetic conservation of the haemagglutinin in H1N1 swine influenza viruses. J Gen Virol. 1993 Jun;74 (Pt 6):1197-200.

76. Veljkovic V, Niman HL, Glisic S, Veljkovic N, Perovic V, Muller CP. Identification of hemagglutinin structural domain and polymorphisms which may modulate swine H1N1 interactions with human receptor. BMC Struct Biol. 2009;9:62.

77. Scalera NM, Mossad SB. The first pandemic of the 21st century: a review of the 2009 pandemic variant influenza A (H1N1) virus. Postgrad Med. 2009 Sep;121(5):43-7.

78. Loens K, Goossens H, Ieven M. Acute respiratory infection due to Mycoplasma pneumoniae: current status of diagnostic methods. European Journal of Clinical Microbiology and Infectious Diseases. 2010;29(9):1055-69.

79. Fleming DM. Influenza diagnosis and treatment: a view from clinical practice. Philos Trans R Soc Lond B Biol Sci. 2001 Dec 29;356(1416):1933-43.

80. Centre for Disease Control and Prevention. Evaluation of rapid influenza diagnostic tests for detection of novel influenza A (H1N1) Virus - United States, 2009. MMWR Morb Mortal Wkly Rep. 2009 Aug 7;58(30):826-9.

81. Centre for Disease Control and Prevention. Interim Recommendations for Clinical Use of Influenza Diagnostic Tests During the 2009-10 Influenza Season Atlanta: Centre for Disease Control (USA); 2009 [cited 2010 27/01/2010]; Available from: http://www.cdc.gov/h1n1flu/guidance/diagnostic_tests.htm.

82. Centre for Disease Control and Prevention. Interim guidance for influenza diagnostic testing during the 2008-2009 influenza season. Atlanta: CDC; 2009 [cited 2009 18/09/2009]; Available from: http://www.cdc.gov/flu/professionals/diagnosis/index.htm.

83. Cox N, Subbarao K. Influenza. Lancet. 1999;354:1277-82.

84. Murdoch D, Jennings L, Bhat N, Anderson T. Emerging advances in diagnostics of respiratory infections. Infect Dis Clin N Am. 2010;24:791-807.

85. Uyeki TM. Influenza diagnosis and treatment and children: a review of studies on clinically useful tests for infleunza. Pediatric Infectious Disease Journal. 2003;22:164-77.

86. Viljoen GJ. Molecular diagnostic PCR handbook Springer; 2005.

87. Leland D, Ginocchio CC. Role of cell culture for virus detection in the age of technology. Clinical Microbiology Reviews. 2007;20(1):49-78.

88. de Vet H, Eisinga A, Aertgeerts B, Pewsner D. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy; version 0.4. Collaboration TC, editor: Elsevier; 2008.

89. Beck ET, Jurgens LA, Kehl SC, Bose ME, Patitucci T, Lague E*, et al.* Development of a Rapid Automated Influenza A, Influenza B, and Respiratory Syncytial Virus A/B Multiplex Real-Time RT-PCR Assay and Its Use during the 2009 H1N1 Swine-Origin Influenza Virus Epidemic in Milwaukee, Wisconsin. Journal of Molecular Diagnostics. 2010 Jan;12(1):74-81.

90. Bose ME, Beck ET, Ledeboer N, Kehl SC, Jurgens LA, Patitucci T*, et al.* Rapid Semiautomated Subtyping of Influenza Virus Species during the 2009 Swine Origin Influenza A H1N1 Virus Epidemic in Milwaukee, Wisconsin. Journal of Clinical Microbiology. 2009 Sep;47(9):2779-86.

91. Ginocchio C, Lotlikar M, Falk L, Arora S, Kowerska M, Bornfreund M*, et al.* Clinical performance of the 3M rapid detection Flu A +B test compared to R-Mix culture, DFA and BinaxNOW Influenza A&B test. Journal of Clinical Virology. 2009;45:146-9.

92. Mokhtari-Azad T, Rezaie-Khollari F, Nadji AR, Salimi V, Noroozbahaiet Z, Mirzaie AR*, et al.* Comparison of multiplex nested RT-PCR with virus isolation for detection of influenza viruses A and B. Iranian Journal of Public Health. 2007;36(3):1-7.

93. Wu CL, Cheng XW, He HF, Lv X, Wang JW, Deng RQ*, et al.* A multiplex real-time RT-PCR for detection and identification of influenza virus types A and B and subtypes H5 and N1. Journal of Virological Methods. 2008 Mar;148(1-2):81-8.

94. Ginocchio CC, Zhang F, Manji R, Arora S, Bornfreund M, Falk L*, et al.* Evaluation of multiple test methods for the detection of the novel 2009 influenza A (H1N1) during the New York City outbreak. Journal of Clinical Virology. 2009 Jul;45(3):191-5.

95. Babin S, Hsieh Y-H, Rothman R, Gaydos C. A meta-analysis of point-of-care laboratory tests in the diagnosis of novel 2009 swine-lineage pandemic influenza A (H1N1). Diagnostic Microbiology and Infectious Disease. 2011;89:410-8.

96. Centre for Disease Control and Prevention. Systematic Reviews: CRD's guidance for undertaking reviews in healthcare. 2008 [cited 2009 9 June].

97. Bouscambert-Duchamp M, Billaud G, Enouf V, Valette M, van der Werf S, Lina B. Evaluation of sensitivity of commercial kits for the diagnosis of A(H1N1)v infection in comparison with RT-PCRs developed by national influenza centers. Journal of Clinical Virology. 2009 Sep;46:S4-S.

