Cochrane Library

Cochrane Database of Systematic Reviews

Rapid diagnostic tests for plague (Review)

Jullien S, Dissanayake HA, Chaplin M

Jullien S, Dissanayake HA, Chaplin M.
Rapid diagnostic tests for plague.
Cochrane Database of Systematic Reviews 2020, Issue 6. Art. No.: CD013459.
DOI: 10.1002/14651858.CD013459.pub2.

www.cochranelibrary.com

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Rapid diagnostic tests for plague

Sophie Jullien[1], Harsha A Dissanayake[2], Marty Chaplin[3]

[1]Barcelona Institute for Global Health, University of Barcelona, Barcelona, Spain. [2]University Medical Unit, National Hospital of Sri Lanka, Colombo, Sri Lanka. [3]Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

Contact address: Sophie Jullien, sophjullien@gmail.com.

Editorial group: Cochrane Infectious Diseases Group.
Publication status and date: New, published in Issue 6, 2020.

Citation: Jullien S, Dissanayake HA, Chaplin M. Rapid diagnostic tests for plague. *Cochrane Database of Systematic Reviews* 2020, Issue 6. Art. No.: CD013459. DOI: 10.1002/14651858.CD013459.pub2.

ABSTRACT

Background

Plague is a severe disease associated with high mortality. Late diagnosis leads to advance stage of the disease with worse outcomes and higher risk of spread of the disease. A rapid diagnostic test (RDT) could help in establishing a prompt diagnosis of plague. This would improve patient care and help appropriate public health response.

Objectives

To determine the diagnostic accuracy of the RDT based on the antigen F1 (F1RDT) for detecting plague in people with suspected disease.

Search methods

We searched the CENTRAL, Embase, Science Citation Index, Google Scholar, the World Health Organization International Clinical Trials Registry Platform and ClinicalTrials.gov up to 15 May 2019, and PubMed (MEDLINE) up to 27 August 2019, regardless of language, publication status, or publication date. We handsearched the reference lists of relevant papers and contacted researchers working in the field.

Selection criteria

We included cross-sectional studies that assessed the accuracy of the F1RDT for diagnosing plague, where participants were tested with both the F1RDT and at least one reference standard. The reference standards were bacterial isolation by culture, polymerase chain reaction (PCR), and paired serology (this is a four-fold difference in F1 antibody titres between two samples from acute and convalescent phases).

Data collection and analysis

Two review authors independently selected studies and extracted data. We appraised the methodological quality of each selected studies and applicability by using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. When meta-analysis was appropriate, we used the bivariate model to obtain pooled estimates of sensitivity and specificity. We stratified all analyses by the reference standard used and presented disaggregated data for forms of plague. We assessed the certainty of the evidence using GRADE.

Main results

We included eight manuscripts reporting seven studies. Studies were conducted in three countries in Africa among adults and children with any form of plague. All studies except one assessed the F1RDT produced at the Institut Pasteur of Madagascar (F1RDT-IPM) and one study assessed a F1RDT produced by New Horizons (F1RDT-NH), utilized by the US Centers for Disease Control and Prevention. We could

not pool the findings from the F1RDT-NH in meta-analyses due to a lack of raw data and a threshold of the test for positivity different from the F1RDT-IPM.

Risk of bias was high for participant selection (retrospective studies, recruitment of participants not consecutive or random, unclear exclusion criteria), low or unclear for index test (blinding of F1RDT interpretation unknown), low for reference standards, and high or unclear for flow and timing (time of sample transportation was longer than seven days, which can lead to decreased viability of the pathogen and overgrowth of contaminating bacteria, with subsequent false-negative results and misclassification of the target condition).

F1RDT for diagnosing all forms of plague

F1RDT-IPM pooled sensitivity against culture was 100% (95% confidence interval (CI) 82 to 100; 4 studies, 1692 participants; very low certainty evidence) and pooled specificity was 70.3% (95% CI 65 to 75; 4 studies, 2004 participants; very low-certainty evidence).

The performance of F1RDT-IPM against PCR was calculated from a single study in participants with bubonic plague (see below).
There were limited data on the performance of F1RDT against paired serology.

F1RDT for diagnosing pneumonic plague

Performed in sputum, F1RDT-IPM pooled sensitivity against culture was 100% (95% CI 0 to 100; 2 studies, 56 participants; very low-certainty evidence) and pooled specificity was 71% (95% CI 59 to 80; 2 studies, 297 participants; very low-certainty evidence).

There were limited data on the performance of F1RDT against PCR or against paired serology for diagnosing pneumonic plague.

F1RDT for diagnosing bubonic plague

Performed in bubo aspirate, F1RDT-IPM pooled sensitivity against culture was 100% (95% CI not calculable; 2 studies, 1454 participants; low-certainty evidence) and pooled specificity was 67% (95% CI 65 to 70; 2 studies, 1198 participants; very low-certainty evidence).

Performed in bubo aspirate, F1RDT-IPM pooled sensitivity against PCR for the *caf1* gene was 95% (95% CI 89 to 99; 1 study, 88 participants; very low-certainty evidence) and pooled specificity was 93% (95% CI 84 to 98; 1 study, 61 participants; very low-certainty evidence).

There were no data providing data on both F1RDT and paired serology for diagnosing bubonic plague.

Authors' conclusions

Against culture, the F1RDT appeared highly sensitive for diagnosing either pneumonic or bubonic plague, and can help detect plague in remote areas to assure management and enable a public health response. False positive results mean culture or PCR confirmation may be needed. F1RDT does not replace culture, which provides additional information on resistance to antibiotics and bacterial strains.

PLAIN LANGUAGE SUMMARY

Rapid diagnostic tests for plague

Why is improving diagnosis of plague important?

Plague is a severe disease associated with high death rates. Pneumonic plague mainly affects the lungs, while bubonic plague present with painful swellings. Not recognizing plague early may result in delayed diagnosis and treatment associated with advanced illness and death, and increased disease spread. A rapid diagnostic test (RDT) could help prompt diagnosis of plague, especially in low-resource settings. This would improve patient care and help appropriate response to avoid the disease spread.

What is the aim of this review?

To assess the accuracy of the F1RDT for detecting plague in people with suspected plague.

What was studied in this review?

F1RDT is a test that detects the F1 antigen, which is part of the outer surface of *Yersinia pestis*, the bacteria causing plague. The test is simple to perform and provides a result within 15 minutes. It can be performed in the pus contained in the buboes (swellings), or in the sputum (mucous coughed up from the respiratory tract) of people with suspected pneumonic plague. We measured the results of F1RDT against culture, molecular test, or serological tests.

What are the main results?

Seven studies (reported in eight manuscripts) provided findings of F1RDT used in people with suspected plague in three African countries.

For any form of plague and when compared to culture, F1RDT registered positive in 100% (sensitivity, which measures a test's ability to correctly identify a positive result for the disease) of people who had plague and registered negative in 70% of people who actually did not

have plague (specificity, which measures a test's ability to correctly generate a negative result for people who do not have the condition that is being tested for).

For pneumonic plague, sensitivity was 100% and specificity 71% compared to culture.

For bubonic plague, sensitivity was 100% and specificity 67% compared to culture. Compared to a molecular test for bubonic plague, sensitivity was 95% and specificity 93%.

How confident are we in the review's results?

Overall, the reliability of the evidence was very low. Results should be interpreted with caution. There were concerns about the quality of the methodology of all included studies. Also, culture might not work well as a reference standard (comparator) when people received antibiotics before sample collection for testing.

What do the results mean?

In a hypothetical population of 1000 people:

• with symptoms of pneumonic plague where 40 of them have the disease confirmed by culture, the utilization of F1RDT would result in: 318 people to be F1RDT-positive, of which 278 would not have pneumonic plague (called false positives); and 682 people to be F1RDT-negative, of which none would have pneumonic plague (called false negatives).

• with symptoms of bubonic plague where 40 of them have the disease confirmed by culture, the utilization of F1RDT would result in: 357 people to be F1RDT-positive, of which 317 would not have bubonic plague (false positives); and 643 people to be F1RDT-negative, of which none would have bubonic plague (false negatives).

• with symptoms of bubonic plague where 40 of them have the disease confirmed by molecular test, the utilization of F1RDT would result in: 105 people to be F1RDT-positive, of which 67 would not have bubonic plague (false positives); and 895 people to be F1RDT-negative, of which two would have bubonic plague (false negatives).

Who do the review's results apply to?

Adults and children with suspected bubonic or pneumonic plague.

What are the implications of this review?

F1RDT appears to be highly sensitive for pneumonic or bubonic plague. As a simple test that can be performed at a patient's bedside in remote and low-resource areas, F1RDT can assist with plague diagnosis for early management, and appropriate preventive measures to avoid spread of the disease.

The number of false positives (people with a positive F1RDT but who do not have plague) indicate that F1RDT may need to be combined with other laboratory evaluations (culture or molecular test) to confirm the diagnosis of plague.

F1RDT does not replace culture, which provides additional information on resistance to antibiotics and bacterial strains.

How up-to-date is this review?

The review authors searched for studies up to 15 May 2019.

SUMMARY OF FINDINGS

Summary of findings 1. F1RDT against culture for all forms of plague

Review question: what is the diagnostic accuracy of F1RDT for the diagnosis of plague in adults and children?

Setting: populations where plague has been known to occur and populations where an outbreak is in progress

Patient or population: adults and children with suspicion of plague

Index test: F1RDT from Institut Pasteur of Madagascar

Reference standard: culture

Study design: retrospective reports

Pooled sensitivity: 1.00 (95% CI 0.82 to 1.00). **Pooled specificity:** 0.70 (95% CI 0.65 to 0.75)

Test result	Number of results per 1000 participants tested (95% CI)[a]			Number of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0.1%	Prevalence 4%	Prevalence 20%		
True positives (participants correctly classified as having plague)	1 (1 to 1)	40 (33 to 40)	200 (164 to 200)	1692 (4)	⊕⊖⊖⊖ **Very low** [b,c,d]
False negatives (participants incorrectly classified as not having plague)	0 (0 to 0)	0 (0 to 7)	0 (0 to 36)		
True negatives (participants correctly classified as not having plague)	699 (649 to 749)	672 (624 to 720)	560 (520 to 600)	2004 (4)	⊕⊖⊖⊖ **Very low** [b,d,e]
False positives (participants incorrectly classified as having plague)	300 (250 to 350)	288 (240 to 336)	240 (200 to 280)		

[a]The table displays normalized frequencies within a hypothetical cohort of 1000 people at three different plague prevalences (pretest probabilities): 0.1%, 4%, and 20%. We selected prevalence values based on the range of prevalence observed across the included studies and on hypothetical field situations. We estimated confidence intervals based on those around the point estimates for pooled sensitivity and specificity.

Abbreviations

CI: confidence interval; F1RDT: F1 antigen rapid diagnostic test.

Explanations

[b]Downgraded two levels for risk of bias: there is high risk of bias on the patient selection and flow and timing domains for all included studies (except Rajerison 2020, with low risk of bias on flow and timing domain). The included studies were retrospective case series (except Chanteau 2003a, which was prospective with unclear sampling methods), which was at high risk of introducing bias for evaluating diagnostic test accuracy.

[c]Not downgraded for inconsistency: there was an outlier corresponding to the pneumonic outbreak from 2017 to 2018 (Rajerison 2020), which includes four participants. The remainder of findings showed consistency of sensitivity estimate towards 100%.

[d]Downgraded two levels for imprecision: wide 95% CI around false negatives and false positives may lead to different decisions depending on whether the upper or lower limit is assumed.
[e]Downgraded one level for indirectness: there was high concern of applicability due to exclusion of people who received antibiotics prior to sample collection (Rajerison 2020). This can lead to false-negative culture, which is the reference standard. In addition, the index test was performed in a central laboratory, which may not reflect the field conditions in case of an outbreak.
GRADE certainty of evidence (GRADEpro GDT 2015)
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Summary of findings 2. F1RDT against culture for pneumonic plague

Review question: what is the diagnostic accuracy of F1RDT for the diagnosis of pneumonic plague in adults and children?

Setting: populations where plague has been known to occur and populations where an outbreak is in progress

Patient or population: adults and children with suspicion of pneumonic plague

Index test: F1RDT from Institut Pasteur of Madagascar performed in sputum

Reference standard: sputum culture

Study design: retrospective reports

Pooled sensitivity: 1.00 (95% CI 0.00 to 1.00). **Pooled specificity:** 0.71 (95% CI 0.59 to 0.80)

Test result	Number of results per 1000 participants tested (95% CI)[a]			Number of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0.1%	Prevalence 4%	Prevalence 20%		
True positives (participants correctly classified as having pneumonic plague)	1 (0 to 1)	40 (0 to 40)	200 (0 to 200)	56 (2)	⊕⊖⊖⊖ **Very low** [b,c,d]
False negatives (participants incorrectly classified as not having pneumonic plague)	0 (0 to 1)	0 (0 to 40)	0 (0 to 200)		
True negatives (participants correctly classified as not having pneumonic plague)	709 (589 to 799)	682 (566 to 768)	568 (472 to 640)	297 (2)	⊕⊖⊖⊖ **Very low** [b,d,e]
False positives (participants incorrectly classified as having pneumonic plague)	290 (200 to 410)	278 (192 to 394)	232 (160 to 328)		

[a]The table displays normalized frequencies within a hypothetical cohort of 1000 people at three different pneumonic plague prevalences (pretest probabilities): 0.1%, 4%, and 20%. We selected prevalence values based on the range of prevalence observed across the included studies and on hypothetical field situations. We estimated confidence intervals based on those around the point estimates for pooled sensitivity and specificity.

Abbreviations

CI: confidence interval; F1RDT: F1 antigen rapid diagnostic test.

Explanations

[b]Downgraded two levels for risk of bias: there was high or unclear risk of bias on the index test and flow and timing domains for both included studies. The design of both studies was retrospective case series, which is at high risk of introducing bias for evaluating diagnostic test accuracy.

[c]Not downgraded for inconsistency: there was an outlier corresponding to the pneumonic outbreak from 2017 to 2018 (Rajerison 2020), which included four participants. The remainder of the findings showed consistency of sensitivity estimate towards 100%.

[d]Downgraded two levels for imprecision: wide 95% CI around false negatives and false positives may lead to different decisions depending on whether the upper or lower limit is assumed. In addition, small number of participants included.

[e]Downgraded one level for indirectness: there was high concern of applicability due to exclusion of people who received antibiotics prior to sample collection. This can lead to false-negative culture, which is the reference standard. In addition, the index test was performed in a central laboratory, which may not reflect the field conditions in case of an outbreak.

GRADE certainty of evidence (GRADEpro GDT 2015)

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Summary of findings 3. F1RDT against culture for bubonic plague

Review question: what is the diagnostic accuracy of F1RDT for the diagnosis of bubonic plague in adults and children?

Setting: populations where plague has been known to occur and populations where an outbreak is in progress

Patient or population: adults and children with suspicion of bubonic plague

Index test: F1RDT from Institut Pasteur of Madagascar performed in bubo aspirate

Reference standard: culture from bubo aspirate

Study design: retrospective reports

Pooled sensitivity: 1.00 (95% CI not calculable). **Pooled specificity:** 0.67 (95% CI 0.65 to 0.70)

Test result	Number of results per 1000 participants tested (95% CI)[a]			Number of participants (studies)	Certainty of the evidence (GRADE)
	Prevalence 0.1%	Prevalence 4%	Prevalence 50%		
True positives (participants correctly classified as having bubonic plague)	1	40	500	1454 (2)	⊕⊕⊖⊖ **Low** [b]

False negatives (participants incorrectly classified as not having bubonic plague)	0	0	0		
True negatives (participants correctly classified as not having bubonic plague)	669 (649 to 699)	643 (624 to 672)	335 (325 to 350)	1198 (2)	⊕⊖⊖⊖ **Very low** [b,c]
False positives (participants incorrectly classified as having bubonic plague)	330 (300 to 350)	317 (288 to 336)	165 (150 to 175)		

[a]The table displays normalized frequencies within a hypothetical cohort of 1000 people at three different bubonic plague prevalences (pretest probabilities): 0.1%, 4%, and 50%. We selected prevalence values based on the range of prevalence observed across the included studies and on hypothetical field situations. We estimated confidence intervals based on those around the point estimates for pooled sensitivity and specificity.

Abbreviations

CI: confidence interval; F1RDT: F1 antigen rapid diagnostic test.

Explanations

[b]Downgraded two levels for risk of bias: there was high risk of bias on the index test domain for both included studies. The design of both studies was retrospective, which is at high risk of introducing bias for evaluating diagnostic test accuracy.

[c]Downgraded one level for indirectness on false-positive tests only. One of the included studies was at high concern of applicability due to exclusion of people who received antibiotics prior to sample collection. This can lead to false-negative culture, which is the reference standard. In addition, the index test was performed in a central laboratory, which may not reflect the field conditions in case of an outbreak.

GRADE certainty of evidence (GRADEpro GDT 2015)

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Summary of findings 4. F1RDT against PCR for bubonic plague

Review question: what is the diagnostic accuracy of F1RDT for the diagnosis of bubonic plague in adults and children?

Setting: populations where plague has been known to occur and populations where an outbreak is in progress

Patient or population: adults and children with suspicion of bubonic plague

Index test: F1RDT from Institut Pasteur of Madagascar performed in bubo aspirate

Reference standard: PCR (*caf1*) from bubo aspirate

Study design: retrospective reports

Pooled sensitivity: 0.95 (95% CI 0.89 to 0.99). **Pooled specificity:** 0.93 (95% CI 0.84 to 0.98)

Test result	Number of results per 1000 participants tested (95% CI)[a]			Number of participants (studies)	Certainty of the evidence (GRADE)

	Prevalence 0.1%	Prevalence 4%	Prevalence 50%		
True positives (participants correctly classified as having bubonic plague)	1 (1 to 1)	38 (36 to 40)	475 (445 to 495)	88 (1)	⊕⊖⊖⊖ **Very low** [b,c]
False negatives (participants incorrectly classified as not having bubonic plague)	0 (0 to 0)	2 (0 to 4)	25 (5 to 55)		
True negatives (participants correctly classified as not having bubonic plague)	929 (839 to 979)	893 (806 to 941)	465 (420 to 490)	61 (1)	⊕⊖⊖⊖ **Very low** [b,c]
False positives (participants incorrectly classified as having bubonic plague)	70 (20 to 160)	67 (19 to 154)	35 (10 to 80)		

[a]The table displays normalized frequencies within a hypothetical cohort of 1000 people at three different bubonic plague prevalences (pretest probabilities): 0.1%, 4%, and 50%. We selected prevalence values based on the range of prevalence observed across the included studies and on hypothetical field situations. We estimated confidence intervals based on those around the point estimates for pooled sensitivity and specificity.

Abbreviations

CI: confidence interval; F1RDT: F1 antigen rapid diagnostic test; PCR: polymerase chain reaction.

Explanations

[b]Downgraded two levels for risk of bias: there was high risk of bias on the index test domain. The estimates were assessed by one retrospective study, which is a study design at high risk of introducing bias for evaluating diagnostic test accuracy.

[c]Downgraded two levels for imprecision: wide 95% CI around false positives (and false negatives for high pretest probability of the disease) may lead to different decisions depending on whether the upper or lower limit is assumed. In addition, small number of participants included.

GRADE certainty of evidence (GRADEpro GDT 2015)

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

BACKGROUND

Target condition being diagnosed

Plague has caused major historic pandemics including the 'Plague of Justinian' in the 6th century, the 'Black Death' in the 14th century (which resulted in the death of one-third of the European population), and the 'Third Pandemic' in the 19th century (Rasmussen 2015). This severe disease remains a current threat in many parts of the world, and has increased over the last three decades (WHO 2009). Between 1989 and 2003, 25 countries reported 38,310 human cases of plague, including 2845 deaths (WHO 2019a). Since 2000, over 95% of the burden associated with plague has been concentrated in Africa, particularly the Democratic Republic of the Congo (DRC), Madagascar, Uganda, and the United Republic of Tanzania (WHO 2016; WHO 2019a). Peru and the USA also regularly report cases. Finally, although Asia is the region with the biggest natural foci of the disease, the reservoir consists of gerbils and marmots; there is a limited at-risk population in contact with these animals, so outbreaks are sporadic (WHO 2016). As of 2017, the DRC, Madagascar, and Peru had the highest incidence of the disease. However, countries that have never experienced plague, or not experienced plague for a long while, can be affected as the limits of the existing foci are not fixed and new foci can emerge. Human plague outbreaks are continuously being reported, from Indonesia in 2007 and the DRC and Tanzania in 2014, to the outbreak of pneumonic plague more recently reported in Madagascar in 2017 (WHO 2009; WHO 2016; WHO 2019a).

Plague is caused by the bacteria *Yersinia pestis*. It is primarily a vector borne zoonotic disease, affecting rodents and other wild and domestic animals. It is most commonly transmitted to humans by rodent fleas, leading to bubonic plague. Less frequently, plague can be transmitted through scratches or bites from infected animals, direct handling of infected animals, and human-to-human transmission by inhalation of droplets from people with pneumonic plague (CDC 2019; Weniger 1984).

Plague can affect both adults and children, with no differences between genders or ethnicities. However, the disease presents more frequently among people involved in activities with an increased exposure to the disease, such as hunters, veterinarians, etc. Poverty is also associated with a greater risk of contracting plague due to increased exposure to rodents.

Plague is always a medical emergency and presents in a variety of forms, with three major clinical syndromes. Bubonic plague is the most common form and is characterized by enlarged lymph nodes with necrotic areas called buboes. Pneumonic plague is most often a fulminant form which affects the lungs and presents with cough and bloody sputum. The pneumonic form can be primary (as a result of inhalation of droplets from infected humans or animals), or secondary (as a result of the haematogenous spread of any other form of plague) (CIDRAP 2013). The third major clinical form is septicaemic plague, which occurs when the infection spreads to the circulatory system; it can be primary (without buboes or pulmonary affectation), or secondary (as a result of spreading bubonic or pneumonic plague). Less commonly, plague can present as meningitis (Prentice 2007).