98. Chan KH, Lai ST, Poon LLM, Guan Y, Yuen KY, Peiris JSM. Analytical sensitivity of rapid influenza antigen detection tests for swine-origin influenza virus (H1N1). Journal of Clinical Virology. 2009 Jul;45(3):205-7.

99. Chan KH, Lam SY, Puthavathana P, Nguyen TD, Long HT, Pang CM*, et al.* Comparative analytical sensitivities of six rapid influenza A antigen detection test kits for detection of influenza A subtypes H1N1, H3N2 and H5N1. J Clin Virol. 2007 Feb;38(2):169-71.

100.   Daum LT, Canas LC, Schadler CA, Ujimori VA, Huff WB, Barnes WJ, *et al.* A rapid, single-step multiplex reverse transcription-PCR assay for the detection of human H1N1, H3N2, and B influenza viruses. J Clin Virol. 2002 Dec;25(3):345-50.

101.   Drexler JF, Helmer A, Kirberg H, Reber U, Panning M, Muller M, *et al.* Poor Clinical Sensitivity of Rapid Antigen Test for Influenza A Pandemic (H1N1) 2009 Virus. Emerging Infectious Diseases. 2009 Oct;15(10):1662-4.

102.   Hall RJ, Peacey M, Huang QS, Carter PE. A rapid method to support the diagnosis of swine-origin influenza virus (S-OIV), by sequencing of real-time PCR amplicon from diagnostic assays. J Clin Microbiol. 2009 Jul 8.

103.   Kessler N, Ferraris O, Palmer K, Marsh W, Steel A. Use of the DNA flow-thru chip, a three-dimensional biochip, for typing and subtyping of influenza viruses. Journal of Clinical Microbiology. 2004 May;42(5):2173-85.

104.   Kubo T, Agoh M, Mai LQ, Fukushima K, Nishimura H, Yamaguchi A, *et al.* Development of a Reverse Transcription-Loop-Mediated Isothermal Amplification Assay for Detection of Pandemic (H1N1) 2009 Virus as a Novel Molecular Method for Diagnosis of Pandemic Influenza in Resource-Limited Settings. J Clin Microbiol. 2010 March 1, 2010;48(3):728-35.

105.   Schweiger B, Zadow R, Heckler H, Timm J, Pauli G. a fluorogenic PCR assay for typing and subtyping of influenza viruses in respiratory. Journal of Clinical Microbiology. 2000 01/01;38(4):1552-8.

106.   Tucker SP, Cox C, Steaffens J. A flu optical immunoassay (ThermoBioStar's FLU OIA): a diagnostic tool for improved influenza management. Philos Trans R Soc Lond B Biol Sci. 2001 Dec 29;356(1416):1915-24.

107.   Whiley DM, Bialasiewicz S, Bletchly C, Faux CE, Harrower B, Gould AR, *et al.* Detection of novel influenza A(H1N1) virus by real-time RT-PCR. Journal of Clinical Virology. 2009 Jul;45(3):203-4.

108.   Agoritsas K, Mack K, Bonsu BK, Goodman D, Salamon D, Marcon MJ. Evaluation of the Quidel QuickVue test for detection of influenza A and B viruses in the pediatric emergency medicine setting by use of three specimen collection methods. J Clin Microbiol. 2006 Jul;44(7):2638-41.

109.   Alexander R, Hurt AC, Lamb D, Wong FY, Hampson AW, Barr IG. A comparison of a rapid test for influenza with laboratory-based diagnosis in a paediatric population. Commun Dis Intell. 2005;29(3):272-6.

110.   Atmar RL, Baxter BD, Dominguez EA, Taber LH. Comparison of reverse transcription-PCR with tissue culture and other rapid diagnostic assays for detection of type A influenza virus. J Clin Microbiol. 1996 Oct;34(10):2604-6.

111.   Bolotin S, De Lima C, Choi KW, Lombos E, Burton L, Mazzulli T, *et al.* Validation of the TaqMan Influenza A Detection Kit and a Rapid Automated Total Nucleic Acid Extraction Method to Detect Influenza A Virus in Nasopharyngeal Specimens. Annals of Clinical and Laboratory Science. 2009 Spr;39(2):155-9.

112.   Boon ACM, French AMF, Fleming DM, Zambon MC. Detection of influenza A subtypes in community-based surveillance. Journal of Medical Virology. 2001 Sep;65(1):163-70.

113.   Booth S, Baleriola C, Rawlinson WD. Comparison of two rapid influenza A/B test kits with reference methods showing high specificity and sensitivity for influenza A infection. J Med Virol. 2006 May;78(5):619-22.

114.   Carraro E, Neto DF, Benfica D, SittaPerosa AH, Granato CF, Bellei NC. Applications of a duplex reverse transcription polymerase chain reaction and direct immunofluorescence assay in comparison with virus isolation for detection of influenza A and B. Diagn Microbiol Infect Dis. 2007 Jan;57(1):53-7.

115.   Cazacu AC, Chung SE, Greer J, Demmler GJ. Comparison of the directigen flu A+B membrane enzyme immunoassay with viral culture for rapid detection of influenza A and B viruses in respiratory specimens. J Clin Microbiol. 2004 Aug;42(8):3707-10.

116. Dunn JJ, Gordon C, Kelley C, Carroll KC. Comparison of the Denka-Seiken INFLU A.B-Quick and BD Directigen Flu A+B kits with direct fluorescent-antibody staining and shell vial culture methods for rapid detection of influenza viruses. J Clin Microbiol. 2003 May;41(5):2180-3.