Although efficient antimicrobials are available, plague still has a high mortality rate as most outbreaks take place in remote places in resource-limited settings, where proper diagnosis and treatment remains challenging (WHO 2009). While bubonic plague is associated with case fatality ratios (CFRs) of 10% to 20%, pneumonic plague is highly fatal, with a CFR close to 100% if left untreated and over 50% when adequately treated with antimicrobials (Prentice 2007).

In addition to the sporadic cases and outbreaks, because of the characteristics of the disease resulting in high mortality, *Y pestis* has been used as a biological weapon and is currently a bioterrorism threat (CDC 2019).

Due to its historical pandemics and high fatality rate, plague continues to cause fear and panic, and is sometimes associated with a disproportionate public health response, which has considerable social and economic consequences (Mavalankar 1995; Mead 2018). A clinical diagnosis of plague is difficult and not reliable. The symptoms of pneumonic plague are not specific and can be present in a person with pneumonia caused by many other pathogens. These include other bacteria such as streptococcus pneumoniae or tuberculosis, that would require different antibiotics than plague, but also viruses, such as influenza, which would require no antibiotics. A person with a swollen lymph node in an endemic area or in the context of an outbreak is more likely to receive an accurate diagnosis of bubonic plague than people with suspected pneumonic plague. However, other diagnoses need to be considered, mainly other infections that cause swollen lymph nodes such as pyogenic abscess, tularaemia, tuberculosis, lymphogranuloma venereum, and cat scratch fever. A point-of-care diagnostic tool that is quick to use and highly accurate would help ensure appropriate response, especially in the context of outbreaks.

Index test(s)

Rapid diagnostic tests (RDTs) detect pathogen-specific antigens in a small quantity of different body fluids through lateral flow immunochromatography. RDTs are widely used in other diseases, such as malaria (WHO 2019b). They are usually easy to use and interpret. Indeed, they can be performed at the bedside of the patient without the requirement of special equipment or laboratory facilities. They give a simple result within around 15 minutes – positive or negative, at thresholds set by the manufacturer – that can easily be interpreted by health workers without advanced training. Therefore, RDTs are useful diagnostic tools for use at the community level and in low-resource settings (WHO 2019b).

In the case of plague, the RDT detects the F1 capsular antigen of *Y pestis* (F1RDT), which is present in large amounts in buboes, blood, and sputum from patients infected with plague. F1RDT is the only RDT for plague that has been developed for clinical purposes that we are aware of. The test gives a semi-quantitative result within 15 minutes according to the intensity of the line (from 1+ to 4+), although it is most commonly used as a qualitative test (positive or negative result) where positivity is interpreted from 1+ (as soon as the line is visible). The threshold for positivity will depend on the manufacturer, and is established by the lowest concentration of the F1 antigen that the test can detect. The F1RDT can be used in bubo aspirate, urine, and sputum (Chanteau 2000a); it is not usually used in blood as the pink result line would be difficult to see. Currently, the F1RDT that is mainly used in the field is produced in Madagascar. The most recent version was developed in 2001 (Chanteau 2003a). Other F1RDTs for plague are produced by New

Horizons in the USA (New Horizons 2019) and in Taiwan (Hsu 2018), but they are not licensed for use in humans or available in the market.

Similar to other RDTs, F1RDT should be administered by trained staff, following the manufacturer's restrictions and warnings of the test. Although no advanced training is required, clear indications must be respected. These include sample collection, sample preparation, and timely and accurate reading of the result. Storage conditions are indicated by each manufacturer and are usually easy to comply with.

Clinical pathway

People of any age affected by plague will present with non-specific symptoms such as fever, chills, headache, or nausea; these are usually associated with lymph node swelling in case of bubonic plague, with or without cough, haemoptysis, and chest pain for pneumonic plague. While a first diagnosis of suspected plague is based on clinical findings, the definitive diagnosis requires laboratory testing. Bacteriological identification of *Y pestis* through microscopy or culture (or both) is the reference standard for a confirmed case of plague (Rajerison 2020). *Y pestis* grows easily in standard culture media, and, while bacteriological isolation has a high specificity, it is also highly sensitive under ideal conditions. However, administration of antibiotics prior to sample collection is likely to lead to a false-negative result. Delay in transportation from the time of collection to the central laboratory (which is not uncommon in low-resource settings), associated or not with poor storage conditions, can lead to decreased viability of the pathogen and overgrowth of contaminating bacteria, with subsequent false-negative results. Another confirmatory diagnosis of plague is obtained by serology with a four-fold difference in F1 antibody titres between paired serum samples, from acute and convalescent phases. While this test is highly specific and allows diagnosis of true cases of plague, a limitation in practice is the collection of the second serum sample during the convalescence phase, as people might return to their home or work setting as soon as they feel well enough. Another technique more recently used in the diagnosis of plague is polymerase chain reaction (PCR), targeting several genes including the *pla* gene, encoding plasminogen activator, and the *caf1* gene, encoding F1 capsule antigen. This technique only requires a small amount of sample, and detects specific genes of *Y pestis* whether the bacteria is alive or dead. A person with plague who has received antibiotics prior to sample collection might present a negative culture (false negative) but a positive PCR.

These diagnostic tests (culture, paired serology, and PCR) require technology and qualified staff, which are rarely available in resource-limited areas. In addition, results from culture and paired serology take several days. These tests do not allow a fast confirmation or exclusion of plague diagnosis, and physicians cannot rely on them for the acute management of patients. Therefore, in practice, as soon as a case of plague is clinically suspected in an area where plague is endemic (or if the case visited an endemic area), the patient is immediately given antibiotics following collection of biological samples (blood, sputum, bubo aspirate, or a combination of these) whenever possible, and managed as a case of plague until microbiological diagnosis is confirmed or excluded, usually based on culture results. It is common that patients finish the treatment course or evolve to a fatal outcome before plague is microbiologically confirmed. In addition, public health measures are established, especially for cases of pneumonic plague, including tracing contacts and distributing chemoprophylaxis.

The use of the F1RDT, performed for all suspected cases of plague during the first contact of the patient with healthcare facilities, would support the clinical suspicion of plague when positive, and guide physicians to consider other diseases when negative, providing valuable guidance for both clinical and public health response.

Prior test

It is very unlikely that patients will have had another diagnostic test for detecting plague prior to presentation to the health facility where the F1RDT would be performed, as the F1RDT should be performed at the first point of presentation with medical facilities.

Role of index test

The role of the F1RDT is to provide bedside rapid results in the identification of people with plague, to allow prompt treatment, and to establish preventive measures in order to limit transmission of the disease to others in case of a positive result for pneumonic plague. A negative F1RDT finding would prompt clinicians to consider other diagnoses for correct management of the patient and to avoid unnecessary preventive measures that would be essential in case of plague. An easy-to-use and accurate F1RDT for plague would, therefore, be of considerable help in daily clinical practice for the management of people with suspicion of plague in endemic areas by providing a fast diagnosis, as microbiological confirmation of plague takes several days. The F1RDT would be used in addition to the reference standard and would not replace culture, which is fundamental for assessing circulating strains and antibiotic resistance testing. According to the specificity of the test, the F1RDT could be used as a triage tool, meaning that a negative result would definitely exclude plague, and that other tests – such as culture – would only be collected in cases with a positive result, and managed accordingly.

The F1RDT cannot be used as a screening tool for plague in asymptomatic people, for example asymptomatic people who have been in contact with people with plague. Indeed, samples to perform the test are mainly bubo aspirate (in case of suspicion of bubonic plague) and sputum (in case of pneumonic plague) and those samples would be non-existent in asymptomatic people.

Alternative test(s)

Direct enzyme-linked immunosorbent assay (ELISA) for detection of the F1 antigen is another test used for the diagnosis of plague. It requires equipment as well as trained personnel to perform it, and it is not easily available in low-resource settings.

Rationale

Plague is a serious illness with high mortality and rapid transmission from fleas or in between humans if control measures are not immediately implemented. Given the non-specific symptoms of plague and the measures to be implemented in the event of a confirmed case (such as surveillance measures, identification of contacts for prophylaxis, and safe burial to avoid spread of the disease), it is crucial to make a formal diagnosis of plague as soon as possible to distinguish it from other infections with similar clinical presentation. This is particularly important for the pneumonic form of plague, where the high prevalence of other

infections presenting with similar respiratory symptoms such as cough and fever makes an accurate clinical diagnosis difficult. The delay in a confirmed diagnosis may have two main repercussions. The first is at the individual level, as confirmation or exclusion of the diagnosis will help optimize the patient's management, including consideration of alternative diseases and treatments. A confirmed diagnosis allows targeted treatment with specific antibiotics. If the diagnosis is not made, patients are likely to be treated with routine empiric antibiotics which are ineffective against plague. The second is at a public health level, as pneumonic plague can be transmitted from human to human, leading to outbreaks, which are often associated with fear, panic, and sometimes with excessive measures that can lead to social and economic disruption, with considerable consequences (Mavalankar 1995).

A highly accurate and fast diagnostic test would undoubtedly be helpful. High accuracy of the test is imperative; indeed, low sensitivity (i.e. high numbers of false negatives) would lead to missed cases of plague in the situation where the test is used as a screening tool (i.e. to exclude plague diagnosis when a negative result is observed and pursue with microbiological analysis when a positive result is observed), and to initial mismanagement of the case until it is confirmed microbiologically. Low specificity would probably lead to a bigger concern. A false-positive case might trigger unnecessary social alert, avoidable anxiety to the patients and their family, and avoidable use of resources, particularly in fragile health systems in countries where plague is endemic (Mavalankar 1995; Mead 2018). A highly specific test with a very low false-positivity rate would allow the adequate management of all negative cases, considering them true-negative cases.

Another important consideration to consider while evaluating the F1RDT is the use and importance of specificity of the test in diagnosing people with suspicion of pneumonic plague in the context of an outbreak. The pneumonic form of plague is a very severe and fatal disease that can be transmitted from human to human. Contrary to the bubonic form, where the presence of buboes might facilitate the suspicion of plague, symptoms presented with the pneumonic plague are less specific. In addition, the obtention of a good sample to run the F1RDT might be more difficult in the pneumonic plague, where it might be challenging to obtain good-quality sputum from children and from severely ill people with decreased consciousness. Performing the test in saliva instead of sputum will certainly lead to different accuracy findings, and this should be taken into account.

The F1RDT is a simple diagnostic tool that can be performed at the bedside of the patient, with a fast result that allows prompt diagnosis and early treatment, as well as timely implementation of control measures to limit the spread of the disease. Therefore, the F1RDT has the potential to be useful to health workers, and could contribute to reducing the high mortality attributed to plague, as well as inadequate public health responses to it. However, there is no systematic review assessing the diagnostic test accuracy of the F1RDT for plague against standard diagnostic tests. Currently, a confirmed case of plague is made either by isolation of *Y pestis* (culture) or by acute and convalescent serological antibody testing (four-fold difference in F1 antibody titres), according to World Health Organization (WHO) definitions (WHO 2019a). Positivity of either test provides a reliable diagnosis of plague and can be used interchangeably to consider a case of plague (although culture is preferred in order to identify the strain and resistance pattern). Therefore, it is reasonable to assess the accuracy of the F1RDT against both these tests. With the increasing inclusion of molecular biology for the diagnosis of many infectious diseases, we thought it was relevant to also assess accuracy of the F1RDT against PCR. However, to date, there is a lack of evidence supporting the superiority of one reference test above another.

The findings of this review will help to develop evidence-based recommendations on the role of the F1RDT in the diagnosis of plague, which could be included in clinical guidelines about the management of plague as well as in guidelines on infection prevention and control.

OBJECTIVES

To determine the diagnostic accuracy of the rapid diagnostic test (RDT) based on the antigen F1 (F1RDT) for detecting plague in people with suspected disease.

Secondary objectives

To assess the effect of forms of plague (bubonic, septicaemic, or pneumonic), specimen tested (bubonic aspirate, urine, or sputum), prior antibiotic treatment, location where the test is performed (field or laboratory studies), and threshold for detecting the disease (as set by the manufacturer) on the accuracy of the F1RDT for detecting the disease.

METHODS

Criteria for considering studies for this review

Types of studies

We included cross-sectional studies that assess the accuracy of the F1RDT for diagnosing plague in the laboratory or in field conditions, where patients were tested for plague with both the F1RDT and at least one of the reference standards (culture, PCR, or serology). We excluded case-control (two-gate cross-sectional) studies as we aim to determine accuracy of the RDT from only one set of participants, all of them with suspected plague.

Participants

We included participants (including children and pregnant women) living in or visiting areas where plague was endemic, who presented to any healthcare facility (primary, secondary, or tertiary care) with clinical suspicion of any form of plague. For studies where only a subgroup of participants was eligible for inclusion in the review, we included the study when there were disaggregated data that we could extract for that subgroup.

Index tests

The index test we assessed was the F1RDT performed in any relevant sample to detect plague, this was bubo aspirate to detect bubonic plague, sputum to detect pneumonic plague, and other samples such as urine to detect septicaemic plague.

Target conditions

The target condition was any form of symptomatic plague (bubonic, septicaemic, or pneumonic).

Reference standards

We included studies that used one of the following reference standards to diagnose plague.

- Isolation of *Y pestis* by culture.
- PCR.
- Serology showing a four-fold difference in F1 antibody titres between two paired samples.

Search methods for identification of studies

We conducted the literature search up to 15 May 2019 (27 August 2019 for PubMed), and identified potential studies regardless of language, publication status, or publication date.

Electronic searches

We searched the following databases using the search terms and strategy described in Appendix 1: the Cochrane Central Register of Controlled Trials (CENTRAL, published in the Cochrane Library, Issue 5, May 2019), MEDLINE (PubMed, from 1966 to 27 August 2019), Embase (Ovid; from 1949 to 15 May 2019), Science Citation Index (Web of Science, from 1900 to 15 May 2019). On 15 May 2019, we also searched Google Scholar (scholar.google.co.uk), and the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/en/), and ClinicalTrials.gov for trials in progress.

Searching other resources

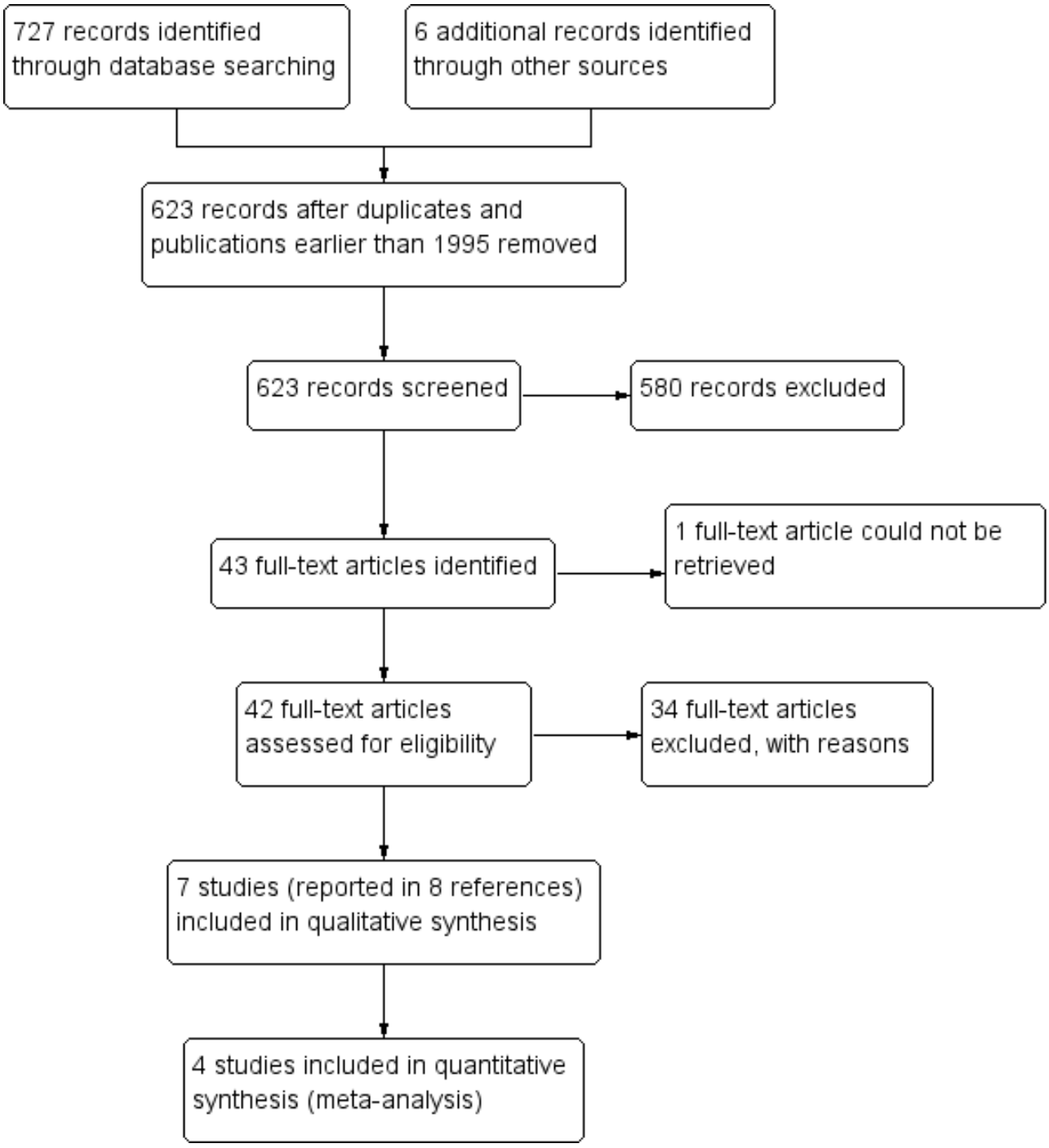
We searched the proceedings and abstracts of relevant conferences from the past five years: the International Symposium on Yersinia, the American Society of Tropical Medicine and Hygiene conference, and the congress of European Microbiologists. We handsearched the reference lists of relevant papers and contacted researchers working in the field. We also searched for related articles to the included studies using the PubMed "similar articles" function, on 27 August 2019.

Data collection and analysis

Selection of studies

Two review authors independently screened all the abstracts retrieved by the search strategy, using the predefined eligibility criteria. We excluded studies that were clearly irrelevant based on the titles and abstracts. We retrieved full-text copies of the remaining studies and applied the predefined criteria for inclusion in the review. We resolved any disagreements in assessment through discussion. We listed all studies excluded after full-text assessment in the Characteristics of excluded studies table. We illustrated the study selection process in a PRISMA diagram (Figure 1).

Figure 1. Study flow diagram.



Data extraction and management

One review author piloted the data extraction form on two studies. Based on the results of the pilot, we modified and finalized the data extraction form. Two review authors independently conducted data extraction and management, using the finalized data extraction form. We compared these data and resolved any disagreement through discussion. For each included study, we gathered information on the following (Appendix 2).

- Setting, design, and duration of the study.
- Baseline characteristics of the study population and sample size.
- Target condition: forms of plague assessed.
- Index test used: name, detection target, need for sample preparation, personnel who conducted the test, training provided to personnel for conducting the test, location where test performed.
- Reference standard: test performed, personnel who conducted the test, training provided to personnel for conducting the test, location where test performed, conditions of storage, and transport.

- Results for both index and reference standard tests: missing cases, uninterpretable results, true and false positives, true and false negatives, sensitivity and specificity of index tests.
- Other relevant details such as source of funding.

Assessment of methodological quality

Two review authors independently assessed the methodological quality of each included study, using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, which is based on four domains: participant selection, index test, reference standard, and flow and timing (Whiting 2011). We have tailored the tool to the context of this review (Appendix 3). We answered each of the signalling questions as 'yes,' 'no,' or 'unclear,' and gave the reason for our judgement. We resolved any disagreements through discussion and through consultation with a third review author in case of persisting disagreement.

Statistical analysis and data synthesis

We stratified all analyses by the reference standard used. Within each stratum, we constructed a two-by-two table (containing the number of true-positive, true-negative, false-positive, and false-negative results) for each study. We entered the two-by-two data into Review Manager 5 (Review Manager 2014). We summarized estimates of sensitivity and specificity from each individual study on forest plots and plotted the estimates using summary receiver operating characteristic (SROC) plots.

We calculated positive (PPV) and negative predictive value (NPV) estimates for various scenarios to help interpret the impact of F1RDT findings. There are no data reporting estimates of pretest probability of plague for the different case scenarios. Therefore, we estimated three case scenarios for each form of plague (bubonic and pneumonic plague), to the best of our knowledge and in consultation with experts in the field. The prevalence of 4% was calculated based on the number of confirmed cases of plague among notified cases of plague during the 2017 outbreak in Madagascar, as a starting point. We estimated a lower pretest probability of the disease at 0.1% for the context where plague was known to occur but where there was no declared ongoing outbreak. Finally, we estimated a higher pretest probability of the test, where the test was used in the context of a declared outbreak. In the case of bubonic plague in a context of known outbreak, a person presenting with a bubo has high probability of having the disease. Therefore, we estimated a high pretest probability of 50%. In the case of pneumonic plague, the differential diagnosis is broader than for bubonic plague, including other respiratory infections that are highly prevalent. Therefore, we estimated a pretest probability lower than for the case of bubonic plague, at 20%.

When meta-analysis was appropriate (given the number of studies and extent of clinical heterogeneity), we pooled results from the included studies. We used the bivariate model to obtain pooled estimates of sensitivity and specificity, as meta-analysis was performed for a single threshold. If data were sparse and the bivariate model would not converge, we pooled data using methods described by Takwoingi and colleagues (Takwoingi 2017). All meta-analyses were performed using the xtmelogit commands in Stata version 14 (Stata 2015). We plotted the pooled estimates of sensitivity and specificity using SROC plots in Review Manager 5 (Review Manager 2014).

Investigations of heterogeneity

We stratified the findings by type of reference standard used: bacterial isolation by culture, PCR, and serology showing a four-fold difference in F1 antibody titres between two samples from acute and convalescent phases. For PCR, we described all the genes that the included studies assessed. However, when we needed to choose one gene (i.e. for meta-analysis), we used *caf1*, which is the most relevant gene as according to experts in the field.

We planned to assess the impact of forms of plague on the accuracy of F1RDT by performing meta-regression. However, there were an insufficient number of studies to obtain reliable results from meta-regression. We instead presented the findings stratified by forms of plague, as this is clinically relevant.

We also planned to assess the impact of prior antibiotic treatment, location of performance of the F1RDT (field and central laboratory), and threshold for detecting the disease on accuracy of F1RDTs by performing subgroup analyses or meta-regression. It was not possible to assess the impact of prior antibiotic treatment or location of performance of the F1RDT test due to scarcity of the data. Furthermore, it was not necessary to assess the impact of threshold as studies were consistent with regards to the threshold used to determine disease status.

Sensitivity analyses

We planned to perform a sensitivity analysis in which we only included studies that had a low risk of bias for the four domains (patient selection, index test, reference standard, and flow and timing), or restricted to those studies at low risk of bias for patient selection only. However, it was not possible to perform this sensitivity analysis as none of the included studies were at low risk of bias for the patient selection domain.

Two manuscripts analyzed two cohorts with overlap of participants (Andrianaivoarimanana 2019; Rajerison 2020). We included Rajerison 2020 in the primary analysis and performed a sensitivity analysis in which we repeated the analysis including Andrianaivoarimanana 2019.

We identified an outlier in the primary analysis assessing RDT against culture for all forms of plague. We performed a sensitivity analysis to explore the impact of this outlier in the pooled estimate.

Assessment of reporting bias

Little is known on how to assess and detect reporting bias for diagnostic test accuracy studies (Macaskill 2010). We decided not to carry out a formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have important limitations when used in reviews of diagnostic test accuracy.

Assessment of the certainty of the evidence

We assessed the certainty of the evidence using GRADE and GRADEpro GDT software (GRADE Handbook 2013; GRADEpro GDT 2015; Schünemann 2020). We rated the certainty of the evidence as high, moderate, low, or very low by assessing four domains (risk of bias, indirectness, inconsistency, and imprecision), as follows.

- Risk of bias: we assessed risk of bias by using the QUADAS-2 tool.

- Indirectness: we used the QUADAS-2 tool to assess applicability concerns and looked for important differences between the populations studied, the setting, and the review question.
- Inconsistency: we explored inconsistency by investigating potential sources of heterogeneity, and we downgraded the certainty of the evidence when we could not explain inconsistency in the accuracy estimates.
- Imprecision: we considered the width of the confidence intervals (CIs) and questioned whether the truth set at the lower or upper limit of the 95% CI would change our decision. We calculated absolute numbers of true positives, true negatives, false positives, and false negatives, with ranges for these values based on the CIs of the pooled estimates of sensitivity and specificity for various prevalences of plague, and we made judgements on imprecision using these calculations.

We constructed 'Summary of findings' tables, which showed the main review findings along with the certainty of the evidence.

RESULTS

Results of the search

The search identified 727 records. We identified six additional papers by handsearching or by contacting experts. After excluding duplicates and manuscripts published before 1995 (before which F1RDT was not available), we selected 623 records for screening. We excluded 580 of them based on title and abstract. Among the 43 publications identified for full-text review, we could not retrieve the full-text of one of them. We assessed the full-text of 42 articles and excluded 34 of them, with the reasons for exclusion listed in the Characteristics of excluded studies table. See Figure 1 for the flow diagram of the study selection process.

Eight manuscripts met our inclusion criteria. Two manuscripts from Bertherat and colleagues reported the same plague outbreak that occurred in the DRC in 2005. As the manuscript published in 2005 did not provide additional data to the manuscript published in 2011, we merged the documents and referred to them in this review under one study ID (Bertherat 2011). Andrianaivoarimanana 2019 and Rajerison 2020 presented data on plague in Madagascar from national surveillance records, with an overlap of participants between the two manuscripts for the period between 2002 and 2007. We described the manuscripts separately in the Characteristics of included studies tables as they applied different exclusion criteria to the cohort of participants. However, we chose to include Rajerison 2020 in the primary analysis, as Rajerison 2020 presented additional data for two outbreaks with disaggregated data for bubonic and pneumonic plague.

We described the seven studies in the Characteristics of included studies table and their key findings in Table 1. Five studies reported findings of F1RDT used in plague epidemics in Madagascar between 2000 and 2019 (Andrianaivoarimanana 2019; Chanteau 2003a; Rajerison 2020; Richard 2015; Riehm 2011). One study assessed data from two outbreaks that occurred in the DRC in 2005 and 2006 (Bertherat 2011), and one study described performance of F1RDT from several clinics both in Madagascar and Uganda, from 2004 to 2017 (Petersen 2018).

The studies included in the primary analyses reported data from 4146 participants (Andrianaivoarimanana 2019 excluded) including adults and children of both sex. Among them, 3989 samples were analyzed with both the index test and at least one reference standard.

One study assessed only bubonic plague (Riehm 2011), two studies assessed only pneumonic plague (Bertherat 2011; Richard 2015), and four studies assessed all forms of plague (Andrianaivoarimanana 2019; Chanteau 2003a; Petersen 2018; Rajerison 2020). Samples used to perform F1RDT were bubo aspirates in participants with suspicion of bubonic plague and sputum in participants with suspicion of pneumonic plague (this was also assumed for Andrianaivoarimanana 2019 but was not clearly reported). F1RDT was also performed in a few postmortem tissue samples in two studies (Chanteau 2003a; Rajerison 2020). The postmortem tissue samples were not relevant to this review question; however, disaggregated data were not available, so we could not exclude them from our analysis. Rajerison 2020 excluded cases for which time of transport of the biological samples from collection to the laboratory was superior to seven days and cases who received antibiotics prior to sample collection.

The studies used two different F1RDT dipsticks. All studies except one (Petersen 2018) used the F1RDT produced at the Institut Pasteur of Madagascar (F1RDT-IPM), which is based on the combination of two anti-F1 antibodies (B18–1 and G6–18), with a lower detection threshold of the F1 antigen of 0.5 ng/mL. Petersen 2018 assessed a F1RDT from New Horizons (F1RDT-NH), utilized by the US Centers for Disease Control and Prevention (CDC), with a lower detection threshold of the F1 antigen of 1 ng/mL (personal communication). Additional and extensive data on performance of this test are due to be published. According to the manufacturers, both F1RDT-IPM and F1RDT-NH can be stored at room temperature.

Three studies performed F1RDT in a central laboratory (Chanteau 2003a; Petersen 2018; Rajerison 2020), and two studies both on site and in a central laboratory (Bertherat 2011; Riehm 2011). It was also performed on site in Chanteau 2003a, but data included for analysis were based from the findings of those performed in the central laboratory. It was unclear where the test was performed in Richard 2015. Andrianaivoarimanana 2019 did not report the location of performance of F1RDT, but we deduced it was in a central laboratory, as there was an overlap of participants from Rajerison 2020. The people who performed the F1RDT were a trained biologist in Bertherat 2011, and trained doctors, nurses, and health workers in Chanteau 2003a. Such information was not reported for the remaining studies.

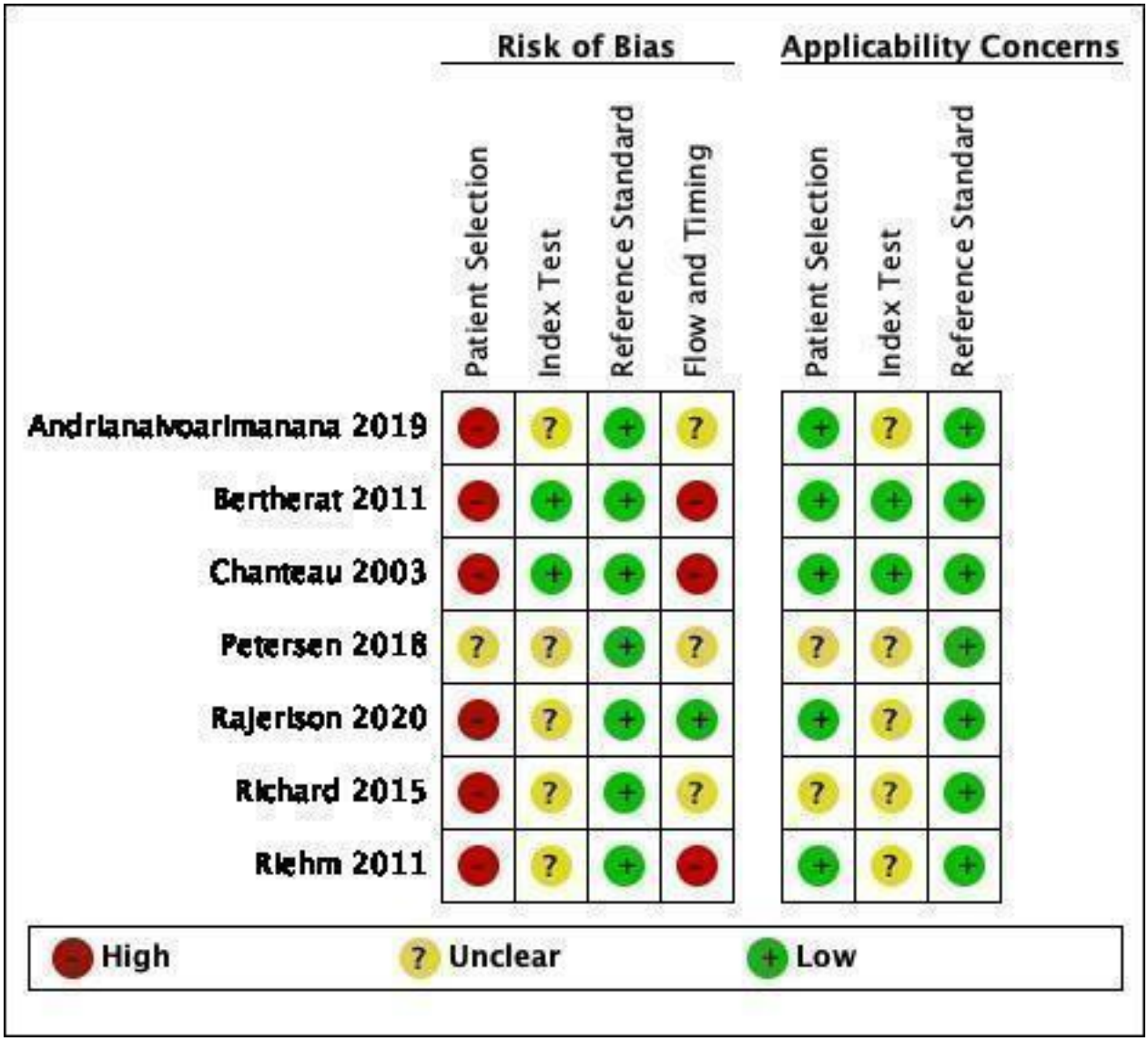
All studies compared F1RDT against culture, three studies against PCR (Bertherat 2011; Richard 2015; Riehm 2011), and two studies against paired serology (Petersen 2018; Richard 2015).

The Institut Pasteur of Madagascar, which produced one of the two dipsticks assessed in this review, supported two studies (Chanteau 2003a; Richard 2015). Other financial support or sponsors were the Institut Pasteur of Paris (Chanteau 2003a), the WHO (Rajerison 2020), and the President's Malaria Initiative/US Agency for International Development and the US Department of Homeland Security (Richard 2015). Four studies did not report the source of funding (Andrianaivoarimanana 2019; Bertherat 2011; Petersen 2018; Riehm 2011).

Methodological quality of included studies

See the Characteristics of included studies table for the assessment of the methodological quality of each included manuscript, and Figure 2 for risk of bias and applicability concerns summaries.

Figure 2. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.



In the patient selection domain, we considered all studies except one (Petersen 2018) at high risk of bias, because recruitment of participants was not consecutive or random but based on selection of participants based on clinicians who declared cases to the national surveillance system, following a retrospective design; in addition, there were unclear exclusion criteria. In some studies, administration of antibiotics prior to sample collection was not taken into account, which could have led to a false-negative culture when the participant truly had plague, and therefore misclassification of the test to detect the disease. We considered Petersen 2018 at unclear risk of bias as there were insufficient data to judge on this domain. Regarding applicability, five studies had low concern in the patient selection domain because the studies included the appropriate participants and settings (outbreak and national surveillance in endemic areas). Two studies were at unclear concern because there were insufficient data from Petersen 2018, and because there were limited information from the only two participants in Richard 2015.

In the index test domain, we judged two studies at low risk of bias because the F1RDT was interpreted without the knowledge of the results of the reference standard, and we considered them to have low concern for applicability (Bertherat 2011; Chanteau 2003a). The other five studies were at unclear risk of bias because we do not know if the result of the F1RDT was interpreted without the knowledge of the results of the reference standard. In addition,

these studies did not report the way that the test was conducted, such as who performed the test, where it was performed, or the need for sample preparation. We therefore considered the conduct and interpretation of the index test to be of unclear concern for applicability in these five studies (Andrianaivoarimanana 2019; Petersen 2018; Rajerison 2020; Richard 2015; Riehm 2011).

In the reference standard domain, we considered all studies at low risk of bias, as they all used culture, and four studies also used PCR, paired serology, or both, which correctly classify the target condition. For most studies, it was unclear whether the interpretation of the reference standard was blinded to the RDT result, but reference standard results are objective findings and we considered this to be at low risk to introduce bias, and at low concern of applicability.

In the flow and timing domain, we considered three studies at high risk of bias because the time of sample transportation was longer than seven days, which can lead to decreased viability of the pathogen and overgrowth of contaminating bacteria, with subsequent false-negative results and misclassification of the target condition (Bertherat 2011; Chanteau 2003a; Riehm 2011). In addition, in one study, all participants did not receive the same reference standard (Bertherat 2011). We considered one study at low risk of bias in this domain as they excluded participants who received antibiotics prior to enrolment and those for whom samples took longer than seven days to reach the central laboratory (Rajerison 2020). This was unclear for the remaining three studies.

Findings

We present the findings for all forms of plague, and then present findings disaggregated for pneumonic and bubonic plague. We could not pool the findings from the two different F1RDT used (F1RDT-IPM and F1RDT-NH) as the tests present different lower detection thresholds for considering the test positive. Furthermore, the only study that assessed the F1RDT-NH used combined reference standards to consider true cases of plague (Petersen 2018).

We present the summary of findings including true positive, false positive, false negative, true negative, sensitivity, and specificity for each study with disaggregated data on F1RDT for diagnosing the different forms of plague against the different reference standard in Table 2.

1. F1RDT for diagnosing all forms of plague

a. F1RDT from Institut Pasteur of Madagascar

F1RDT-IPM versus culture for diagnosing all forms of plague

Six studies reported data on F1RDT-IPM and culture findings for diagnosing all forms of plague. We used Rajerison 2020 in our primary analysis and performed sensitivity analysis with Andrianaivoarimanana 2019.

Primary analysis

Sensitivity estimates ranged from 25% to 100% and specificity estimates ranged from 51% to 82% (Table 2).

Two data set were not included in the meta-analysis; Richard 2015 reported zero true-positive and zero true-negative cases and, therefore, had no estimable sensitivity or specificity, and Bertherat 2011 reported a true-positive value of zero for the 2005 outbreak and, therefore, had no estimable sensitivity.

Four studies contributed to the final primary meta-analysis (Bertherat 2011 (2006 outbreak only); Chanteau 2003a; Rajerison 2020; Riehm 2011) (Figure 3). For the diagnosis of all forms of plague, F1RDT-IPM pooled sensitivity against culture was 100% (95% CI 82 to 100; 4 studies, 1692 participants; very low-certainty evidence) and specificity was 70% (95% CI 65 to 75; 4 studies, 2004 participants; very low-certainty evidence) (Summary of findings 1) (Figure 4). These results were from two univariate random-effects meta-analyses, one for sensitivity and one for specificity.

Figure 3. Forest plot of 1 F1 antigen rapid diagnostic test (F1RDT) versus culture for all forms of plague, primary analysis. The data for Bertherat 2011 refer to the 2006 outbreak only. CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive.

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bertherat 2011	4	19	0	73	1.00 [0.40, 1.00]	0.79 [0.70, 0.87]
Chanteau 2003a	151	128	31	381	0.83 [0.77, 0.88]	0.75 [0.71, 0.79]
Rajerison 2020 (2002-2007 Bubonic)	1328	324	0	667	1.00 [1.00, 1.00]	0.67 [0.64, 0.70]
Rajerison 2020 (2002-2007 Pneumonic)	40	45	0	64	1.00 [0.91, 1.00]	0.59 [0.49, 0.68]
Rajerison 2020 (2017-2018 Bubonic)	22	2	0	9	1.00 [0.85, 1.00]	0.82 [0.48, 0.98]
Rajerison 2020 (2017-2018 Pneumonic)	1	23	3	40	0.25 [0.01, 0.81]	0.63 [0.50, 0.75]
Rajerison 2020 (2018-2019 Bubonic)	57	23	0	71	1.00 [0.94, 1.00]	0.76 [0.66, 0.84]
Rajerison 2020 (2018-2019 Pneumonic)	8	6	0	27	1.00 [0.63, 1.00]	0.82 [0.65, 0.93]
Riehm 2011	47	41	0	61	1.00 [0.92, 1.00]	0.60 [0.50, 0.69]

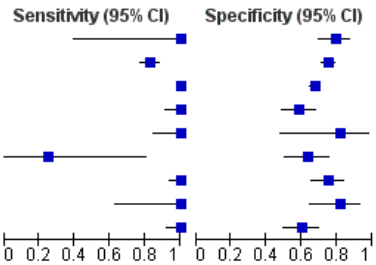
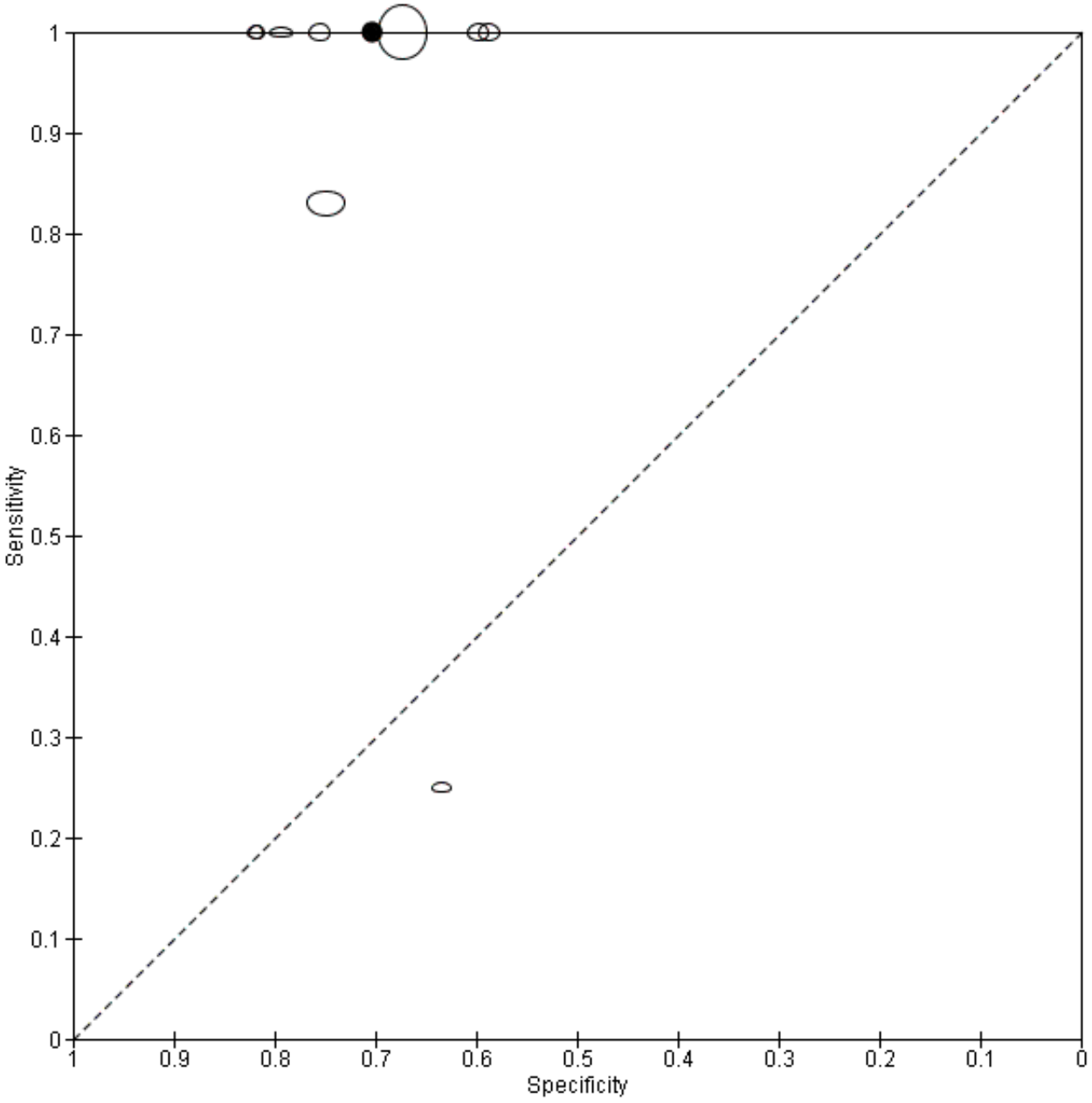


Figure 4. Summary receiver operating characteristic (SROC) plot of one F1 antigen rapid diagnostic test (F1RDT) versus culture for all forms of plague, primary analysis. Pooled sensitivity: 100% (95% CI 82 to 100); pooled specificity: 70% (95% CI 65 to 75). The solid black dot corresponds to the pooled estimate of sensitivity and specificity; the black circles show individual study results and their size corresponds to the sample size of the study contributing to the analysis.