117. Ellis JS, Fleming DM, Zambon MC. Multiplex reverse transcription-PCR for surveillance of influenza A and B viruses in England and Wales in 1995 and 1996. J Clin Microbiol. 1997 Aug;35(8):2076-82.

118. Herrmann B, Larsson C, Zweygberg BW. Simultaneous detection and typing of influenza viruses A and B by a nested reverse transcription-PCR: comparison to virus isolation and antigen detection by immunofluorescence and optical immunoassay (FLU OIA). J Clin Microbiol. 2001 Jan;39(1):134-8.

119. Kehl SC, Henrickson KJ, Hua W, Fan J. Evaluation of the Hexaplex assay for detection of respiratory viruses in children. J Clin Microbiol. 2001 May;39(5):1696-701.

120. Mehlmann M, Bonner AB, Williams JV, Dankbar DM, Moore CL, Kuchta RD, *et al.* Comparison of the MChip to viral culture, reverse transcription-PCR, and the QuickVue influenza A plus B test for rapid diagnosis of influenza. Journal of Clinical Microbiology. 2007 Apr;45(4):1234-7.

121. Pachucki CT. Rapid Tests for Influenza. Curr Infect Dis Rep. 2005 May;7(3):187-92.

122. Quach C, Newby D, Daoust G, Rubin E, McDonald J. QuickVue influenza test for rapid detection of influenza A and B viruses in a pediatric population. Clin Diagn Lab Immunol. 2002 Jul;9(4):925-6.

123. Rahman M, Vandermause MF, Kieke BA, Belongia EA. Performance of Binax NOW Flu A and B and direct fluorescent assay in comparison with a composite of viral culture or reverse transcription polymerase chain reaction for detection of influenza infection during the 2006 to 2007 season. Diagn Microbiol Infect Dis. 2008 Oct;62(2):162-6.

124. Ruest A, Michaud S, Deslandes S, Frost EH. Comparison of the Directigen flu A+B test, the QuickVue influenza test, and clinical case definition to viral culture and reverse transcription-PCR for rapid diagnosis of influenza virus infection. J Clin Microbiol. 2003 Aug;41(8):3487-93.

125. Simmerman JM, Chittaganpitch M, Erdman D, Sawatwong P, Uyeki TM, Dowell SF. Field performance and new uses of rapid influenza testing in Thailand. Int J Infect Dis. 2007 Mar;11(2):166-71.

126. Smith AB, Mock V, Melear R, Colarusso P, Willis DE. Rapid detection of influenza A and B viruses in clinical specimens by Light Cycler real time RT-PCR. J Clin Virol. 2003 Sep;28(1):51-8.

127. Uyeki TM, Prasad R, Vukotich C, Stebbins S, Rinaldo C, Ferng Y, *et al.* Low sensitivity of rapid diagnostic test for influenza. Clinical Infectious Diseases. 2009;48:e89-e92.

128. Vabret A, Sapin G, Lezin B, Mosnier A, Cohen JM, Burnouf L, *et al.* Comparison of three non-nested RT-PCR for the detection of influenza A viruses. Journal of Clinical Virology. 2000 Sep;17(3):167-75.

129. Carr M, Rory G, Alasdair M, Suzie C, Margaret F, Mary S. Development of a real-time RT-PCR for the detection of swine-lineage influenza A (H1N1) virus infections. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology. 2009 06/23.

130. Choi Y, Kim H, Park J, Oh M, Nam H, Kim Y, *et al.* Evaluation of new rapid antigen test for detection of pandemic influenza A/H1N1 2009 virus. Journal of Clinical Microbiology. 2010;48(6):2260-2.

131. Ciblak M, Kanturvardar M, Asar S, Bozkaya E, Yenen O, Badur S. Sensitivity of rapid influenza antigen tests in the diagnosis of pandemic (H1N1) 2009 compared with the standard of rRT-PCR technique during the 2009 pandemic in Turkey. Scandinavian Journal of Infectious Diseases. 2010:e1-e4.

132. de la Tabla V, Antequera P, Masia M, Ros P, Martin C, Gazquez G, *et al.* Clilnical evaluation of rapid point of care testing for detecting of novel influenza A (H1N1) virus in a population based study in Spain. Clinical Microbiology and Infection. 2009.

133.   Faix DJ, Sherman SS, Waterman SH. Rapid-Test Sensitivity for Novel Swine-Origin Influenza A (H1N1) Virus in Humans. New England Journal of Medicine. 2009 Aug;361(7):728-9.

134.   He J, Bose ME, Beck ET, Fan J, Tiwari S, Metallo J*, et al.* Rapid Multiplex RT-PCR typing of influenza A and B, and subtyping of influenza A into H1, 2, 3, 5, 7, 9, N1 (human), N1 (animal), N2 and N7 including typing of novel swine-origin Influenza A (H1N1) Virus during current 2009 outbreak in Milwaukee, Wisconsin. J Clin Microbiol. 2009 Jul 29.

135.   Herzum I, Lutz T, Koch F, Geisel R, Gehrt A. Diagnostic performance of rapid influenza antigen assays in patients infected with the new influenza A (H1N1) virus. Clinical Chemistry and Laboratory Medicine. 2010 Jan;48(1):53-6.

136.   Jenny S, Hu Y, Overduin P, Meijer A. Evaluation of the Xpert Flu A panel nucleic acid amplification based point of care test for influenza A virus detection and pandemic H1 subtyping. Journal of Clinical Virology. 2010;49:85-9.