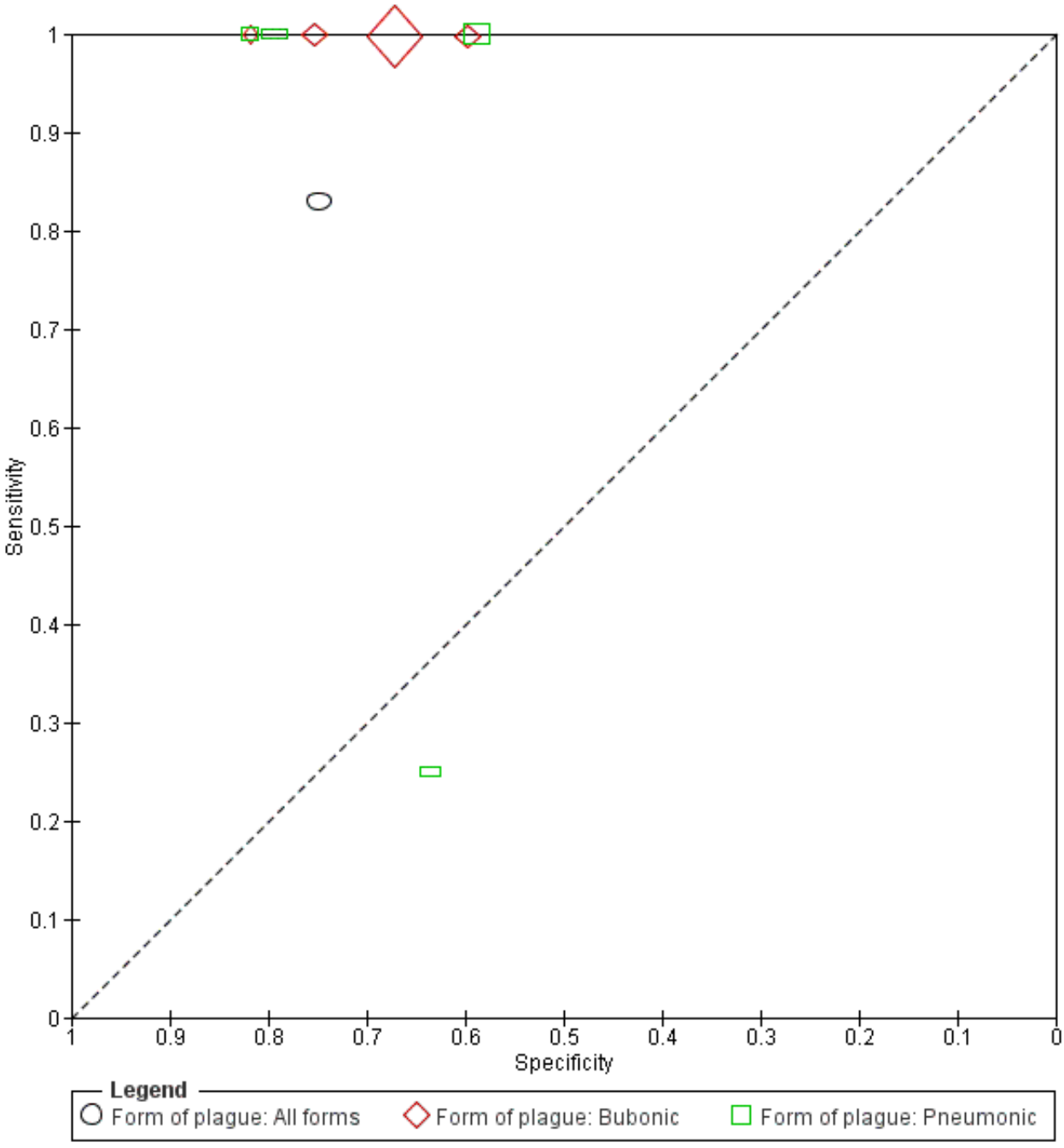
By considering the forest plot, we found that sensitivity from participants with pneumonic plague in the 2017 to 2018 period reported by Rajerison 2020 were not in line with the rest of the sensitivity estimates by the other studies. There was a limited number of four cases in this period, all four had positive culture but only one had a positive RDT. This could have been due to several factors such as the quality of sputum samples. We decided to keep this outlier in the meta-analysis as it did not reflect any situation inherently different from the other studies.

Out of the 691 biological samples evaluated by Chanteau 2003a, RDT was performed on 35 postmortem samples from lung or liver puncture. The authors did not present disaggregated data for those samples or by forms of plague. As these samples constituted 5% of the overall samples evaluated in this study, we decided to include these findings.

We did not investigate whether type of plague explained heterogeneity by introducing a covariate to the bivariate model, as there were an insufficient number of studies to produce

reliable results. However, we did produce an SROC plot which allowed visual examination of how sensitivity and specificity varied between subgroups (Figure 5). There was no clear pattern within the limited number of studies to suggest that differences existed between the different forms of plague.

Figure 5. Summary receiver operating characteristic (SROC) plot of one F1 antigen rapid diagnostic test (F1RDT) versus culture for all forms of plague, primary analysis, for investigation of heterogeneity. Each circle (all forms of plague with no disaggregated data), square (pneumonic plague), and diamond (bubonic plague) corresponds to individual study results. Their size corresponds to the sample size that contributed to the estimate of sensitivity and specificity.

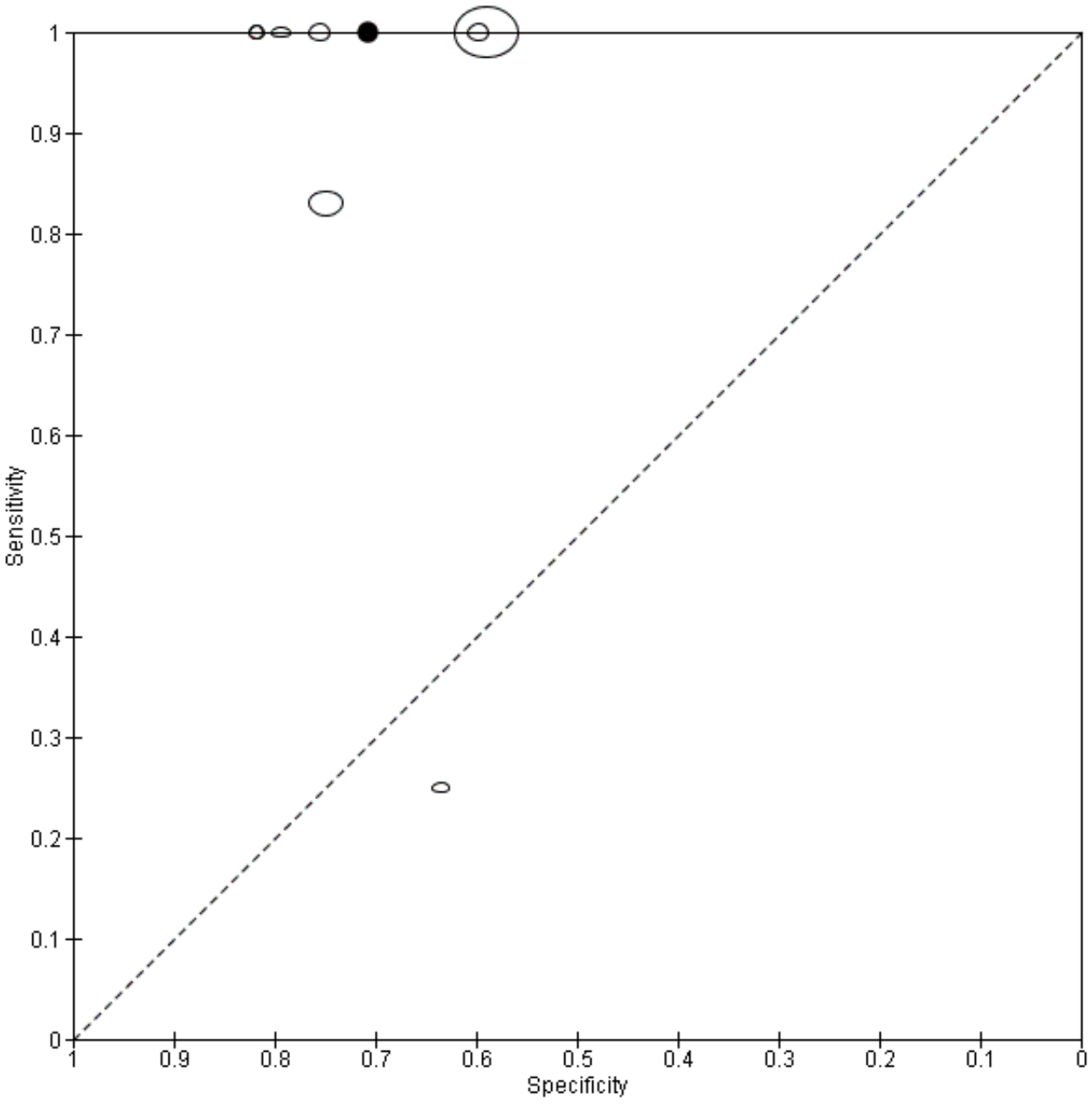


Sensitivity analyses

We performed sensitivity analysis by excluding Rajerison 2020 and including Andrianaivoarimanana 2019. Both studies were conducted from surveillance data in Madagascar, but Rajerison 2020 presented disaggregated data for participants with bubonic and pneumonic forms of plague for three different times (2002 to 2007, 2017 to 2018, and 2018 to 2019), and Andrianaivoarimanana 2019 presented sensitivity and specificity estimates from all forms of plague during the 2002 to 2007 period. In addition, and contrary to Andrianaivoarimanana 2019, Rajerison 2020 excluded from analysis cases for which time of transport of biological samples was more than seven days, and cases who received antibiotics prior to sample collection, which would ensure a better performance of the standard test for comparison.

We found very similar findings to the primary analysis, with a pooled sensitivity of 100% (95% CI 70 to 100) and pooled specificity of 71% (95% CI 65 to 77) (Figure 6). These results were from two univariate random-effects meta-analyses, one for sensitivity and one for specificity.

Figure 6. Summary receiver operating characteristic (SROC) plot of 2 F1 antigen rapid diagnostic test (F1RDT) versus culture for all forms of plague, sensitivity analysis. Pooled sensitivity: 100% (95% CI 70 to 100); pooled specificity: 71% (95% CI 65 to 77). The solid black dot corresponds to the pooled estimate of sensitivity and specificity; the black circles show individual study results and their size corresponds to the sample size of the study contributing to the analysis.



We also performed sensitivity analysis by excluding the data on pneumonic plague from Rajerison 2020 for the 2017 to 2018 period, identified as an outlier. Although the CI for sensitivity was wider, we found no difference to the pooled results as compared with the primary analysis (pooled sensitivity of 100%, 95% CI 49 to 100 and pooled specificity of 70%, 95% CI 65 to 74). These results were from one bivariate random-effects meta-analysis,

F1RDT-IPM versus PCR for diagnosing all forms of plague

Three studies provided data on both F1RDT-IPM and PCR findings (Bertherat 2011, 2005 outbreak; Richard 2015; Riehm 2011). We summarized the outcomes in Table 2. Sensitivity estimates ranged from 72% to 95% and specificity estimates ranged from 50% to 93%. The specificity estimate of 50% was based from a sample size of two participants (Bertherat 2011, 2005 outbreak).

Riehm 2011 assessed the performance of three different genes by PCR in the same set of 149 participants. We calculated a sensitivity of 95% and specificity of 93% for the *caf1* gene, 72% and 93% for the *pla* gene, and 92% and 87% for the *Ymt* gene.

Bertherat 2011 (2005 outbreak) and Richard 2015 provided data on two participants each. It was not possible to estimate sensitivity in Bertherat 2011 (2005 outbreak) as the reference standard identified no cases of plague, and it was not possible to assess specificity in Richard 2015 as the reference standard only identified cases of plague. With one study remaining (Riehm 2011), we could not perform meta-analysis to assess the performance of F1RDT-IPM versus PCR.

F1RDT-IPM versus paired serology for diagnosing all forms of plague

Only one study reported data with F1RDT-IPM and paired serology findings (Richard 2015). Both tests (F1RDT-IPM and paired serology) were performed on sputum in two participants, and were positive in both participants (Table 2).

b. F1RDT from New Horizons

One study reported findings of the F1RDT-NH for diagnosing plague (Petersen 2018).

Confirmed cases of plague were defined as isolation of *Y pestis* by culture in bubo, blood, or sputum, or 4-fold change in titre between acute and convalescent serum samples. Participants confirmed as non-plague were defined as *Y pestis* not isolated from culture and no significant change in titre between acute and convalescent serum samples.

From the limited available data, there were 118 confirmed cases of plague (34 from Madagascar and 84 from Uganda); among which 109 where cases of bubonic plague. There were 136 participants tested and confirmed as non-plague (61 from Madagascar and 75 from Uganda), among which 116 were suspected bubonic plague.

F1RDT-NH gives a semi-quantitative result according to the intensity of the line in the dipstick, from 1+ to 4+. Where positivity of F1RDT-NH was interpreted from 1+ (this is as soon as the line is visible), the test presented a sensitivity of 90.6% (95% CI 83.8 to 95.2) and a specificity of 88.0% (95% CI 82.9 to 93.3), as reported by the study authors. When evaluating F1RDT-NH with positivity results from 2+, the test presented a sensitivity of 87.4% (95% CI 80.1 to 93.0) and a specificity of 97.7% (95% CI 93.4 to 99.5).

Comprehensive findings of this F1RDT-NH should be published soon (personal communication).

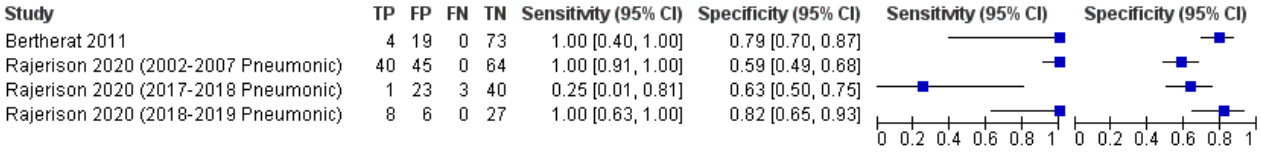
2. F1RDT for diagnosing pneumonic plague

a. F1RDT from Institut Pasteur of Madagascar

F1RDT-IPM versus culture for diagnosing pneumonic plague

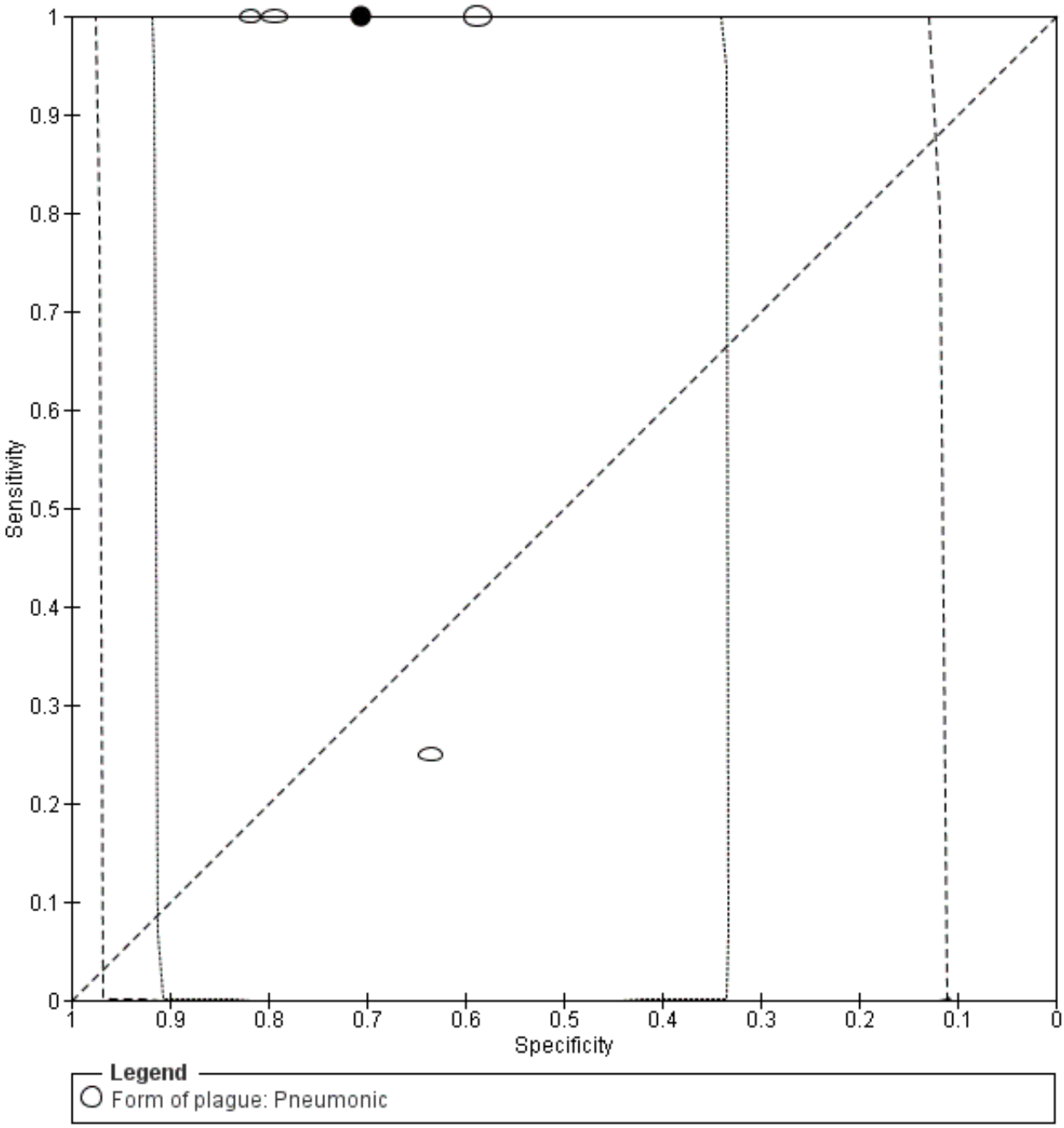
Three studies provided findings from both F1RDT-IPM and culture for diagnosing pneumonic plague (Bertherat 2011; Rajerison 2020; Richard 2015) (Table 2). Sensitivity was 100% with broad 95% CI (the lowest lower limit was 40%), except for the 2017 to 2018 period reported by Rajerison 2020 that showed a sensitivity of 25% (Figure 7). There was a limited number of four cases during this period, which had positive culture but only one of them had a positive RDT. This could be due to several factors such as technical issues in preparing the sputum sample for analysis or in running the F1RDT. Specificity estimates ranged from 51% to 82%.

Figure 7. Forest plot of 1 F1RDT versus culture for pneumonic plague. CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive.



Only two studies provided data that allowed estimation of both sensitivity and specificity of the test (Bertherat 2011; Rajerison 2020), and so these two studies contributed to the meta-analysis. Performed in sputum, F1RDT-IPM pooled sensitivity against culture was 100% (95% CI 0 to 100; 2 studies, 56 participants; very low-certainty evidence) and pooled specificity was 71% (95% CI 59 to 80; 2 studies, 297 participants) (Summary of findings 2) (Figure 8). These results were from one bivariate random-effects meta-analysis.

Figure 8. Summary receiver operating characteristic (SROC) plot of 1 F1 antigen rapid diagnostic test (F1RDT) versus culture for pneumonic plague. Pooled sensitivity: 100% (95% CI 0 to 100); pooled specificity: 71% (95% CI 59 to 80). The solid black dot corresponds to the pooled estimate of sensitivity and specificity; the black circles show individual study results and their size corresponds to the sample size of the study contributing to the analysis.



F1RDT-IPM versus PCR for diagnosing pneumonic plague

Bertherat 2011 (2005 outbreak) and Richard 2015 provided findings of the F1RDT-IPM for diagnosing pneumonic plague. However, it was not possible to assess sensitivity in Bertherat 2011 (2005 outbreak) as there were no cases of plague identified by PCR (sensitivity cannot be calculated when there are no cases identified as this would involve dividing by zero), and it was not possible to assess specificity in Richard 2015 as PCR only identified cases of plague (specificity cannot be calculated when only cases are identified as this would involve dividing by zero). Therefore, we did not pool data from these two studies in a meta-analysis.

F1RDT-IPM versus paired serology for diagnosing pneumonic plague

Richard 2015 provided findings of the F1RDT-IPM for diagnosing two participants with pneumonic plague, but it was not possible to assess specificity as the paired serology only identified cases of plague.

b. F1RDT from New Horizons

There were no disaggregated findings for pneumonic plague from the F1RDT-NH test.

3. F1RDT for diagnosing bubonic plague

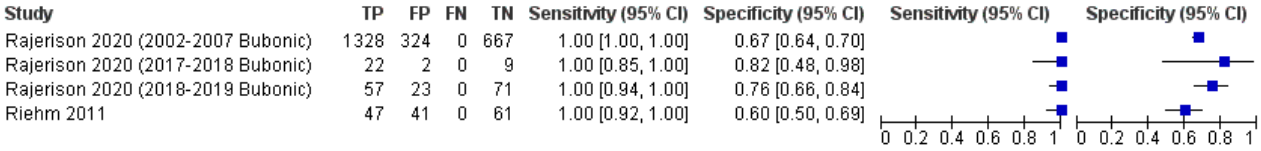
There was no disaggregated findings for bubonic plague from the New Horizons test. The findings below are from studies assessing the F1RDT-IPM.

a. F1RDT from Institut Pasteur of Madagascar

F1RDT-IPM versus culture for diagnosing bubonic plague

Two studies performed both the F1RDT-IPM and culture in participants with suspicion of bubonic plague (Rajerison 2020; Riehm 2011) (Table 2). Sensitivity estimates were 100% for all the time periods presented by the two studies, thereby it was not possible to pool data for sensitivity estimates (Figure 9). Specificity estimates ranged from 60% to 82%. F1RDT-IPM performed in bubo aspirate for diagnosing bubonic plague showed a pooled specificity of 67% (95% CI 65 to 70; 2 studies; 1198 participants).

Figure 9. Forest plot of 1 F1RDT versus culture for bubonic plague. CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive.



F1RDT-IPM versus PCR for diagnosing bubonic plague

One study assessed the performance of three different genes by PCR in the same set of 149 participants (Riehm 2011). We calculated the sensitivity and specificity of the F1RDT-IPM against each of these genes, and found sensitivity of 95% (95% CI 89 to 99) and specificity of 93% (95% CI 84 to 98) for the *caf1* gene, 72% (95% CI 63 to 80) and 93% (95% CI 77 to 99) for the *pla* gene, and 92% (95% CI 84 to 97) and 87% (95% CI 76 to 94) for the *Ymt* gene (Table 2).

F1RDT-IPM versus paired serology for diagnosing bubonic plague

We found no studies assessing the accuracy of F1RDT-IPM against paired serology for the diagnosis of bubonic plague.

b. F1RDT from New Horizons

There were no disaggregated findings for bubonic plague from the F1RDT-NH test.

DISCUSSION

Summary of main results

This systematic review of the diagnostic accuracy of F1RDT for diagnosing any form of plague in people with suspected disease summarizes the current literature and includes seven studies described in eight manuscripts. Three studies described the performance of F1RDT for diagnosing pneumonic plague and two studies for diagnosing bubonic plague, with no disaggregated data for forms of plague in three studies. The evidence came from three African countries (Madagascar, the DRC, and Uganda) with findings of F1RDT used during outbreaks or surveillance system in settings where plague is endemic. All studies except one used the F1RDT produced at the Institut Pasteur of Madagascar (F1RDT-IPM), while the remaining study assessed a F1RDT from New Horizons (F1RDT-NH), which is utilized by the US CDC. All studies were considered at high risk of bias for the patient domain, mainly due to the retrospective design of the studies, the absence of consecutive or random sampling with clear inclusion and exclusion criteria, or both. The major findings of this review include the following.

For the diagnosis of any form of plague, the accuracy of F1RDT-IPM is:

- against culture, sensitivity of 100% (95% CI 82 to 100; 4 studies, 1692 participants; very low-certainty evidence) and specificity of 70% (95% CI 65 to 75; 4 studies, 2004 participants; very low certainty evidence) (Summary of findings 1);
- against PCR, sensitivity estimates between 72% and 95% and specificity estimates between 50% and 93%, but meta-analysis was not possible;
- against paired serology, very limited evidence from two participants in one study.

For the diagnosis of pneumonic plague, the accuracy of F1RDT-IPM is:

- against culture, sensitivity of 100% (95% CI 0 to 100; 2 studies, 56 participants; very low-certainty evidence) and specificity of 71% (95% CI 59 to 80; 2 studies, 297 participants; very low-certainty evidence) (Summary of findings 2);
- against PCR, very limited evidence from four participants in two studies;
- against paired serology, very limited evidence from two participants in one study.

The performance of F1RDT-IPM for diagnosing bubonic plague:

- against culture, sensitivity of 100% (CI not calculable; 2 studies, 1454 participants; low-certainty evidence), and specificity of

67% (95% CI 65 to 70; 2 studies, 1198 participants; very low-certainty evidence) (Summary of findings 3);

- against PCR, sensitivity of 95% (95% CI 89 to 99; 1 study, 88 participants; very low-certainty evidence), and specificity of 93% (95% CI 84 to 98; 1 study, 61 participants; very low-certainty evidence) (Summary of findings 4);
- against paired serology, no data.

The performance of the F1RDT-NH for diagnosing any form of plague against combined culture or paired serology showed sensitivity of 91% (95% CI 84 to 95) and specificity of 88% (95% CI 83 to 93), as reported by the study authors. When evaluating the F1RDT-NH with positivity results from 2+, the sensitivity decreased to 87% (95% CI 80 to 93) and the specificity increased to 98% (95% CI 93 to 100).

F1RDT for diagnosing pneumonic plague

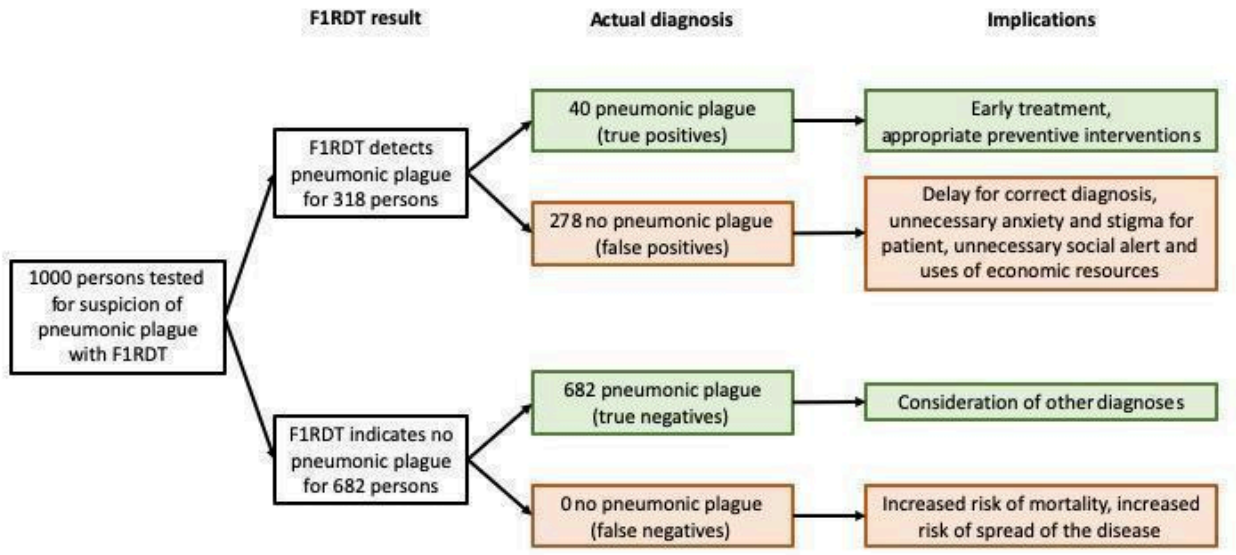
Early diagnosis of pneumonic plague is critical so that prompt treatment is started (pneumonic plague is associated with high fatality rate if left untreated) and so that preventive measures are established to limit transmission of the disease (pneumonic plague can be transmitted from human to human by inhalation of respiratory droplets produced by coughing).

Against culture results, sensitivity appeared high (100%) but three participants testing negative in one outbreak in a meta-analysis that included data for 56 cases resulted in very wide CIs (0 to 100). The specificity was 71% (95% CI 59 to 80). However, these estimates came from a small number of participants and there was very low-certainty evidence.

We used different case scenarios where pretest probability of plague varied from 0.1% to 4% to 20% to simulate scenarios where F1RDT could be used in plague endemic areas and in situations where an outbreak is in progress (Summary of findings 2). This numerical approach should be interpreted with caution due to the limitations cited above.

- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where one (0.1%) of them actually had pneumonic plague identified by culture, we estimated that the test would correctly diagnose the one person with pneumonic plague (1 true positive, 95% CI 0 to 1) and would not miss any people with pneumonic plague (0 false negatives, 95% CI 0 to 1), but would diagnose 290 people with pneumonic plague who were culture negative (290 false positives, 95% CI 200 to 410).
- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where 40 (4%) of them actually had pneumonic plague identified by culture, we estimated that the test would correctly diagnose all 40 people (40 true positives, 95% CI 0 to 40) and would not miss any people with plague (0 false negative, 95% CI 0 to 40), but would diagnose 278 people with pneumonic plague who were culture negative (278 false positives, 95% CI 192 to 394). In this case scenario, the PPV would be 12.6%. This means that among 318 people with positive F1RDT, 12.6% of them would actually have pneumonic plague (see Figure 10). The NPV would be 100%, which means that among the 682 people with negative F1RDT, all of them truly will not have pneumonic plague.
- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where 200 (20%) of them actually had pneumonic plague identified by culture, we estimated that the test would correctly diagnose all 200 people (200 true positives, 95% CI 0 to 200) and would not miss any people with plague (0 false negative, 95% CI 0 to 200), but would diagnose 232 people with pneumonic plague who were culture negative (232 false positives, 95% CI 160 to 328).

Figure 10. Flow diagram summarizing the main results in a hypothetical cohort with 4% of people with pneumonic plague (adapted from van Hoving 2019). F1RDT: F1 antigen rapid diagnostic test.



The high sensitivity associated with F1RDT for diagnosing pneumonic plague means that the test would detect all cases of pneumonic plague and would not miss any people with diagnosis of pneumonic plague. The high NPV means that if result of F1RDT is negative, pneumonic plague can be ruled out. However, the relatively low specificity and low PPV associated with F1RDT are of concern for the potential significant repercussions in the context of this disease. False positives mean that people who do not have plague will be considered as having plague, with repercussion for both the person (missing the true diagnosis, unnecessary anxiety and stigma related to the diagnosis of plague) and for society (unnecessary social alert and use of economic resources) (Figure 10).

The high apparent level of false positives with culture as the reference standard may actually be a result of people with plague testing culture negative; prior antibiotic use, for example, could cause this. Culture is known to be an imperfect reference standard for plague, as reminded in Andrianaivoarimanana 2019 where authors mentioned that "negative culture results might have resulted from administration of antimicrobial drugs by local health officials before sampling, whereas F1RDT results remained positive >3 weeks after treatment initiation." This would result in an underestimation of the specificity.

There were insufficient data to make any estimates against PCR or paired serology as reference standard.

F1RDT for diagnosing bubonic plague

Early and correct diagnosis of bubonic plague is essential for correct management of the patient and adequate public health measures and rodent control.

Two studies assessed F1RDT against culture, with a mean sensitivity of 100% (95% CI not calculable), and specificity of 67% (95% CI 65 to 70). However, the evidence was low certainty for sensitivity and very-low certainty for specificity.

One study assessed F1RDT against PCR for three genes (*caf1*, *pla*, *Ymt*). For *caf1*, which is considered the most relevant gene according to experts in the field, sensitivity was 95% (95% CI 89 to 99) and specificity was 93% (95% CI 84 to 98).

We used different case scenarios where pretest probability of plague varied from 0.1% to 4% to 50% to simulate scenarios where F1RDT could be used in plague endemic areas and in situations where an outbreak is in progress. This numerical approach should be interpreted with caution due to the limitations cited above.

When true cases of bubonic plague were diagnosed with culture, the main findings were the following (Summary of findings 3).

- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where one (0.1%) of them actually had bubonic plague identified by culture, we estimated that the test would correctly diagnose the one person with bubonic plague (1 true positive, 95% CI not calculable) and would not miss any people with bubonic plague (0 false negatives, 95% CI not calculable), but would diagnose 330 people with bubonic plague who were culture negative (330 false positives, 95% CI 300 to 350).
- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where 40 (4%) of them actually had bubonic plague identified by culture, we estimated that the test would correctly diagnose all 40 people (40 true positives, 95% CI not calculable) and would not miss any people with bubonic plague (0 false negatives, 95% CI not calculable), but would diagnose

317 people with bubonic plague who were culture negative (317 false positives, 95% CI 288 to 336). In this case scenario, the (PPV would be 11.2%. This means that among 357 people with positive F1RDT, 11.2% of them would actually have bubonic plague. The NPV would be 100%, which means that among the 643 people with negative F1RDT, all of them truly would not have bubonic plague.

- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where 500 (50%) of them actually had bubonic plague identified by culture, we estimated that the test would correctly diagnose all 500 people (500 true positives, 95% CI not calculable) and would not miss any people with bubonic plague (0 false negative, 95% CI not calculable), but would diagnose 165 people with bubonic plague who were culture negative (165 false positives, 95% CI 150 to 175).

When true cases of bubonic plague are diagnosed with PCR targeting the *caf1* gene, the main findings are the following (Summary of findings 4).

- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where one (0.1%) of them actually had bubonic plague identified by PCR, we estimated that the test would correctly diagnose the one person with bubonic plague (1 true positive, 95% CI 1 to 1) and would not miss any people with bubonic plague (0 false negatives, 95% CI 0 to 0), but would diagnose 70 people with bubonic plague who were PCR negative (70 false positives, 95% CI 20 to 160).
- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where 40 (4%) of them actually had bubonic plague identified by PCR, we estimated that the test would correctly diagnose 38 people (38 true positives, 95% CI 36 to 40), but would miss two people with bubonic plague (2 false negatives, 95% CI 0 to 4) and would diagnose 67 people with bubonic plague who were PCR negative (67 false positives, 95% CI 19 to 154). In this case scenario, the PPV would be 36.1%. This means that among 105 people with positive F1RDT, 36.1% of them would actually have bubonic plague. The NPV would be 99.8%, which means that among the 895 people with negative F1RDT, 99.8% of them truly would not have bubonic plague.
- If F1RDT was used in a hypothetical cohort of 1000 people with symptoms where 500 (50%) of them actually had bubonic plague identified by PCR, we estimated that the test would correctly diagnose 475 people (475 true positives, 95% CI 445 to 495), but it would miss 25 people with bubonic plague (25 false negatives, 95% CI 5 to 55) and would diagnose 35 people with bubonic plague who were PCR negative (35 false positives, 95% CI 10 to 80).

Similar to its performance for diagnosing pneumonic plague, the high sensitivity associated with F1RDT for diagnosing bubonic plague means that the test would detect all cases (when compared to culture) or most cases (when compared to PCR) of bubonic plague. The high NPV means that when F1RDT shows a negative finding, bubonic plague can be ruled out.

We found higher specificity and PPV estimates when F1RDT was compared to PCR than to culture. This further demonstrates that culture is an imperfect reference standard. In the case of prior use of antibiotics, culture is likely to become negative while PCR might still be positive by detecting DNA of *Y pestis* after several doses of antibiotics.

We found no studies that determined the performance of F1RDT against paired serology for diagnosing bubonic plague.

Strengths and weaknesses of the review

Strengths and weaknesses of the included studies

The small number of studies and participants included in the analyses is a major limitation of the review with regards to estimation of accuracy. The sensitivity estimates had broad CIs, which underlies the imprecision of the estimates. The lower and higher values of the CIs provide very different scenarios that would lead to different decisions in practice. The study design of the included studies contributing to the high risk of selection bias further weakened our confidence in the results – we judged all included studies at high risk of bias in the patient selection domain when applying the QUADAS-2 tool. We found no study that assessed the accuracy of F1RDT for diagnosing plague by prospectively recruiting consecutive patients with clear inclusion and exclusion criteria. There are also some limitations with regards to the reference standards. Although they were considered at low risk of introducing bias due to the objective results they provided, culture could be considered as an imperfect reference standard when antibiotics are given prior to sample collection, and there is a lack of standardized targeted gene for identification *Y pestis* with PCR.

Strengths and weaknesses of the review process

The findings of this review are based on a comprehensive literature search with no restriction in language, strict inclusion criteria, duplicate data extraction, and rigorous assessment of risk of bias using the QUADAS-2 tool tailored to our review question. The inclusion of unpublished data at the time of conducting the review constitutes a strength of the review process.

We attempted to contact study authors when there was poor reporting that limited data extraction or our judgement on applicability of the findings.

The prevalence of plague is not well established. For the hypothetical cohort scenarios, we presented above and in the 'Summary of finding' tables, we chose different values from 0.1% to 50% corresponding to pretest probability in order to contemplate different scenarios. For this, we considered using RDT where there was no ongoing outbreak (lower pretest probability) or where there was ongoing outbreak (higher pretest probability), as well as the form of plague (signs and symptoms of bubonic plague are more specific than those of pneumonic plague, therefore the pretest probability of plague when conducting F1RDT in a person with fever and a bubo in an endemic area is likely to be higher than when conducting F1RDT in a person with fever and productive cough).

Sources of heterogeneity could not be explored due to the scarcity of the included studies and participants. However, we calculated the test accuracy and presented the findings disaggregated by form of plague and against different reference standards that are commonly used in practice.

Applicability of findings to the review question

Inclusion criteria of this review were broad and, as such, were a good representation of the real scenarios in which F1RDT would be used. Overall, we had low concern of the applicability of findings

from the included studies to our review question, when assessed with the tailored QUADAS-2 tool.

The participant characteristics and settings matched our review question. For the reference standard domain, all the studies had low concern for applicability. With regards to the index test domain, F1RDT was conducted in central laboratories for the majority of participants who contributed to the primary analyses. In most cases, it was unclear whether the index test was performed by specialized and trained staff or by healthcare workers who may not have been adequately trained or may not have used a rapid test by immunochromatography previously. Therefore, it is possible that the accuracy of the test is lower when used on the bedside of the patients in the field. Another aspect to consider is the sample collection. Collection of sputum from a sick person during an outbreak is often performed in a context of panic and can be challenging, leading to collection of saliva instead. In the case of people with suspected bubonic plague, the obtention of pus from the bubo is usually simpler but can also be challenging depending on the size of the bubo. However, this is likely to affect the yield of all diagnostic tests performed in the relevant sample (sputum or bubo aspirate), including F1RDT, culture, and PCR, with no expected repercussion on the accuracy of the F1RDT against culture or against PCR. Poor-quality samples (sputum or bubo aspirate) may lead to a decreased sensitivity when comparing the F1RDT against paired serology (performed in blood samples).

AUTHORS' CONCLUSIONS

Implications for practice

The use of antigen F1 rapid diagnostic test (F1RDT) needs to be contextualized in order to assess its implications for practice. Two main factors may change considerably the interpretation of the accuracy findings of F1RDT.

The first factor is the form of plague as consequences of false positives and false negatives might differ between bubonic and pneumonic plague, the latter being more severe and transmissible from human to human. Furthermore, clinical diagnosis of pneumonic plague can be more difficult than diagnosis of bubonic plague, as pneumonic plague shares signs and symptoms of other, highly prevalent, respiratory infections. This introduces uncertainty to the assessment of the pretest clinical probability; the corresponding expected numbers of false positives and false negatives would also be uncertain. In addition, the difficulties of sputum collection and obtaining good-quality sputum samples leads to limitations of the yield of all the diagnostic techniques performed on sputum samples. These include F1RDT, but also culture and polymerase chain reaction. Diagnostic tests performed in poor-quality samples would not detect some cases of pneumonic plague. But because this affects both rapid diagnostic tests and the reference standards (culture and polymerase chain reaction), this does not affect our pooled estimates of F1RDT against these reference standards.

The second factor is whether the test is used during an outbreak or not, leading to different prevalence of the disease in which the test will be used, with subsequent variation in the absolute number of false positives and false negatives.

When F1RDT is used where plague is known to occur but there is no declared ongoing outbreak, the test can have a key role in detecting the beginning of a potential outbreak. In such situations, it seems clear that F1RDT, which is easy-to-use, cheap, and provides fast results even in remote or resource-restricted areas would be desirable. The specificity of the test was estimated to be 70% for detecting pneumonic plague. The number of false-positive results is likely to lead to unnecessary social alarm with non-negligible repercussions. However, F1RDT is being considered as a diagnostic test that may be used as an add-on to clinical suspicion. Therefore, clinicians must consider the need of additional testing in case of positive findings. False-negative results might contribute to the spread of the disease, and increase risk of mortality for the missed case. F1RDT showed sensitivity and NPV of 100% to detect pneumonic plague. A negative result would, therefore, rule out the disease. The broad confidence interval and the very low-certainty evidence in the estimate oblige us to consider adding additional testing in order to rule out plague, and to take into consideration the clinical suspicion and the real epidemiological scenario.

When F1RDT is used in a context of a declared outbreak, the test was found to present high NPV and consequently negative results help to rule out the disease and prompt us to consider alternative diagnoses. This is significant in panic situations where considerable numbers of people will present with the fear of having the disease. However, it is important that the clinician interprets the result while considering the clinical presentation and epidemiological context. False negatives might lead to a less negative impact than in the previous scenario, as the patient is likely to receive treatment independently of the result of the test (or to be closely followed) if there is high clinical suspicion of plague. However, false positives might be more of a concern as the patients will receive the diagnosis of plague and this is likely to prevent clinicians considering alternative diagnoses and adequate treatment might be delayed. False positives are also associated with unnecessary anxiety and stigma for the patients and their families. With a higher pretest probability (within a context of an outbreak), cases of false positives will increase. Therefore, for positive RDTs, the test needs to be combined with other laboratory evaluations to confirm the diagnosis, but treatment can be started.

This review contributed to provide accuracy data for the World Health Organization (WHO) Guideline Development Group Meeting in September 2019. Both the report of the meeting and the WHO plague guidelines are expected to be in the public domain later in 2020.

Implications for research

Large, prospective, well-designed studies that recruit people with suspicion of plague with clear inclusion and exclusion criteria will help elucidate the true accuracy of F1RDT for the detection and diagnosis of plague. False-positive rate is likely to have been overestimated due to the imperfect reference standard for plague. Research for clarification of the true false-positive rate of F1RDT for both pneumonic and bubonic plague is required. Clarification of estimate of pretest probability of plague for different case scenario, for RDT and other diagnostic methods will also be very valuable when determining the implication of F1RDT in practice. Research for new diagnostic techniques that do not require sputum would help overcome the limitations arising from poor-quality samples.