137.   Likitnukul S, Boonsiri K, Tangsuksant Y. Evaluation of Sensitivity and Specificity of Rapid Influenza Diagnostic Tests for Novel Swine-Origin Influenza A (H1N1) Virus. Pediatric Infectious Disease Journal. 2009 Nov;28(11):1038-9.

138.   Louie J, Guevara H, Boston E, Dahlke M, Nevarez M, Kong T*, et al.* Rapid antigen test for diagnosis of pandemic (H1N1) 2009. Emerging Infectious Diseases. 2010;16(5):824-6.

139.   Troppan K, Bozic M, Santner B, Kressler H. Evaluation of four molecular assays for detection of pandemic influenza A (H1N1) 2009 virus in the routine diagnostic laboratory. Journal of Clinical Virology. 2010;49:82-4.

140.   Vasoo S, Stevens J, Singh K. Rapid Antigen Tests for Diagnosis of Pandemic (Swine) Influenza A/H1N1. Clinical Infectious Diseases. 2009 Oct;49(7):1090-3.

141.   Arya SC, Agarwal N. Apropos 'performance of influenza point-of-care tests in the detection of pandemic (H1N1) 2009 influenza viruses'. Influenza and Other Respiratory Viruses. 2009 Nov;3(6):265-6.

142.   Chidlow G, Harnett G, Williams S, Levy A, Speers D, Smith DW. Duplex Real-Time Reverse Transcriptase PCR Assays for Rapid Detection and Identification of Pandemic (H1N1) 2009 and Seasonal Influenza A/H1, A/H3, and B Viruses. J Clin Microbiol. 2010 March 1, 2010;48(3):862-6.

143.   Hornsleth A, Jankowski M. Sensitive enzyme immunoassay for the rapid diagnosis of influenza A virus infections in clinical specimens. Res Virol. 1990 May-Jun;141(3):373-84.

144.   Rebelo-de-Andrade H, Zambon MC. Different diagnostic methods for detection of influenza epidemics. Epidemiol Infect. 2000 Jun;124(3):515-22.

145.   Stone B, Burrows J, Schepetiuk S, Higgins G, Hampson A, Shaw R*, et al.* Rapid detection and simultaneous subtype differentiation of influenza A viruses by real time PCR. Journal of Virological Methods. 2004 May;117(2):103-12.

146.   Wallace LA, McAulay KA, Douglas JD, Elder AG, Stott DJ, Carman WF. Influenza diagnosis: from dark isolation into the molecular light. West of Scotland Respiratory Virus Study Group. J Infect. 1999 Nov;39(3):221-6.

Chapter 8 Appendices

Appendix I Initial search terms

Initial search terms and concepts:

Swine flu

Swine influenza

H1N1

h1n1

Influenza A

Influenza pandemic

Diagnostic

Diagnostic test

Diagnostic test accuracy

Sensitivity

Specificity

Diagnosis

Viral culture

PCR

Rapid diagnostic test

Immunofluorescence

Accuracy

Validity

Positive predictive value

Negative predictive value

Likelihood ratio

Appendix II   Electronic databases

Electronic databases searched:

PubMed

Embase

CINAHL

DARE

MedNar

Global health

CDC

WHO

Scopus

ProQuest dissertations and theses

NTIS

Health technology assessment database

Turning research into practice

MEDION

Web of science

Appendix III   Critical appraisal tool – the QUADAS checklist [57]

1.   Was the spectrum of patients representative of the patients who will receive the test in practice?

2.   Were selection criteria clearly described?

3.   Is the reference standard likely to correctly classify the target condition?

4.   Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?

5.   Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?

6.   Did patients receive the same reference standard regardless of the index test result?

7.   Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?

8.   Was the execution of the index test described in sufficient detail to permit replication of the test?

9.   Was the execution of the reference standard described in sufficient detail to permit its replication?

10.   Were the index test results interpreted without knowledge of the results of the reference standard?

11.   Were the reference standard results interpreted without knowledge of the results of the index test?

12.   Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?

13.   Were uninterpretable/ intermediate test results reported?

14.   Were withdrawals from the study explained?

Appendix IV   Example of search terms

Search terms tailored for the PubMed database.

1) Exp diagnostic test
2) "sensitivity AND specificity"
3) Exp Influenza A
4) Exp H1N1 AND Swine flu*
5) 1 AND 2
6) 3 AND 4
7) 5 AND 6

Filters: Human only subjects

                English language

Appendix V   Data extraction tool – the STARD checklist [32]

NOTE:
This appendix is included on pages
93-94 of the print copy of the thesis
held in the University of Adelaide
Library.

Appendix VI Table of included studies

Study	Population	Reference test	Index test	Results	Author conclusions and Comments
Beck ET, Jurgens LA, Kehl SC, Bose ME, Patitucci T, LaGue E, Darga P, Wilkinson K, Witt LM, Fan J, He J, Kumar S, Henrickson KJ. Journal of Molecular Diagnostics 201; 12 (1) 74-81	100 nasopharyngeal swab samples from the Children's Hospital of Wisconsin. Swabs were transported in M4 transport medium and stored at minus 80oC until testing No further patient details.	Details of the reference test (viral culture) - such as culture conditions, cell types, length of culture - not provided. Details of what constituted a positive result (i.e. a cut-off point) for viral culture were not provided. Influenza A subtyping was performed by the Midwest Respiratory Virus Group (Wisconsin)	Jaguar PCR assay (a fully automated PCR assay, Handylab Inc.) Patient samples were considered H1N1 positive if the MS2 internal control was positive and the FAM count was <40, with an appropriate amplification curve and melt temperature was within expected values (63oC +/-2oC). Samples were considered negative if they did not meet these criteria.	Total samples= 100 Samples positive for H1N1 by viral culture = 16 Samples negative for H1N1 by viral culture =84 Samples positive for H1N1 by PCR = 15 Samples negative for H1N1 by PCR = 78 Indeterminate = 7 These results were excluded from the sensitivity and specificity calculations Sensitivity of Jaguar assay 100 (CI 74-100). Specificity of Jaguar assay 100 (CI 96-100) %CI not detailed PPV = 100, NPV = 99 – however no details on prevalence.	The aim of the study was to establish validity (sensitivity and specificity) of new PCR assays compared with viral culture and an FDA approved assay prior to use as high throughput diagnostic tools. The study concludes that both sensitivity and specificity are high, for the Jaguar assay using frozen nasopharyngeal swab specimens. However, no details were given on viral culture methods, limiting comparisons that can be made.

Bose ME, Beck ET, Ledeboer N, Kehl SC, Jurgens LA, Patitucci T, Witt L, LaGue E, Darga P, He J, Fan, J, Kumar W and Henrickson KJ Journal of Clinical Microbiology 2009 47 (9) 2779-2786	315 nasopharyngeal swabs were collected at Albuquerque. Transported in M5 VTM. No further patient details.	R-Mix cells were used to grow virus from patient samples. Samples containing influenza A virus, were then typed and subtyped using type specific serum, however this technique is not described. Influenza A subtyping was performed by the Midwest Respiratory Virus Group (Wisconsin) Details of what constituted a positive result (i.e. a cut-off point) for viral culture were not provided.	PCR Assay NucliSENS easyMAG system was used for extraction and amplification using a Raider thermocycler. H1N1 reaction products were made in real time through fluorophore (AP-593) detection on a spectrophotometer. PCR samples were considered H1N1 positive if the MS2 internal control was positive and the threshold cycle was <40, with an appropriate amplification curve and melt temperature was within expected values ($60^{o}C$ +/-$2^{o}C$). Samples were considered negative if they did not meet these criteria.	Total samples= 315 Samples positive for H1N1 by viral culture = 50 Samples negative for H1N1 by viral culture =265 Samples positive for H1N1 by PCR = 50 Samples negative for H1N1 by PCR = 264 Indeterminate =1 sample could not be sub typed or sequenced and may represent a virus with a low titre or an incomplete genome. Sensitivity 95 (CI 87-99) Specificity 95 (CI 92-98) CI = 95% PPV = 83, NPV = 99 – however no details on prevalence.	The aim of the study was to establish validity of new PCR assays compared with viral culture to use as high throughput diagnostic tools. The study concludes that both sensitivity and specificity are high, for the Seasonal assay using frozen nasopharyngeal swab specimens. However, no details were given on viral culture methods, limiting comparisons that can be made. Discrepancies were present in the data: The Flu A or B type was determined for 315 samples. However a total of 65 Flu A - only 50 H1N1 +ve, 10 H3 +ve, 5 samples not accounted for

Ginocchio CC, Zhang F, Manji R, Arora S, Bornfreund M, Falk L, Lotlikar M, Kowerska M, Becker G, Korologos D, deGeronimo M and Crawford JM. Journal of Clinical Virology 2009; 45 191-195	Nasopharyngeal swab samples collected in universal viral transport medium and frozen at minus 70ºC until being tested. Patients presented with influenza-like symptoms to a hospital (either as in or out patients) in New York city (USA). Patients ranged in age from 4 days to 98 years. A total of 288 samples were processed for all 4 tests.	Routine viral culture methods were used; however these methods were not described. R-Mix cells (diagnostic hybrids (DHI) Athens, OH, USA) were used. Techniques used to subtype Influenza A viruses were not described. Details of what constituted a positive result (i.e. a cut-off point) for viral culture were not provided.	The diagnostic tests were conducted according to manufacturers' instructions. BinaxNOW Influenza A+B test (Inverness, Waltham, MA, USA) 3MA - 3M Rapid detection Flu A +B test (3M Medical diagnostics, St Paul, MN, USA) DFA - direct immuno fluorescence (D3 respiratory virus agents, diagnostic hybrids (DHI) Athens, OH, USA) RVP – Luminex xTAG respiratory panel virus (Luminex molecular diagnostics, Toronto, Canada) Rapid antigen testing – a combination of BinaxNOW 3M Rapid detection Flu A +B test	Total samples= 288 Samples positive for H1N1 by viral culture = 132 Samples negative for H1N1 by viral culture =156 **Binax,** Sensitivity= 9.6 Specificity= 100 PPV =100 NPV = 88.9 **3MA** Sensitivity = 40.0 Specificity = 97.8 PPV = 73.9 NPV = 96.3 **DFA** Sensitivity = 46.7 specificity = 94.5 PPV = 91.3 NPV = 58.9 **RVP** Sensitivity = 97.8 Specificity = 100 PPV = 100 NPV = 97.3 **Rapid antigen testing (Binax + 3MA)** Sensitivity = 17.8 Specificity = 93.6	The authors conclude that rapid diagnostic tests demonstrated poor sensitivities in this study. DFA performed better, with PCR being the most sensitive of the index tests, compared with viral culture. The PCR assay also had the added benefit that typing and subtyping of a virus can occur at the same stage, saving time. Binax and DFA tests were significantly less sensitive for novel H1N1 than viral culture (p<0.0001) or PCR. 3MA+B was significantly less sensitive than viral culture (p=0.0006), with the suggestion this test gave 8 false positive readings.