ACKNOWLEDGEMENTS

The CIDG Academic Editor is Professor Mical Paul, and DTA Editor is Dr Mia Schmidt-Hansen.

We are grateful to Vittoria Lutje, the Information Specialist of the Cochrane Infectious Diseases Group (CIDG), for her help with the literature search strategy, and to the Diagnostic Test Accuracy (DTA) editorial team for their input. We acknowledge Minoarisoa Rajerison and Jeanine Petersen for their expertise and support, and for providing unpublished data. We are grateful to Xin Wang for her help in screening Chinese manuscripts. We thank Paul Garner, CIDG Co-ordinating Editor, for his feedback and for contributing to the assessment of risk of bias.

Marty Chaplin and the CIDG editorial base are funded by UK aid from the UK government for the benefit of low- and middle-income countries (project number 300342-104). The views expressed do not necessarily reflect the UK government's official policies.

REFERENCES

References to studies included in this review

Andrianaivoarimanana 2019 *{published data only}*

Andrianaivoarimanana V, Piola P, Wagner DM, Rakotomanana F, Maheriniaina V, Andrianalimanana S, et al. Trends of human plague, Madagascar, 1998–2016. *Emerging Infectious Diseases* 2019;**25**(2):220-8.

Bertherat 2011 *{published data only}*

Bertherat E, Lamine KM, Formenty P, Thullier P, Mondonge V, Mitifu A, et al. Major pulmonary plague outbreak in a mining camp in the Democratic Republic of Congo: brutal awakening of an old scourge [Epidémie de peste pulmonaire dans un camp minier de la République Démocratique du Congo: Le réveil brutal d'un vieux fléau]. *Medecine Tropicale* 2005;**65**(6):511-4.

* Bertherat E, Thullier P, Shako JC, England K, Kone ML, Arntzen L, et al. Lessons learned about pneumonic plague diagnosis from two outbreaks, Democratic Republic of the Congo. *Emerging Infectious Diseases* 2011;**17**(5):778-84.

Chanteau 2003a *{published data only}*

Chanteau S, Rahalison L, Ralafiarisoa L, Foulon J, Ratsitorahina M, Ratsifasoamanana L, et al. Development and testing of a rapid diagnostic test for bubonic and pneumonic plague. *Lancet* 2003;**361**(9353):211-6.

Petersen 2018 *{unpublished data only}*

Petersen J. Status of plague diagnostics for vaccine evaluation. www.who.int/blueprint/what/norms-standards/Plague_vaccines_workshop-23-april-2018/en/ (accessed 28 August 2019).

Rajerison 2020 *{published and unpublished data}*

Rajerison M, Melocco M, Andrianaivoarimanana V, Rahajandraibe S, Rakotoarimanana F, Spiegel A, et al. Performance of plague rapid diagnostic test compared to bacteriology: a retrospective analysis of the data collected in Madagascar. *BMC Infectious Diseases* 2020;**20**:90.

Rajerison 2020 (2002-2007 Bubonic) *{published and unpublished data}*

Rajerison M, Melocco M, Andrianaivoarimanana V, Rahajandraibe S, Rakotoarimanana F, Spiegel A, et al. Performance of plague rapid diagnostic test compared to bacteriology: a retrospective analysis of the data collected in Madagascar. *BMC Infectious Diseases* 2020;**20**:90.

Rajerison 2020 (2002-2007 Pneumonic) *{published and unpublished data}*

Rajerison M, Melocco M, Andrianaivoarimanana V, Rahajandraibe S, Rakotoarimanana F, Spiegel A, et al. Performance of plague rapid diagnostic test compared to bacteriology: a retrospective analysis of the data collected in Madagascar. *BMC Infectious Diseases* 2020;**20**:90.

Rajerison 2020 (2017-2018 Bubonic) *{published and unpublished data}*

Rajerison M, Melocco M, Andrianaivoarimanana V, Rahajandraibe S, Rakotoarimanana F, Spiegel A, et al. Performance of plague rapid diagnostic test compared to bacteriology: a retrospective analysis of the data collected in Madagascar. *BMC Infectious Diseases* 2020;**20**:90.

Rajerison 2020 (2017-2018 Pneumonic) *{published and unpublished data}*

Rajerison M, Melocco M, Andrianaivoarimanana V, Rahajandraibe S, Rakotoarimanana F, Spiegel A, et al. Performance of plague rapid diagnostic test compared to bacteriology: a retrospective analysis of the data collected in Madagascar. *BMC Infectious Diseases* 2020;**20**:90.

Rajerison 2020 (2018-2019 Bubonic) *{published and unpublished data}*

Rajerison M, Melocco M, Andrianaivoarimanana V, Rahajandraibe S, Rakotoarimanana F, Spiegel A, et al. Performance of plague rapid diagnostic test compared to bacteriology: a retrospective analysis of the data collected in Madagascar. *BMC Infectious Diseases* 2020;**20**:90.

Rajerison 2020 (2018-2019 Pneumonic) *{published and unpublished data}*

Rajerison M, Melocco M, Andrianaivoarimanana V, Rahajandraibe S, Rakotoarimanana F, Spiegel A, et al. Performance of plague rapid diagnostic test compared to bacteriology: a retrospective analysis of the data collected in Madagascar. *BMC Infectious Diseases* 2020;**20**:90.

Richard 2015 *{published data only}*

Richard V, Riehm JM, Herindrainy P, Soanandrasana R, Ratsitoharina M, Rakotomanana F, et al. Pneumonic plague outbreak, northern Madagascar, 2011. *Emerging infectious diseases* 2015;**21**(1):8-15.

Riehm 2011 *{published data only}*

Riehm JM, Rahalison L, Scholz HC, Thoma B, Pfeffer M, Razanakoto LM, et al. Detection of Yersinia pestis using real-time PCR in patients with suspected bubonic plague. *Molecular and Cellular Probes* 2011;**25**(1):8-12.

References to studies excluded from this review

Abedi 2018 *{published data only}*

Abedi AA, Shako JC, Gaudart J, Sudre B, Ilunga BK, Shamamba SK, et al. Ecologic features of plague outbreak areas, Democratic Republic of the Congo, 1004–2014. *Emerging Infectious Diseases* 2018;**24**(2):210-20.

Anish 2013 *{published data only}*

Anish C, Guo X, Wahlbrink A, Seeberger PH. Detection by anti-carbohydrate antibodies. *Angewandte Chemie (International Ed. In English)* 2013;**52**:9524-8.

Asaku 2014 *{published data only}*

Asaku ST, Apangu T, Mead P. Surveillance and response: integrated indigenous health systems-based sentinel detection of human plague in a plague endemic West Nile region of Uganda (2008–2013). *International Journal of Infectious Diseases* 2014;**1**:141.

Bertherat 2007 *{published data only}*

Bertherat E, Bekhoucha S, Chougrani S, Razik F, Duchemin JB, Houti L, et al. Plague reappearance in Algeria after 50 years, 2003. *Emerging Infectious Diseases* 2007;**13**(10):1459-62.

Bosch unpublished *{unpublished data only}*

Bosch Q, Andrianaivoarimanana V, Ramasindrazana B, Mikaty G, Rakotonanahary RJ, Nikolay B, et al. A framework to assess and optimize the use of diagnostics during epidemics: application to the 2017 plague outbreak in Madagascar. Data on file .

Cabanel 2013 *{published data only}*

Cabanel N, Leclercq A, Chenal-Francisque V, Annajar B, Rajerison M, Bekkhoucha S, et al. Plague outbreak in Libya, 2009, unrelated to Plague in Algeria. *Emerging Infectious Diseases* 2013;**19**(2):230-6.

Chanteau 2000a *{published data only}*

Chanteau S, Rahalison L, Ratsitorahina M, Mahafaly, Rasolomaharo M, Boisier P, et al. Early diagnosis of bubonic plague using F1 antigen capture ELISA assay and rapid immunogold dipstick. *International Journal of Medical Microbiology* 2000;**290**(3):279-83.

Chanteau 2000b *{published data only}*

Chanteau S, Ratsitorahina M, Rahalison L, Rasoamanana B, Chan F, Boisier P, et al. Current epidemiology of human plague in Madagascar. *Microbes and Infection* 2000;**2**(1):25-31.

Chanteau 2003b *{published data only}*

Chanteau S, Nato F, Migliani R. Importance of rapid immunochromatography tests for the monitoring of characteristic diseases epidemic in developing countries: the example of the plague in Madagascar. *Medecine Tropicale* 2003;**63**(6):574-6.

Chanteau 2005 *{published data only}*

Chanteau S, Nato F. Rapid diagnosis of outbreak-prone bacterial diseases. *Medecine et Maladies Infectieuses* 2005;**35 Suppl 2**:S100-2.

Choi 2017 *{published data only}*

Choi N, Lee J, Ko J, Jeon JH, Rhie GE, deMello AJ, et al. Integrated SERS-based microdroplet platform for the automated immunoassay of F1 antigens in Yersinia pestis. *Analytical Chemistry* 2017;**89**(16):8413-20.

da Silva 2012 *{published data only}*

da Silva LC, Vasconcelos SS, Mendes-Marques CL, de Almeida AM, Leal NC. Plague diagnosis STNPCR and MPCR kits assembly, reaction reproducibility, and reagent stability. *Advances in Experimental Medicine and Biology* 2012;**954**:143-7.

Dennis 2003 *{published data only}*

Dennis DT, Chu MC. A major new test for plague. *Lancet* 2003;**361**(9353):191-2.

Goel 2015 *{published data only}*

Goel S, Raj Kessar R, Raj S, Raj P, Singh A. Relevance of rapid diagnostic tests for plague in the field settings: a review. *International Journal of Contemporary Microbiology* 2015;**1**(1):156-64.

Hai 2007 *{published data only}*

Hai R. Development and application of plague diagnostic technique in China. *Chinese Journal of Endemiology* 2007;**26**(5):473-4.

Migliani 2006 *{published data only}*

Migliani R, Chanteau S, Rahalison L, Ratsitorahina M, Boutin JP, Ratsifasoamanana L, et al. Epidemiological trends for human plague in Madagascar during the second half of the 20th century: a survey of 20,900 notified cases. *Tropical Medicine & International Health* 2006;**11**(8):1228-37.

MMWR 2009 *{published data only}*

Morbidity and Mortality Weekly Report. Bubonic and pneumonic plague – Uganda, 2006. MMWR Morbidity and Mortality Weekly Report 2009;**58**(28):778-81.

Ramasindrazana 2017 *{published data only}*

Ramasindrazana B, Andrianaivoarimanana V, Rakotondramanga JM, Birdsell DN, Ratsitorahina M, Rajerison M. Pneumonic plague transmission, Moramanga, Madagascar, 2015. *Emerging Infectious Diseases* 2017;**23**(3):521-4.

Randremanana 2019 *{published data only}*

Randremanana R, Andrianaivoarimanana V, Nikolay B, Ramasindrazana B, Paireau J, ten Bosch QA, et al. Epidemiological characteristics of an urban plague epidemic in Madagascar, August–November, 2017: an outbreak report. *Lancet Infectious Diseases* 2019;**19**:537-45.

Ratsitorahina 2000 *{published data only}*

Ratsitorahina M, Chanteau S, Rahalison L, Ratsifasoamanana L, Boisier P. Epidemiological and diagnostic aspects of the outbreak of pneumonic plague in Madagascar. *Lancet* 2000;**355**(9198):111-3.

Simon 2013 *{published data only}*

Simon S, Demeure C, Lamourette P, Filali S, Plaisance M, Creminon C, et al. Fast and simple detection of Yersinia pestis applicable to field investigation of plague foci. *PloS One* 2013;**8**(1):e54947.

Splettstoesser 2004 *{published data only}*

Splettstoesser WD, Rahalison L, Grunow R, Neubauer H, Chanteau S. Evaluation of a standardized F1 capsular antigen capture ELISA test kit for the rapid diagnosis of plague. *FEMS Immunology and Medical Microbiology* 2004;**41**(2):149-55.

Tomaso 2007 *{published data only}*

Tomaso H, Thullier P, Seibold E, Guglielmo V, Buckendahl A, Rahalison L, et al. Comparison of hand-held test kits, immunofluorescence microscopy, enzyme-linked immunosorbent assay, and flow cytometric analysis for rapid presumptive identification of Yersinia pestis. *Journal of Clinical Microbiology* 2007;**45**(10):3404-7.

Tsui 2015 *{published data only}*

Tsui PY, Tsai HP, Chiao DJ, Liu CC, Shyu RH. Rapid detection of Yersinia pestis recombinant fraction 1 capsular antigen. *Applied Microbiology and Biotechnology* 2015;**99**(18):7781-9.

UCLA 2003 *{published data only}*

UCLA. Simple test found to quickly spot plague infection. *Clinical Infectious Diseases* 2003;**36**(5):i.

Wang 2005 *{published data only}*

Wang P, Guo Y, Shi LY, Dong XQ, Yu GL. Labeling monoclonal antibodies to the F1 antigen of Yersinia pestis with colloidal gold particles and testing F1 antigen by dot immunogold filtration assay. *Chinese Journal of Endemiology* 2005;**24**(5):479-81.

Xu 2008 *{published data only}*

Xu DL, Wei SZ, Shi XM, Zhang ZB, Cai H, Tang YJ, et al. Estimating rapid detection of F1 antigen of Yersinia pestis with gold-immunochromatography. *Chinese Journal of Endemiology* 2008;**27**(3):326-8.

Yan 2006 *{published data only}*

Yan ZQ, Zhou L, Zhao YK, Wang J, Huang LH, Hu KX, et al. Rapid quantitative detection of Yersinia pestis by lateral-flow immunoassay and up-converting phosphor technology-based biosensor. *Sensors and Actuators. B, Chemical* 2006;**119**(2):656-63.

Yang 2006 *{published data only}*

Yang XY, Li M, Wang L, Mu Y, Xie H, Wang M, et al. The application of gold immunochromatographic assay for a rapid diagnosis for Yersinia pestis. *Chinese Journal of Endemiology* 2006;**25**(5):568-70.

Yao 2013 *{published data only}*

Yao MQ, Shi ML, Lei G, Turdi R, Xia XW, Jia HM, et al. Epidemiological investigation of human plague in the southern Junggar Basin: a field survey. *Chinese Journal of Endemiology* 2013;**32**(6):642-4.

Zasada 2015 *{published data only}*

Zasada AA, Forminska K, Zacharczuk K, Jacob D, Grunow R. Comparison of eleven commercially available rapid tests for detection of Bacillus anthracis, Francisella tularensis and Yersinia pestis. *Letters in Applied Microbiology* 2015;**60**(5):409-13.

Zasada 2018 *{published data only}*

Zasada AA, Zacharczuk K, Forminska K, Wiatrzyk A, Ziolkowski R, Malinowska E. Isothermal DNA amplification combined with lateral flow dipsticks for detection of biothreat agents. *Analytical Biochemistry* 2018;**560**:60-6.

Zhang 2014 *{published data only}*

Zhang P, Liu X, Wang C, Zhao Y, Hua F, Li C, et al. Evaluation of up-converting phosphor technology-based lateral flow strips for rapid detection of Bacillus anthracis Spore, Brucella spp., and Yersinia pestis. *PloS One* 2014;**9**(8):e105305.

Zhu 2006 *{published data only}*

Zhu H, Zhang CH, Tan H, He J, Zhao B, Duan Q. Development of an immunochromatography assay method for the detection of Yersinia pestis. *Chinese Journal of Epidemiology* 2006;**27**(3):253-5.

Additional references

CDC 2019

Centers for Disease Control and Prevention. Plague. www.cdc.gov/plague/index.html (accessed 19 April 2019).

CIDRAP 2013

Center for Infectious Disease Research and Policy. Plague. Agent and pathogenesis. 2013. www.cidrap.umn.edu/infectious-disease-topics/plague (accessed 15 April 2019).

GRADE Handbook 2013

Schünemann H, Brożek J, Guyatt G, Oxman A, editor(s). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. (updated October 2013). GRADE Working Group, 2013. Available from gdt.guidelinedevelopment.org/app/handbook/handbook.html (accessed 17 May 2019).

GRADEpro GDT 2015 [Computer program]

McMaster University (developed by Evidence Prime) GRADEpro GDT. Version accessed 17 May 2019. Hamilton (ON): McMaster University (developed by Evidence Prime), 2015.Available at gradepro.org.

Hsu 2018

Hsu HL, Chuang CC, Liang CC, Chiao DJ, Wu HL, Wu YP, et al. Rapid and sensitive detection of Yersinia pestis by lateral-flow assay in simulated clinical samples. *BMC Infectious Diseases* 2018;**18**:402.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: srdta.cochrane.org/.

Mavalankar 1995

Mavalankar DV. Indian 'plague' epidemic: unanswered questions and key lessons. *Journal of the Royal Society of Medicine* 1995;**88**:547-51.

Mead 2018

Mead PS. Plague in Madagascar – a tragic opportunity for improving public health. *New England Journal of Medicine* 2018;**378**(2):106-8.

New Horizons 2019

New Horizons. Yersinis pestis (F1). www.nhdiag.com/ypestis.shtml (accessed 22 August 2019).

Prentice 2007

Prentice MB, Rahalison L. Plague. *Lancet* 2007;**369**:1196-207.

Rasmussen 2015

Rasmussen S, Allentoft ME, Nielsen K, Nielsen R, Kristiansen K, Willerslev E. Early divergent strains of Yersinia pestis in Eurasia article early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell* 2015;**163**:571-82.

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Schünemann 2020

Schünemann HJ, Mustafa R, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 2. Inconsistency, Imprecision, publication bias and other domains for rating the certainty of evidence for test accuracy and presenting it in evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 2020;**122**:142-52. [DOI: DOI: 10.1016/j.jclinepi.2019.12.021]

Stata 2015 [Computer program]

Stata. Version 14. College Station (TX): StataCorp, 2015.Available at www.stata.com.

Takwoingi 2017

Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Statistical Methods in Medical Research* 2017;**26**(4):1896-911.

van Hoving 2019

van Hoving DJ, Griesel R, Meintjes G, Takwoingi Y, Maartens G, Ochodo EA. Abdominal ultrasound for diagnosing abdominal tuberculosis or disseminated tuberculosis with abdominal involvement in HIV-positive individuals. *Cochrane Database of Systematic Reviews* 2019, Issue 9. [DOI: 10.1002/14651858.CD012777]

Weniger 1984

Weniger BG, Warren AJ, Forseth V, Creelman T, Gorton J, Barnes AM. Human bubonic plague transmitted by a domestic cat scratch. *JAMA* 1984;**251**(7):927-8.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36. [DOI: 10.7326/0003-4819-155-8-201110180-00009]

WHO 2009

World Health Organization SEARO . Operational guidelines on plague surveillance, diagnosis, prevention and control. 2009. apps.who.int/iris/handle/10665/205593 (accessed prior to 3 June 2020).

WHO 2016

World Health Organization. Plague around the world, 2010–2015. Weekly epidemiological record 2016;**91**(8):89-93.

WHO 2019a

World Health Organization. Plague. www.who.int/csr/disease/plague/en/ (accessed 1 February 2019).

WHO 2019b

World Health Organization. Rapid diagnostic tests. www.who.int/malaria/areas/diagnosis/rapid_diagnostic_tests/en/ (accessed 22 August 2019).

References to other published versions of this review

Jullien 2019

Jullien S, Dissanayake HA, Chaplin M. Rapid diagnostic tests for plague. *Cochrane Database of Systematic Reviews* 2019, Issue 10. [DOI: 10.1002/14651858.CD013459]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Andrianaivoarimanana 2019

Study characteristics	
Patient Sampling	Retrospective study. The study authors included the cases of suspected plague that were declared to the national surveillance system. As such, sampling was not consecutive or random.
Patient characteristics and setting	**Country**: Madagascar. **Plague endemicity**: endemic to the central and northern highlands.

Andrianaivoarimanana 2019 *(Continued)*

	Clinical setting: cases of plague from all over the country. **Study dates**: 1998–2016, but cases with RDT compared to reference standard only from 2002 to 2007. **Inclusion criteria**: people with suspected plague declared to the national surveillance. **Exclusion criteria**: none mentioned. **Target condition**: any form of plague. **Sample size**: 4221 (from 2002 to 2007), including 411 with unknown RDT findings. **Samples collected**: not reported. **Age (median)**: for all participants included from 1998 to 2016: Bubonic plague: 11 years (IQR 6 to 20) for suspected case; 13 years (IQR 8 to 24) for confirmed and presumptive cases. Pneumonic plague: 26 years (IQR 17 to 40) for suspected case; 29 years (IQR 20 to 42) for confirmed and presumptive cases. **Sex ratio (M:F)**: for all participants included from 1998 to 2016: Bubonic plague: 1.44 for suspected cases; 1.38 for confirmed and presumptive cases. Pneumonic plague: 1.08 for suspected cases; 1.28 for confirmed and presumptive cases. **Signs and symptoms presented**: not reported. **Antibiotic treatment prior to enrolment**: not quantified. Authors stated that, "negative culture results might have resulted from administration of antimicrobial drugs by local health officials before sampling, whereas F1RDT results remained positive >3 weeks after treatment initiation."
Index tests	**Type**: RDT produced at the IPM. **Brand name**: not specified. **Threshold for positive result**: not specified in the paper. However, this is the same RDT described in Chanteau 2003a, with a lower detection threshold of 0.5 ng/mL, and used as a positive/negative test. **Place where the test was performed**: not reported. **Transport and storage conditions of the RDT**: not reported. **Transport and storage conditions of the samples to be tested**: not reported. **Need for sample preparation**: not reported. **Who performed the test**: not reported. **Blinding of operator to the results of the reference standard**: not reported. **Special training provided to personnel performing the test**: not reported.
Target condition and reference standard(s)	**Definitions of cases of plague**: Confirmed: *Y pestis* isolated by culture or mouse inoculation. Presumptive: no isolation of Y pestis but F1RDT or microscopy positive. Suspected: no samples available for testing or all test results negative.