				PPV = 77.4 NPV = 47.9 Indeterminate =none reported No confidence intervals given Prevalence is reported as being 40.8% for H1N1; however it is unclear whether this relates to the hospital or the population at large.	
Mokhtari-Azad T, Rezaie-Khollari F, Nadji AR, Salimi V,Noroozbabaie Z, Naseri M, Miraie AR, Saadatmand Z, Gouya MM, Esteghamati AR, Hamkar R. Iranian Journal of Public Health 2007 36(3) 1-7	Samples were collected from 689 patients, presenting with influenza-like symptoms. 527 were throat washes and 161 were swabs. No patient demographic details.	Virus from patient sample was isolated using MDCK cells, supplemented with 10% calf serum. The cells were grown in tubes as monolayers in serum-free conditions for 3-7 days at 34°C. Details of what constituted a positive result (i.e. a cut off point) for viral culture were not provided.	PCR Assay Unique primers were designed for each Flu virus of interest – based on the haemagglutinin gene (H1, H3 and H5). No cross reaction was shown between the primers. Negative controls also showed no cross reaction with other common respiratory viruses. The presence of an amplified PCR product of predicted size (944bp for H1) on an ethidium bromide stained 1.5% agarose gel constituted a positive PCR result. The PCR product was not quantified further.	Total samples= n= 698 samples (527 washes and 161 swabs). Viral culture: 17 were H1 positive, 678 negative and 3 were H3 positive. PCR: 26 were H1 positive, 648 negative and 24 H3 positive. No details on whether the H1 positive samples were washes, swabs or a combination. Indeterminate = 52 samples were positive by PCR only, no exploration of this by authors.	The aim of this study was to evaluate an in house multiplex RT PCR assay for both swab and wash samples. The multiplex PCR assay allowed multiple influenza A virus subtypes to be detected at the same time, in the same reaction mix - saving both time and resources. Although viral culture was used as the reference standard, PCR gave 9 more positive results, suggesting either superior identification or false positive results. Unable to make conclusions without information about viral culture cut off criteria.

Wu C, Cheng X, He J, Lv X, Wang J, Deng R, Long Q, Wang X. Journal of Virological Methods 2008 148; 81-88	189 throat swabs from patients presenting with respiratory illness, provided by the CDC, China. No patient details.	Conventional culture methods were used. Not detailed but a reference to Schweiger et al 2000 was given for methods Virus from patient sample was grown in either MDCK cells or embryonated chicken eggs. The cells were grown in tubes as monolayers in serum-free conditions at 33°C. A haemagglutination assay was used for virus subtyping. Details of what constituted a positive result (i.e. a cut- off point) for viral culture were not provided.	PCR Assay Unique primers were designed for each Flu virus of interest based on the neuraminidase gene. A multiplex real time RT-PCR. RNA extraction was done manually using a kit and the amplification was done using an automated real time system for 40 cycles. Each probe used a different chromophore, allowing the products for each virus to be examined separately as the reaction occurred. A H1N1 positive sample gave a dual labelled (FAM and HEX) fluorescent signal after 40 cycles of amplification. The study also examined the effect of routinely used collection media on PCR product amplification (PBS, MEM and Hanks BSS) on stock viruses.	Total samples= 189 Samples positive for H1N1 by viral culture = 5 Samples negative for H1N1 by viral culture =184 Samples positive for H1N1 by PCR = 11 Samples negative for H1N1 by PCR = 178 Indeterminate = none reported There was no significant difference between the number of PCR amplification cycles needed to obtain a PCR product with MEM (21.15 +/-0.14) or HBSS (21.03 +/-0.17). However, PBS (20.75+/- 0.17) was significantly ($p<0.03$) lower than the other 2 media.	The aim of this study was to evaluate an in house multiplex PCR assay, with the goal of developing a quicker/cheaper, test with a higher throughput than the viral culture. The probes did not appear to cross talk with other viruses or Flu A subtypes. The authors conclude that as the PCR assay identifies more H1N1 positive samples that viral culture, it is therefore more accurate – however there was no exploration of potential sources of false positive results or reasons for increased accuracy. A secondary conclusion was that the transport media can affect amplification of PCR products in this assay, at least for throat swabs.

Appendix VI   List of excluded studies

Studies excluded because the focus of the study was analytical sensitivity and not diagnostic accuracy (n=11): [97-107]

Bouscambert-Duchamp M, Billaud G, et al. Evaluation of sensitivity of commercial kits for the diagnosis of A (H1N1) v infection in comparison with RT-PCRs developed by national influenza centers. Journal of Clinical Virology. 2009 Sep;46:S4-S.

Chan KH, Lam SY, et al. Comparative analytical sensitivities of six rapid influenza A antigen detection test kits for detection of influenza A subtypes H1N1, H3N2 and H5N1. J Clin Virol. 2007 Feb;38(2):169-71.

Chan KH, Lai ST, et al. Analytical sensitivity of rapid influenza antigen detection tests for swine-origin influenza virus (H1N1). Journal of Clinical Virology. 2009 Jul;45(3):205-7.

Daum LT, Canas LC, et al. A rapid, single-step multiplex reverse transcription-PCR assay for the detection of human H1N1, H3N2, and B influenza viruses. J Clin Virol. 2002 Dec;25(3):345-50.

Drexler JF, Helmer A, et al. Poor Clinical Sensitivity of Rapid Antigen Test for Influenza A Pandemic (H1N1) 2009 Virus. Emerging Infectious Diseases. 2009 Oct;15(10):1662-4.

Hall RJ, Peacey M, et al. A rapid method to support the diagnosis of swine-origin influenza virus (S-OIV), by sequencing of real-time PCR amplicon from diagnostic assays. J Clin Microbiol. 2009 Jul 8.

Kessler N, Ferraris O, et al. Use of the DNA flow-thru chip, a three-dimensional biochip, for typing and subtyping of influenza viruses. Journal of Clinical Microbiology. 2004 May;42(5):2173-85.

Kubo T, Agoh M, et al. Development of a Reverse Transcription-Loop-Mediated Isothermal Amplification Assay for Detection of Pandemic (H1N1) 2009 Virus as a Novel Molecular Method for Diagnosis of Pandemic Influenza in Resource-Limited Settings. J Clin Microbiol. 2010 March 1, 2010;48(3):728-35.

Schweiger B, Zadow R, et al. a fluorogenic PCR assay for typing and subtyping of influenza viruses in respiratory. Journal of Clinical Microbiology. 2000 01/01;38(4):1552-8.

Tucker SP, Cox C, et al. A flu optical immunoassay (ThermoBioStar's FLU OIA): a diagnostic tool for improved influenza management. Philos Trans R Soc Lond B Biol Sci. 2001 Dec 29;356(1416):1915-

Whiley DM, Bialasiewicz S, et al. Detection of novel influenza A(H1N1) virus by real-time RT-PCR. Journal of Clinical Virology. 2009 Jul;45(3):203-4.

Studies excluded because the focus of the study was Influenza A and not specifically H1N1 (n=22): [91, 108-128]

Agoritsas K, Mack K, et al. Evaluation of the Quidel QuickVue test for detection of influenza A and B viruses in the pediatric emergency medicine setting by use of three specimen collection methods. J Clin Microbiol. 2006 Jul;44(7):2638-41.

Alexander R, Hurt AC, et al. A comparison of a rapid test for influenza with laboratory-based diagnosis in a paediatric population. Commun Dis Intell. 2005;29(3):272-6.

Atmar RL, Baxter BD, et al. Comparison of reverse transcription-PCR with tissue culture and other rapid diagnostic assays for detection of type A influenza virus. J Clin Microbiol. 1996 Oct;34(10):2604-6.

Bolotin S, De Lima C, et al. Validation of the TaqMan Influenza A Detection Kit and a Rapid Automated Total Nucleic Acid Extraction Method to Detect Influenza A Virus in Nasopharyngeal Specimens. Annals of Clinical and Laboratory Science. 2009 Spr;39(2):155-9.

Boon ACM, French AMF, et al. Detection of influenza A subtypes in community-based surveillance. Journal of Medical Virology. 2001 Sep;65(1):163-70.

Booth S, Baleriola C, et al. Comparison of two rapid influenza A/B test kits with reference methods showing high specificity and sensitivity for influenza A infection. J Med Virol. 2006 May;78(5):619-22.

Carraro E, Neto DF, et al. Applications of a duplex reverse transcription polymerase chain reaction and direct immunofluorescence assay in comparison with virus isolation for detection of influenza A and B. Diagn Microbiol Infect Dis. 2007 Jan;57(1):53-7.

Cazacu AC, Chung SE, et al. Comparison of the Directigen flu A+B membrane enzyme immunoassay with viral culture for rapid detection of influenza A and B viruses in respiratory specimens. J Clin Microbiol. 2004 Aug;42(8):3707-10.

Dunn JJ, Gordon C, et al. Comparison of the Denka-Seiken INFLU A.B-Quick and BD Directigen Flu A+B kits with direct fluorescent-antibody staining and shell vial culture methods for rapid detection of influenza viruses. J Clin Microbiol. 2003 May;41(5):2180-3.

Ellis JS, Fleming DM, et al. Multiplex reverse transcription-PCR for surveillance of influenza A and B viruses in England and Wales in 1995 and 1996. J Clin Microbiol. 1997 Aug;35(8):2076-82.

Ginocchio C, Lotlikar M, et al. Clinical performance of the 3M rapid detection Flu A +B test compared to R-Mix culture, DFA and BinaxNOW Influenza A&B test. Journal of Clinical Virology. 2009;45:146-9.

Herrmann B, Larsson C, et al. Simultaneous detection and typing of influenza viruses A and B by a nested reverse transcription-PCR: comparison to virus isolation and antigen detection by immunofluorescence and optical immunoassay (FLU OIA). J Clin Microbiol. 2001 Jan;39(1):134-8.

Kehl SC, Henrickson KJ, et al. Evaluation of the Hexaplex assay for detection of respiratory viruses in children. J Clin Microbiol. 2001 May;39(5):1696-701.

Mehlmann M, Bonner AB, et al. Comparison of the MChip to viral culture, reverse transcription-PCR, and the QuickVue influenza A plus B test for rapid diagnosis of influenza. Journal of Clinical Microbiology. 2007 Apr;45(4):1234-7.

Pachucki CT. Rapid Tests for Influenza. Curr Infect Dis Rep. 2005 May;7(3):187-92.

Quach C, Newby D, et al. QuickVue influenza test for rapid detection of influenza A and B viruses in a pediatric population. Clin Diagn Lab Immunol. 2002 Jul;9(4):925-6.

Rahman M, Vandermause MF, et al. Performance of Binax NOW Flu A and B and direct fluorescent assay in comparison with a composite of viral culture or reverse transcription polymerase chain reaction for detection of influenza infection during the 2006 to 2007 season. Diagn Microbiol Infect Dis. 2008 Oct;62(2):162-6.