Andrianaivoarimanana 2019 *(Continued)*

	Reference standard(s) used in this review: isolation of *Y pestis* by culture (and microscopy, not considered for the purpose of this review). **Place where the reference standard(s) was (were) performed**: Central Laboratory for Plague of the Malagasy Ministry of Health, hosted at the IPM. **Who performed the reference standard(s)**: not reported. **Blinding of operator to RDT result**: not reported. However, there is low risk of introducing bias in the interpretation of the reference standard (objective finding) by knowing the RDT result.
Flow and timing	**Time between collection of sample for RDT and reference standard(s)**: not reported. **Administration of antibiotics between sample collection for index test and reference standard**: not reported. **Time of sample transportation**: not reported. We judged risk of bias for the RDTs performed between 2002 and 2007.
Comparative	
Notes	**Source of funding**: not reported.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it prespecified?	Yes		

Andrianaivoarimanana 2019 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Bertherat 2011

Study characteristics	
Patient Sampling	Retrospective study. The study authors included all the cases with suspicion of plague during the outbreaks, following the WHO definition and based on medical records.

Bertherat 2011 *(Continued)*

Patient characteristics and setting	**Country**: Democratic Republic of the Congo. **Plague endemicity**: endemic. The country "has the most active focus of plague worldwide. In the north-eastern region of Ituri, >1000 suspected cases are reported each year." (quote from manuscript) **Clinical setting**: the 2005 outbreak occurred "in a diamond mining camp in a remote area of the Oriental Province" where "no previous cases of plague had been reported in this region 25 km from the village of Zobia, Bas-Uele." The 2006 outbreak "occurred in a gold mining camp 200 km from the Zobia camp, near Bolebole, Haut Uele." (quotes from manuscript) **Study dates**: 2 outbreaks reported, from 15 December 2004 to 11 March 2005, and from 21 August to 21 October 2006. **Inclusion criteria**: suspected cases of plague following WHO definition based on registries and not following reported cases during the outbreaks. **Exclusion criteria**: none mentioned. **Target condition**: any form of plague, but outbreaks of pneumonic plague. During the first outbreak, 2 cases of septicaemic cases were mentioned. **Sample size**: outbreak 2005: 130 cases (5 confirmed, 10 probable, 115 suspected), biological samples collected for 87 cases (173 specimens collected according to Bertherat 2005), RDT performed in 37 cases (35 against culture, 2 against PCR). Outbreak 2006: 162 cases (23 confirmed, 22 probable, 117 suspected), biological samples collected for 117 cases, RDT performed in 96 cases (against culture). **Samples collected**: sputum **Age**: not reported. **Sex**: not reported. **Signs and symptoms presented**: not reported. **Antibiotic treatment prior to enrolment**: mentioned but not quantified.
Index tests	**Type**: RDT used included in sampling kits from the IPM. The authors referenced Chanteau 2003a for the RDT used, produced at the IPM. **Brand name**: RDT developed by the Naval Medical Research Institute (Bethesda, MD, USA), as described in Chanteau 2003a. **Threshold for positive result**: not stated in the paper, but 0.5 ng/mL as specified in the referenced study of Chanteau 2003a, and used as a positive/negative test. **Place where the test was performed**: on site for both outbreaks. Repeated at IPM for both outbreaks, and in Kinsangani during 2005 outbreak. **Transport and storage conditions of the RDT**: dipstick contained in an individually vacuum-sealed package with desiccant to maintain stability and sterility. **Transport and storage conditions of the samples to be tested**: for RDT performed on site, no sample transport required as performed on site immediately during both outbreaks. For culture and other tests, sputum samples were stored with a swab in Cary Blair medium at 4 °C. During the 2005 outbreak, samples were first sent to the Kisangani laboratory. The remaining aliquot was stored at 4 °C (serum) and 28 °C to 30 °C (sputum) and sent to Kinshasa and then Madagascar, South Africa, France, and Germany. During the 2006 outbreak, samples were directly sent to the IPM. **Need for sample preparation**: sputum (0.5 mL) was diluted and homogenized in 1 mL of phosphate-buffered saline using a sterile syringe. 200 µL was then placed in a sterile tube test to apply the dipstick. **Who performed the test**: a trained biologist from the National Plague Reference Laboratory collected the sputum sample and performed the RDT.

Bertherat 2011 *(Continued)*

	Blinding of operator to the results of the reference standard: not reported, but probably yes as test was performed on site, while reference standard was performed elsewhere. **Special training provided to personnel performing the test**: "trained" biologist with no details on the training received. Although the test was performed by a trained biologist who also collected the sputum, we judged that there would probably be low concern of applicability in the field if minimum training was provided to health workers.
Target condition and reference standard(s)	**Definitions of cases of plague**: using WHO definitions but not clearly stated. **Reference standard(s) used in this review**: Culture + standard biochemical tests and phage lysis RT-PCR (only for the 2005 outbreak) ELISA for serology from paired samples, collected 10 days apart, a 4-fold increase in antibody titre considered positive. (Microscopy and direct fluorescent antibody staining from sputum during the 2005 outbreak were also used, but not for the purpose of this review.) **Place where the reference standard(s) was (were) performed**: Culture and microscopy: at a temporary second-line regional laboratory in Kisangani, which is 2 hours by air from the outbreak area, and at IPM for the 2005 outbreak; only at IPM for the 2006 outbreak. RT-PCR: in Germany for the 2005 outbreak. ELISA: at Centre de Recherche du Service de Santé des Armées in Grenoble, France. **Who performed the reference standard(s)**: a biologist from the IPM. **Blinding of operator to RDT result**: not reported. However, there is low risk of introducing bias in the interpretation of the reference standard (objective finding) by knowing the RDT result.
Flow and timing	**Time between collection of sample for RDT and reference standard(s)**: both RDT and reference standard (culture and RT-PCR) were conducted from the same biological sample. **Administration of antibiotics between sample collection for index test and reference standard**: performed on same samples. **Time of sample transportation**: shipments from the affected area to Madagascar took 8–40 days (median 18.5 days) during the 2005 outbreak, and 4–5 days during the 2006 outbreak.
Comparative	
Notes	**Source of funding**: not reported.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Bertherat 2011 *(Continued)*

Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condi-			Low concern

Bertherat 2011 *(Continued)*

tion as defined by the reference standard does not match the question?		
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	No	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	No	
Did all patients receive a reference standard?	Yes	
Could the patient flow have introduced bias?		High risk

Chanteau 2003a

Study characteristics	
Patient Sampling	Prospective study as part of the national plague control programme. The study authors included the cases of suspected plague that were declared to the national surveillance system, but sampling was unclear, and samples included postmortem tissues.
Patient characteristics and setting	**Country**: Madagascar. **Plague endemicity**: endemic areas. **Clinical setting**: central laboratory in the capital Antananarivo; and 26 pilot sites (6 district hospitals and 20 healthcare centres) in remote areas. **Study dates**: 1 December 2000 to 30 May 2001 in the capital; 1 December 2000 to 25 January 2001 to test RDT in remote areas. **Inclusion criteria**: people with suspected plague. **Exclusion criteria**: none mentioned. **Target condition**: any form of plague, although only bubonic and pneumonic plague are mentioned. **Sample size**: 671 suspected cases from national surveillance system (central laboratory only) + 128 participants from the remote sites (RDT performed both at bedside of the participant and at central laboratory). **Samples collected**: samples sent to the central laboratory: 691 clinical samples from 671 participants (bubonic aspirate 643 (93%); sputum 13 (2%); postmortem lung or liver puncture 35 (5%)). Samples from remote sites: 128 cases with number of samples not specified (bubonic aspirate 123 (96%); sputum: 5 (4%)).

Chanteau 2003a *(Continued)*

	Although 5% of the samples were postmortem tissues, raising questions about applicability, it was a small number and, therefore, we judged applicability as low concern. **Age**: not reported. **Sex**: not reported. **Signs and symptoms presented**: not reported. **Antibiotic treatment prior to enrolment**: not reported.
Index tests	**Type**: RDT produced at the IPM using a combination of B18–1 and G6–18 antibodies. **Brand name**: RDT developed by the Naval Medical Research Institute (Bethesda, MD, USA). **Threshold for positive result**: the lower detection threshold of the RDT was 0.5 ng/mL, the test was used as a positive/negative test. **Place where the test was performed**: at the participant's bedside; repeated at the central laboratory. **Transport and storage conditions of the RDT**: immunostrips trimmed to a width of 5 mm and stored in a waterproof bag at 4 °C, or in 5 mL disposable plastic tubes at room temperature at non-central sites. The RDTs were stored at room temperature during the study (20–30 °C). **Transport and storage conditions of the samples to be tested**: samples collected in remote areas were sent on a swab in Cary Blair agar, at room temperature, to the central laboratory at the IPM. **Need for sample preparation**: on arrival at the central laboratory, the specimen was washed out of the swab by incubation in 1 mL phosphate-buffered saline, and was tested with the 2 reference methods and with RDT. Sputum samples were diluted with saline or phosphate-buffered saline (1 in 2 to 1 in 10 dilution). **Who performed the test**: at the remote sites, 29 medical doctors, 19 nurses, and 9 health workers. **Blinding of operator to the results of the reference standard**: at the remote sites, blinded as RDT performed before sending the sample for reference standard. At central laboratory, blinded as RDT performed before availability of results from reference standard. **Special training provided to personnel performing the test**: at the remote sites, onsite training for 3 hours as to how to obtain clinical samples and to use, read, and archive the dipsticks. An illustrated instruction guidebook, in French and Malagasy, was given to each centre.
Target condition and reference standard(s)	**Definitions of cases of plague**: Confirmed: culture positive Presumptive: culture negative but microscopy positive Negative: culture and microscopy negative **Reference standard(s) used in this review**: Gram staining and isolation of *Y pestis* (either directly from the participants' samples or after mouse inoculation). (Immunocapture ELISA for detection of F1 Ag was also used, but not included as reference standard in this review.) **Place where the reference standard(s) was (were) performed**: central laboratory at the IPM. **Who performed the reference standard(s)**: "Skilled technicians." **Blinding of operator to RDT result**: blinded for RDT results from the field, but not blinded for the RDT results conducted at the central laboratory (communication with the authors). There is low risk of introducing bias in the interpretation of the reference standard (objective finding) by knowing the RDT result.

Chanteau 2003a *(Continued)*

Flow and timing	**Time between collection of sample for RDT and reference standard(s)**: the same sample was used for index test and reference standard for participants from national surveillance; unclear for the remote site samples. **Administration of antibiotics between sample collection for index test and reference standard**: administration of antibiotics was not reported; however, at the central laboratory, both RDT and reference standards were performed from the same sample. **Time of sample transportation**: the median transport time to the central laboratory of the 691 samples was 8 days, ranging from 0 to 66 days (25th percentile 5 days, 75th percentile 14 days). However, there was no significant difference in the mean transport time between bacteriologically confirmed and negative specimens.
Comparative	
Notes	**Source of funding**: Institut Pasteur, Paris (Grant PTR 2000–11), IPM, and the Ministry of Health of Madagascar (World Bank IDA 3302 MAG). Sponsors had no role in study design, data collection, data analysis, data interpretation, or writing of report.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Chanteau 2003a *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
Could the patient flow have introduced bias?		High risk	

Petersen 2018

Study characteristics	
Patient Sampling	Prospective study. Sampling not described.
Patient characteristics and setting	**Country**: Madagascar and Uganda. **Plague endemicity**: endemic areas. **Clinical setting**: 35 clinics in Madagascar and 12 clinics in Uganda. **Study dates**: 2004–2006 in Madagascar, 2004–2017 in Uganda. **Inclusion criteria**: not reported. **Exclusion criteria**: none mentioned. **Target condition**: any form of plague. **Sample size**: ≥ 254. **Samples collected**: bubo aspirates and sputum. **Age**: not reported. **Sex**: not reported. **Signs and symptoms presented**: not reported. **Antibiotic treatment prior to enrolment**: not reported.
Index tests	**Type**: RDT produced by New Horizons. **Brand name**: not reported. **Threshold for positive result**: lower detection threshold of RDT was 1 ng/mL (personal communication), the test was used as semi-quantitative test (from 1+ to 4+) and considered positive for analysis from 1+ and from 2+. Those are the findings presented and we do not know if this was prespecified in the protocol and methods of the study. **Place where the test was performed**: central laboratory. **Transport and storage conditions of the RDT**: not reported. **Transport and storage conditions of the samples to be tested**: not reported. **Need for sample preparation**: not reported. **Who performed the test**: not reported. **Blinding of operator to the results of the reference standard**: not reported. **Special training provided to personnel performing the test**: not reported.
Target condition and reference standard(s)	**Definitions of cases of plague**: Confirmed case: 1 of the following criteria: isolate from a clinical source identified as *Y pestis* (colony morphology and 2/4 tests positive: bacteriophage lysis, F1 detection, PCR, *Y pestis* biochemical profile); with/without 4-fold difference in F1 antibody titre between paired serum samples; with/without F1 Ag detection (bubo, sputum, blood) by immunochromatography in endemic focus when no other confirmatory tests can be performed.

Petersen 2018 *(Continued)*

	Presumptive case: detection by 2 target RT-PCR. Suspect case: rapid onset of fever of at least 38 °C, and 1 of the following: ≥ 1 buboes, defined as a tender lymph node swelling > 1 cm in diameter, or clinical suspicion of pneumonic plague (e.g. prostration, cough, increased respiratory rate, haemoptysis, purulent sputum, or a combination of these), or clinical suspicion of plague and epidemiological link to other cases. **Reference standard(s) used in this review**: Culture Serological analysis with a 4-fold increase in titre in the second serum sample. **Place where the reference standard(s) was (were) performed**: culture at the central laboratory, paired serology at the US CDC. **Who performed the reference standard(s)**: not reported. **Blinding of operator to RDT result**: not reported. However, there is low risk of introducing bias in the interpretation of the reference standard (objective finding) by knowing the RDT result.
Flow and timing	**Time between collection of sample for RDT and reference standard(s)**: collected on same day. **Administration of antibiotics between sample collection for index test and reference standard**: unlikely although not reported. **Time of sample transportation**: not reported.
Comparative	
Notes	**Source of funding**: not reported. **Collaborators**: the CDC, Ministry of Health and IPM and Ministry of Health and Uganda Virus Research Institute from Uganda.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear

Petersen 2018 *(Continued)*

DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it prespecified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Did all patients receive a reference standard?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Rajerison 2020

Study characteristics	
Patient Sampling	Retrospective study. The study authors included cases with suspicion of plague who were declared to the national surveillance system. As such, sampling was not consecutive or random.
Patient characteristics and setting	**Country**: Madagascar. **Plague endemicity**: endemic. **Clinical setting**: surveillance data. **Study dates**: 3 study periods: Period 1: 2002–2007 for all clinical forms (same period as in Andrianaivoarimanana 2019, but excluding those with transport > 7 days and those who received antibiotics prior to sample collection). Period 2: 11 September to 3 October 2017 for all pulmonary plague cases from non-endemic zones. Period 3: 2018 to 3 April 2019 for all clinical forms. **Inclusion criteria**: all clinically suspected cases reported to the national surveillance system with RDT performed at the Central Laboratory for Plague, when RDT and culture were performed independently and systematically (periods when RDTs were used as a screening test for bacteriology testing were not considered). **Exclusion criteria**: cases for which time of transport > 7 days, and cases who have declared to have taken antibiotics prior to sample collection. **Target condition**: all forms of plague **Sample size**: after exclusion and after removing missing cases: Period 1: 2468 (2319 bubonic + 149 pulmonic cases) Period 2: 100 (33 + 67) Period 3: 192 (151 + 41) **Samples collected**: bubo aspirate, sputum, postmortem organ puncture. **Age**: not reported. **Sex**: not reported. **Signs and symptoms presented**: not reported. **Antibiotic treatment prior to enrolment**: none; this was an exclusion criteria.
Index tests	**Type**: RDT produced at the IPM. **Brand name**: not specified. **Threshold for positive result**: not specified in manuscript. However, this is the same RDT described in Chanteau 2003a, and therefore the threshold is 0.5 ng/mL, and used as a positive/negative test. **Place where the test was performed**: Central Laboratory for Plague at the IPM. **Transport and storage conditions of the RDT**: transport not required as RDT conducted where there are produced. Storage conditions not reported. **Transport and storage conditions of the samples to be tested**: not reported. **Need for sample preparation**: not reported.

Rajerison 2020 *(Continued)*

	Who performed the test: not clearly reported, but by trained staff at the central laboratory as study authors mentioned "RDT performed at Central Laboratory for Plague to avoid bias due to handling effect in Health Care Centred despite training." **Blinding of operator to the results of the reference standard**: not reported. **Special training provided to personnel performing the test**: not reported.
Target condition and reference standard(s)	**Definitions of cases of plague**: Confirmed: clinically suspected cases with positive RDT and positive molecular biology, or positive culture. Probable: clinically suspected cases with positive RDT or positive molecular biology and culture negative or not performed. Suspected: all clinically suspected plague cases that meet the clinical and epidemiological criteria as per WHO recommendations (compatible clinical presentation (fever, sepsis syndrome, lymphadenopathy, acute pneumonitis, or a combination of these) and epidemiological features (such as exposure to infected animals or humans, evidence of flea bites, residence in or travel to a known endemic focus within the previous 10 days, or a combination of these)). **Reference standard(s) used in this review**: 'bacteriology,' this is isolation of *Y pestis* by culture and or mouse inoculation (communication with authors). **Place where the reference standard(s) was (were) performed**: Central Laboratory for Plague at the IPM. **Who performed the reference standard(s)**: not reported. **Blinding of operator to RDT result**: not reported. However, there is low risk of introducing bias in the interpretation of the reference standard (objective finding) by knowing the RDT result.
Flow and timing	**Time between collection of sample for RDT and reference standard(s)**: not reported. **Administration of antibiotics between sample collection for index test and reference standard**: none, as per exclusion criteria. **Time of sample transportation**: ≤ 7 days, as per exclusion criteria.
Comparative	
Notes	**Source of funding**: financial support from the Wellcome Trust/Department for International Development (contract no 211309/Z/18/Z) to perform the analyses. Support USAID (Grant no AID-687-G-13-0003) for the implementation of the RT-PCR technique, secured transportation of the biological samples as well as financing for the additional human resources needed by the Central Laboratory for Plague and the Epidemiology Units at IPM. WHO provided funding for the acquisition of new equipment to accelerate RDT production during the 2017 epidemic. The funding source had no role in study design or collection, analysis, and interpretation of data.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Rajerison 2020 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference			Low concern

Rajerison 2020 *(Continued)*

standard does not match the question?				
DOMAIN 4: Flow and Timing				
Was there an appropriate interval between index test and reference standard?	Yes			
Did all patients receive the same reference standard?	Yes			
Were all patients included in the analysis?	Yes			
Did all patients receive a reference standard?	Yes			
Could the patient flow have introduced bias?		Low risk		

Rajerison 2020 (2002-2007 Bubonic)

Study characteristics				
Patient Sampling	See Rajerison 2020.			
Patient characteristics and setting	See Rajerison 2020.			
Index tests	See Rajerison 2020.			
Target condition and reference standard(s)	See Rajerison 2020.			
Flow and timing	See Rajerison 2020.			
Comparative				
Notes	See Rajerison 2020.			
Methodological quality				
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**	
DOMAIN 1: Patient Selection				
Was a consecutive or random sample of patients enrolled?	Unclear			
Was a case-control design avoided?	Unclear			
Did the study avoid inappropriate exclusions?	Unclear			
Did the study considered prior administration of antibiotics?	Unclear			
Could the selection of patients have introduced bias?		Unclear risk		

Rajerison 2020 (2002-2007 Bubonic) *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Did all patients receive a reference standard?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Rajerison 2020 (2002-2007 Pneumonic)

Study characteristics	
Patient Sampling	See Rajerison 2020.
Patient characteristics and setting	See Rajerison 2020.
Index tests	See Rajerison 2020.
Target condition and reference standard(s)	See Rajerison 2020.