Ruest A, Michaud S, et al. Comparison of the Directigen flu A+B test, the QuickVue influenza test, and clinical case definition to viral culture and reverse transcription-PCR for rapid diagnosis of influenza virus infection. J Clin Microbiol. 2003 Aug;41(8):3487-93.

Simmerman JM, Chittaganpitch M, et al. Field performance and new uses of rapid influenza testing in Thailand. Int J Infect Dis. 2007 Mar;11(2):166-71.

Smith AB, Mock V, et al. Rapid detection of influenza A and B viruses in clinical specimens by Light Cycler real time RT-PCR. J Clin Virol. 2003 Sep;28(1):51-8.

Uyeki TM, Prasad R, et al. Low sensitivity of rapid diagnostic test for influenza. Clinical Infectious Diseases. 2009;48:e89-e92.

Vabret A, Sapin G, et al. Comparison of three non-nested RT-PCR for the detection of influenza A viruses. Journal of Clinical Virology. 2000 Sep;17(3):167-75.

Studies excluded because viral culture was not used as the reference test (n=12): [129-140]
Carr M, Rory G, et al. Development of a real-time RT-PCR for the detection of swine-lineage influenza A (H1N1) virus infections. Journal of clinical virology: the official publication of the Pan American Society for Clinical Virology. 2009 06/23.

Choi Y, Kim H, et al. Evaluation of new rapid antigen test for detection of pandemic influenza A/H1N1 2009 virus. Journal of Clinical Microbiology. 2010;48(6):2260-2.

Ciblak M, Kanturvardar M, et al. Sensitivity of rapid influenza antigen tests in the diagnosis of pandemic (H1N1) 2009 compared with the standard of rRT-PCR technique during the 2009 pandemic in Turkey. Scandinavian Journal of Infectious Diseases. 2010:e1-e4.

de la Tabla V, Antequera P, et al. Clinical evaluation of rapid point of care testing for detecting of novel influenza A (H1N1) virus in a population based study in Spain. Clinical Microbiology and Infection. 2009.

Faix DJ, Sherman SS, et al. Rapid-Test Sensitivity for Novel Swine-Origin Influenza A (H1N1) Virus in Humans. New England Journal of Medicine. 2009 Aug;361(7):728-9.

He J, Bose ME, et al. Rapid Multiplex RT-PCR typing of influenza A and B, and subtyping of influenza A into H1, 2, 3, 5, 7, 9, N1 (human), N1 (animal), N2 and N7 including typing of novel swine-origin Influenza A (H1N1) Virus during current 2009 outbreak in Milwaukee, Wisconsin. J Clin Microbiol. 2009 Jul 29.

Herzum I, Lutz T, et al. Diagnostic performance of rapid influenza antigen assays in patients infected with the new influenza A (H1N1) virus. Clinical Chemistry and Laboratory Medicine. 2010 Jan;48(1):53-6.

Jenny S, Hu Y, et al. Evaluation of the Xpert Flu A panel nucleic acid amplification based point of care test for influenza A virus detection and pandemic H1 subtyping. Journal of Clinical Virology. 2010;49:85-9.

Likitnukul S, Boonsiri K, et al. Evaluation of Sensitivity and Specificity of Rapid Influenza Diagnostic Tests for Novel Swine-Origin Influenza A (H1N1) Virus. Pediatric Infectious Disease Journal. 2009 Nov;28(11):1038-9.

Louie J, Guevara H, et al. Rapid antigen test for diagnosis of pandemic (H1N1) 2009. Emerging Infectious Diseases. 2010;16(5):824-6.

Troppan K, Bozic M, et al. Evaluation of four molecular assays for detection of pandemic influenza A (H1N1) 2009 virus in the routine diagnostic laboratory. Journal of Clinical Virology. 2010;49:82-4.

Vasoo S, Stevens J, et al. Rapid Antigen Tests for Diagnosis of Pandemic (Swine) Influenza A/H1N1. Clinical Infectious Diseases. 2009 Oct;49(7):1090-3.

Studies excluded because they contained insufficient data (n=7): [105, 141-146]

Arya SC, Agarwal N. Apropos 'performance of influenza point-of-care tests in the detection of pandemic (H1N1) 2009 influenza viruses'. Influenza and Other Respiratory Viruses. 2009 Nov;3(6):265-6.

Chidlow G, Harnett G, et al. Duplex Real-Time Reverse Transcriptase PCR Assays for Rapid Detection and Identification of Pandemic (H1N1) 2009 and Seasonal Influenza A/H1, A/H3, and B Viruses. J Clin Microbiol. 2010 March 1, 2010;48(3):862-6.

Hornsleth A, Jankowski M. Sensitive enzyme immunoassay for the rapid diagnosis of influenza A virus infections in clinical specimens. Res Virol. 1990 May-Jun;141(3):373-84.

Rebelo-de-Andrade H, Zambon MC. Different diagnostic methods for detection of influenza epidemics. Epidemiol Infect. 2000 Jun;124(3):515-22.

Stone B, Burrows J, et al. Rapid detection and simultaneous subtype differentiation of influenza A viruses by real time PCR. Journal of Virological Methods. 2004 May;117(2):103-12.

Wallace LA, McAulay KA, et al. Influenza diagnosis: from dark isolation into the molecular light. West of Scotland Respiratory Virus Study Group. J Infect. 1999 Nov;39(3):221-6.