Rajerison 2020 (2002-2007 Pneumonic) *(Continued)*

Flow and timing	See Rajerison 2020.
Comparative	
Notes	See Rajerison 2020.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Rajerison 2020 (2002-2007 Pneumonic) *(Continued)*

Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Did all patients receive a reference standard?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Rajerison 2020 (2017-2018 Bubonic)

Study characteristics			
Patient Sampling	See Rajerison 2020.		
Patient characteristics and setting	See Rajerison 2020.		
Index tests	See Rajerison 2020.		
Target condition and reference standard(s)	See Rajerison 2020.		
Flow and timing	See Rajerison 2020.		
Comparative			
Notes	See Rajerison 2020.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		

Rajerison 2020 (2017-2018 Bubonic) *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Did all patients receive a reference standard?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Rajerison 2020 (2017-2018 Pneumonic)

Study characteristics	
Patient Sampling	See Rajerison 2020.
Patient characteristics and setting	See Rajerison 2020.
Index tests	See Rajerison 2020.
Target condition and reference standard(s)	See Rajerison 2020.
Flow and timing	See Rajerison 2020.
Comparative	
Notes	See Rajerison 2020.
Methodological quality	

Rajerison 2020 (2017-2018 Pneumonic) *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Did all patients receive a reference standard?			
Could the patient flow have introduced bias?		Unclear risk	

Rajerison 2020 (2018-2019 Bubonic)

Study characteristics			
Patient Sampling	See Rajerison 2020.		
Patient characteristics and setting	See Rajerison 2020.		
Index tests	See Rajerison 2020.		
Target condition and reference standard(s)	See Rajerison 2020.		
Flow and timing	See Rajerison 2020.		
Comparative			
Notes	See Rajerison 2020.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Rajerison 2020 (2018-2019 Bubonic) *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Did all patients receive a reference standard?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Rajerison 2020 (2018-2019 Pneumonic)

Study characteristics			
Patient Sampling	See Rajerison 2020.		
Patient characteristics and setting	See Rajerison 2020.		
Index tests	See Rajerison 2020.		
Target condition and reference standard(s)	See Rajerison 2020.		
Flow and timing	See Rajerison 2020.		
Comparative			
Notes	See Rajerison 2020.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	Unclear		

Rajerison 2020 (2018-2019 Pneumonic) *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Did all patients receive a reference standard?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Richard 2015

Study characteristics	
Patient Sampling	Retrospective study. Study authors included all cases with suspicion of plague during outbreaks. Consequently, there was no consecutive or random sampling. Furthermore, RDT was only performed in 2 participants, as biological samples could not be retrieved for the first cases when plague was not suspect-

Richard 2015 *(Continued)*

	ed yet, and other cases were suspected on recovery, with no biological samples taken in acute phase.
Patient characteristics and setting	**Country**: Madagascar. **Plague endemicity**: endemic in Madagascar, although the remote region where the outbreak occurred was supposedly free of *Y pestis*. **Clinical setting**: 7 villages along a field path in the communes of Ambarakaraka and Anaborano, Ambilobe District, in the north of Madagascar. **Study dates**: outbreak in January 2011. **Inclusion criteria**: all participants with suspected plague during this outbreak. **Exclusion criteria**: none mentioned. **Target condition**: any form of plague, but outbreak of pneumonic plague. **Sample size**: 20 cases in the outbreak, but RDT was only performed on 2 participants. **Samples collected**: sputum. **Age**: not reported. **Sex**: not reported. **Signs and symptoms presented**: sudden onset of fever, cough, haemoptysis, and chest pain. **Antibiotic treatment prior to enrolment**: both participants received antibiotics, but unclear whether before or after biological sample collection. We judged there were unclear concerns regarding applicability, as the test was performed on only 2 people, with limited information.
Index tests	**Type**: the authors referenced Chanteau 2003a for the RDT used, produced at the IPM. **Brand name**: RDT developed by the Naval Medical Research Institute (Bethesda, MD, USA), as described in Chanteau 2003a. **Threshold for positive result**: not stated in the paper, but Ag threshold of 0.5 ng/mL as specified in the referenced study Chanteau 2003a, and used as a positive/negative test. **Place where the test was performed**: unclear, probably bedside of the participant. **Transport and storage conditions of the RDT**: not reported. **Transport and storage conditions of the samples to be tested**: not reported. **Need for sample preparation**: not reported. **Who performed the test**: not reported. **Blinding of operator to the results of the reference standard**: not reported. It is likely that RDT was performed at the bedside of the participants, and, therefore, interpreted before sending sample for reference standard. **Special training provided to personnel performing the test**: not reported.
Target condition and reference standard(s)	**Definitions of cases of plague**: Confirmed: 4-fold increase in titre of antibodies against F1 Ag in paired serum samples or a positive culture. Presumptive: positive serological result for antibodies anti F1 Ag.

Richard 2015 *(Continued)*

	Suspected: specific clinical symptoms. **Reference standard(s) used in this review**: Culture + inoculation in mice. Serological analysis with a 4-fold increase in titre in the second serum sample. Molecular analysis: PCR (specific for the *Y. pestis* plasminogen activator and capsule Ag fraction 1 genes). **Place where the reference standard(s) was (were) performed**: WHO Collaborating Center for Plague at the IPM in Antananarivo. **Who performed the reference standard(s)**: not reported. **Blinding of operator to RDT result**: not reported. However, there is low risk of introducing bias in the interpretation of the reference standard (objective finding) by knowing the RDT result.
Flow and timing	**Time between collection of sample for RDT and reference standard(s)**: not reported. **Administration of antibiotics between sample collection for index test and reference standard**: not reported. **Time of sample transportation**: distance > 900 km, time of transportation not reported. The judgements we make below are considering only the 2 participants for whom RDT was performed.
Comparative	
Notes	**Source of funding**: supported by the IPM, the President's Malaria Initiative/US Agency for International Development, and the US Department of Homeland Security (project no. DHS-09-ST-108-001/MGN3EL7-01).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear

Richard 2015 *(Continued)*

DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Riehm 2011

Study characteristics	
Patient Sampling	Retrospective study. Study authors included all cases with symptoms consistent with the clinical diagnosis of bubonic plague during the study period. Consequently, there was no consecutive or random sampling.
Patient characteristics and setting	**Country**: Madagascar. **Plague endemicity**: endemic. **Clinical setting**: endemic plague areas in Madagascar. **Study dates**: February 2007 to December 2008. **Inclusion criteria**: suspicion of bubonic plague. **Exclusion criteria**: none mentioned. **Target condition**: bubonic plague. **Sample size**: 149. **Samples collected**: bubo aspirates. **Age**: range 1–72 years; mean 17 years. **Sex**: 59 (39.6%) females. **Signs and symptoms presented**: fever and lymphadenopathy. **Antibiotic treatment prior to enrolment**: not reported.
Index tests	**Type**: F1RDT produced at the IPM, not clearly mentioned by study authors but in references. **Brand name**: not reported. **Threshold for positive result**: not reported. Test was regarded positive if ≥ 1 of tests conducted either onsite or at the IPM was positive. **Place where the test was performed**: onsite instantly, and repeated at the IPM. **Transport and storage conditions of the RDT**: not reported. **Transport and storage conditions of the samples to be tested**: not required as sample tested instantly onsite. **Need for sample preparation**: obtention of bubo aspirate after injection of 1 mL sterile saline solution before treatment. **Who performed the test**: not reported. **Blinding of operator to the results of the reference standard**: not reported. **Special training provided to personnel performing the test**: not reported.
Target condition and reference standard(s)	**Definitions of cases of plague**: Confirmed: culture positive.

Riehm 2011 *(Continued)*

	Presumptive: F1-Ag immunochromatography positive or specific PCR assay positive, or both. Suspicion: clinical diagnosis. **Reference standard(s) used in this review**: Culture. RT-PCR. **Place where the reference standard(s) was (were) performed**: IPM in Antananarivo. **Who performed the reference standard(s)**: not reported. **Blinding of operator to RDT result**: not reported. However, there is low risk of introducing bias in the interpretation of the reference standard (objective finding) by knowing the RDT result.
Flow and timing	**Time between collection of sample for RDT and reference standard(s)**: not reported. **Administration of antibiotics between sample collection for index test and reference standard**: not reported. **Time of sample transportation**: due to unfavourable economic circumstances, the transport of samples to the central reference laboratory for human plague, Institut Pasteur Antananarivo, Madagascar, took several days or even weeks at ambient air temperature (5 °C to 30 °C).
Comparative	
Notes	**Source of funding**: not reported.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Did the study considered prior administration of antibiotics?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			

Riehm 2011 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Did all patients receive a reference standard?	Yes		
Could the patient flow have introduced bias?		High risk	

Ag: antigen; CDC: Centers for Disease Control and Prevention; ELISA: enzyme-linked immunosorbent assay; F: female; IQR: interquartile range; IPM: Institut Pasteur of Madagascar; M: male; PCR: polymerase chain reaction; RDT: rapid diagnostic test; RT-PCR: real-time polymerase chain reaction; USAID: United States Agency for International Development; WHO: World Health Organization.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Abedi 2018	Although this study fulfilled the inclusion criteria, it described 5 outbreaks including the 2 outbreaks described in Bertherat 2011, without adding any data. For the other 3 outbreaks, there are insufficient data to draw 2×2 tables.
Anish 2013	Ineligible index test (array).
Asaku 2014	Ineligible index test. Conference abstract that potentially met inclusion criteria. Contacted author to request full-text publications. 2 publications were sent with no data regarding plague RDT.
Bertherat 2007	Outbreak of bubonic plague in Algeria. Insufficient data to draw 2×2 tables.
Bosch unpublished	Ineligible study design (statistical modelling to estimate the performance of diagnostic tests for plague).
Cabanel 2013	Described a plague outbreak in Libya. Insufficient data to draw 2×2 tables.
Chanteau 2000a	Outdated RDT from Institut Pasteur of Madagascar.
Chanteau 2000b	Ineligible index test (rapid test based on ELISA).
Chanteau 2003b	Ineligible study design (narrative review).
Chanteau 2005	Ineligible study design (narrative review).
Choi 2017	Experimental setting (in vitro).
da Silva 2012	Experimental setting (in vitro).
Dennis 2003	Ineligible study design (commentary).
Goel 2015	Ineligible study design (narrative review).
Hai 2007	Ineligible study design (commentary).
Migliani 2006	Ineligible index test (no use of RDT during the reported outbreak).
MMWR 2009	Ineligible index test (no use of RDT during the reported outbreak).
Ramasindrazana 2017	Ineligible index test (no RDT was used in humans only in mice).
Randremanana 2019	People from an outbreak already included in Rajerison 2020, with no data that we could extract on RDT findings to draw a 2×2 table.
Ratsitorahina 2000	Outdated RDT from Institut Pasteur of Madagascar.
Simon 2013	Experimental setting (in vitro).
Splettstoesser 2004	Ineligible index test (rapid test based on ELISA).
Tomaso 2007	Experimental setting (in vitro).
Tsui 2015	Experimental setting (simulated samples of *Y pestis* within human serum).
UCLA 2003	Ineligible study design (commentary).

Study	Reason for exclusion
Wang 2005	Experimental setting (in vitro).
Xu 2008	Ineligible participants (animals).
Yan 2006	Ineligible participants (animals).
Yang 2006	Experimental setting, using animal samples and human samples (including postmortem tissues) to assess the efficacy of several diagnostic tests for plague.
Yao 2013	Ineligible index test (immunochromatography test based on the detection of F1 antibodies).
Zasada 2015	Experimental setting (in vitro).
Zasada 2018	Experimental setting (in vitro).
Zhang 2014	Experimental setting (in vitro).
Zhu 2006	Experimental setting (in vitro).

ELISA: enzyme-linked immunosorbent assay; RDT: rapid diagnostic test.

D A T A

Presented below are all the data for all of the tests entered into the review.

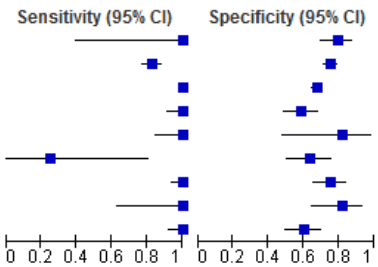
Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 F1RDT versus culture for all forms of plague, primary analysis	9	3696
2 F1RDT versus culture for all forms of plague, sensitivity analysis 1	8	5038
3 F1RDT versus culture for all forms of plague, sensitivity analysis 2	8	3629

Test 1. F1RDT versus culture for all forms of plague, primary analysis

F1RDT versus culture for all forms of plague, primary analysis

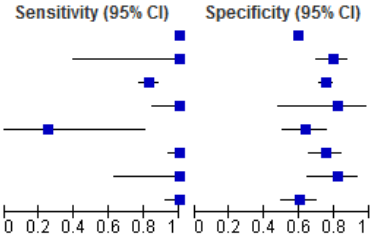
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bertherat 2011	4	19	0	73	1.00 [0.40, 1.00]	0.79 [0.70, 0.87]
Chanteau 2003a	151	128	31	381	0.83 [0.77, 0.88]	0.75 [0.71, 0.79]
Rajerison 2020 (2002-2007 Bubonic)	1328	324	0	667	1.00 [1.00, 1.00]	0.67 [0.64, 0.70]
Rajerison 2020 (2002-2007 Pneumonic)	40	45	0	64	1.00 [0.91, 1.00]	0.59 [0.49, 0.68]
Rajerison 2020 (2017-2018 Bubonic)	22	2	0	9	1.00 [0.85, 1.00]	0.82 [0.48, 0.98]
Rajerison 2020 (2017-2018 Pneumonic)	1	23	3	40	0.25 [0.01, 0.81]	0.63 [0.50, 0.75]
Rajerison 2020 (2018-2019 Bubonic)	57	23	0	71	1.00 [0.94, 1.00]	0.76 [0.66, 0.84]
Rajerison 2020 (2018-2019 Pneumonic)	8	6	0	27	1.00 [0.63, 1.00]	0.82 [0.65, 0.93]
Riehm 2011	47	41	0	61	1.00 [0.92, 1.00]	0.60 [0.50, 0.69]



Test 2. F1RDT versus culture for all forms of plague, sensitivity analysis 1

F1RDT versus culture for all forms of plague, sensitivity analysis 1

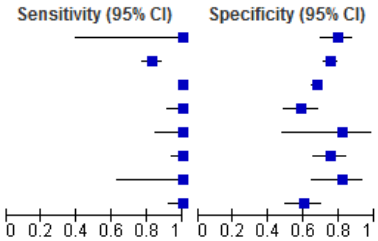
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Andrianaivoarimanana 2019	1420	979	0	1411	1.00 [1.00, 1.00]	0.59 [0.57, 0.61]
Bertherat 2011	4	19	0	73	1.00 [0.40, 1.00]	0.79 [0.70, 0.87]
Chanteau 2003a	151	128	31	381	0.83 [0.77, 0.88]	0.75 [0.71, 0.79]
Rajerison 2020 (2017-2018 Bubonic)	22	2	0	9	1.00 [0.85, 1.00]	0.82 [0.48, 0.98]
Rajerison 2020 (2017-2018 Pneumonic)	1	23	3	40	0.25 [0.01, 0.81]	0.63 [0.50, 0.75]
Rajerison 2020 (2018-2019 Bubonic)	57	23	0	71	1.00 [0.94, 1.00]	0.76 [0.66, 0.84]
Rajerison 2020 (2018-2019 Pneumonic)	8	6	0	27	1.00 [0.63, 1.00]	0.82 [0.65, 0.93]
Riehm 2011	47	41	0	61	1.00 [0.92, 1.00]	0.60 [0.50, 0.69]



Test 3. F1RDT versus culture for all forms of plague, sensitivity analysis 2

F1RDT versus culture for all forms of plague, sensitivity analysis 2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bertherat 2011	4	19	0	73	1.00 [0.40, 1.00]	0.79 [0.70, 0.87]
Chanteau 2003a	151	128	31	381	0.83 [0.77, 0.88]	0.75 [0.71, 0.79]
Rajerison 2020 (2002-2007 Bubonic)	1328	324	0	667	1.00 [1.00, 1.00]	0.67 [0.64, 0.70]
Rajerison 2020 (2002-2007 Pneumonic)	40	45	0	64	1.00 [0.91, 1.00]	0.59 [0.49, 0.68]
Rajerison 2020 (2017-2018 Bubonic)	22	2	0	9	1.00 [0.85, 1.00]	0.82 [0.48, 0.98]
Rajerison 2020 (2018-2019 Bubonic)	57	23	0	71	1.00 [0.94, 1.00]	0.76 [0.66, 0.84]
Rajerison 2020 (2018-2019 Pneumonic)	8	6	0	27	1.00 [0.63, 1.00]	0.82 [0.65, 0.93]
Riehm 2011	47	41	0	61	1.00 [0.92, 1.00]	0.60 [0.50, 0.69]



ADDITIONAL TABLES

Table 1. Summary of included studies

Study ID	Setting, date	n	Form of plague	Samples for RDT	RDT manufacturer	Location of performance of RDT	Reference standards
Andrianaivoarimanana 2019	Madagascar Surveillance 2002–2007[a]	4221 (3810 analyzed, 411 unknown)	BP[b] PP[b]	Not reported	Institut Pasteur Madagascar	Not reported	Culture
Bertherat 2011	DRC 2 outbreaks: 2005 and 2006	2005 outbreak: 130, RDT performed in 37 2006 outbreak: 162, RDT performed in 96	PP	Sputum	Institut Pasteur Madagascar	Unclear[c]	Culture PCR
Chanteau 2003a	Madagascar 2000–2001	671, RDT performed in 691 samples	BP (642) PP (20) Unknown (9)	Bubo aspirate Sputum Postmortem organ puncture	Institut Pasteur Madagascar	Central laboratory[d]	Culture
Petersen 2018	Madagascar 2004–2006 Uganda 2004–2017	≥ 254	At least: BP (225) PP (26) SP (2) Cutaneous (2)	Bubo aspirate Sputum	New Horizons	Central laboratory	Culture Paired serology
Rajerison 2020	Madagascar 3 periods: (a) 2002–2007 (b) 2017–2018 (c) 2018–2019	(A) 2468 (B) 100 (C) 192	BP (2503) PP (257)	Bubo aspirate Sputum Postmortem organ puncture	Institut Pasteur Madagascar	Central laboratory	Culture
Richard 2015	Madagascar Outbreak 2011	20, RDT performed in 2	PP	Sputum	Institut Pasteur Madagascar	Not reported	Culture PCR Paired serology

Table 1. Summary of included studies *(Continued)*

Riehm 2011	Madagascar 2007–2008	149	BP	Bubo aspirate	Institut Pasteur Madagascar	Bedside and central laboratory	Culture PCR

Abbreviations: BP: bubonic plague; n: number; PP: pneumonic plague; PCR: polymerase chain reaction; RDT: rapid diagnostic test; SP: septicaemic plague.
[a]Additional data are reported for 1998 to 2001 and 2008 to 2018. However, for these two periods, there were no data available in order to assess diagnostic test accuracy of the RDT.
[b]No disaggregated data between BP and PP.
[c]RDTs were performed onsite and repeated at a central laboratory. However, it was unclear which findings authors used for analysis.
[d]RDTs were performed at the participant's bedside and repeated at the central laboratory. However, authors used the findings from the test performed at the central laboratory for analysis.
[e]The authors determined positive disease status when either RDT performed onsite or at the central laboratory was positive. Five cases had a positive RDT when performed onsite but negative when repeated at the central laboratory; and two cases had a negative RDT when performed onsite but positive when repeated at the central laboratory.

Table 2. Disaggregated findings of F1RDT for diagnosing plague

Study ID	F1RDT	Form of plague	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
F1RDT versus culture								
Andrianaivoarimanana 2019	IPM	Bubonic and pneumonic (NDD)	1420	979	0	1411	100% (100 to 100)	59% (57 to 61)
Bertherat 2011 (2005 outbreak)	IPM	Pneumonic	0	17	0	18	Not estimable	51% (34 to 69)
Bertherat 2011 (2006 outbreak)	IPM	Pneumonic	4	19	0	73	100% (40 to 100)	79% (70 to 87)
Chanteau 2003a	IPM	Bubonic and pneumonic (NDD)	151	128	31	381	83% (77 to 88)	75% (71 to 79)
Rajerison 2020 (2002–2007)	IPM	Pneumonic	40	45	0	64	100% (91 to 100)	59% (49 to 68)
		Bubonic	1328	324	0	667	100% (100 to 100)	67% (64 to 70)
Rajerison 2020 (2017–2018)	IPM	Pneumonic	1	23	3	40	25% (1 to 81)	63% (50 to 75)
		Bubonic	22	2	0	9	100% (85 to 100)	82% (48 to 98)
Rajerison 2020 (2018–2019)	IPM	Pneumonic	8	6	0	27	100% (63 to 100)	82% (65 to 93)

Table 2. Disaggregated findings of F1RDT for diagnosing plague *(Continued)*

		Bubonic	57	23	0	71	100% (94 to 100)	76% (66 to 84)
Richard 2015	IPM	Pneumonic	0	2	0	0	Not estimable	Not estimable
Riehm 2011 [a]	IPM	Bubonic	47	41	0	61	100% (92 to 100)	60% (50 to 69)
F1RDT versus PCR								
Bertherat 2011 (2005 outbreak)[b]	IPM	Pneumonic	0	1	0	1	Not estimable	50% (1 to 99)
Richard 2015 (*pla* and *caf1*)	IPM	Pneumonic	2	0	0	0	Not estimable	Not estimable
Riehm 2011 (*caf1*)[c]	IPM	Bubonic	84	4	4	57	95% (89 to 99)	93% (84 to 98)
Riehm 2011 (*pla*)[c]	IPM	Bubonic	86	2	34	27	72% (63 to 80)	93% (77 to 99)
Riehm 2011 (*Ymt*)[c]	IPM	Bubonic	80	8	7	54	92% (84 to 97)	87% (76 to 94)
F1RDT versus paired serology								
Richard 2015	IPM	Pneumonic	2	0	0	0	100% (16 to 100)	Not estimable
F1RDT versus culture or paired serology								
Petersen 2018	NH	Bubonic, pneumonic, septicaemic, cutaneous	NR[d]	NR[d]	NR[d]	NR[d]	91% (84 to 95)[d]	88% (83 to 93)[d]

Abbreviations: CI: confidence interval; F1RDT: F1 antigen rapid diagnostic test; FN: false negative; FP: false positive; IPM: Institut Pasteur of Madagascar; NDD: no disaggregated data; NH: New Horizons; NR: not reported; PCR: polymerase chain reaction; TN: true negative; TP: true positive.
Note: we presented findings for positivity of F1RDT interpreted from 1+.
[a]The authors determined positive disease status when either RDT performed onsite or at the central laboratory was positive. Five cases had a positive RDT when performed onsite but negative when repeated at the central laboratory; and two cases had a negative RDT when performed onsite but positive when repeated at the central laboratory.
[b]There were no participants with both PCR and F1RDT findings reported from the 2006 outbreak. The gene tested by PCR during the 2005 outbreak was not specified.
[c]Three genes (*caf1*, *pla*, *Ymt*) were tested for the same group of participants. Therefore, findings for these three rows corresponded to the same cohort of participants.
[d]There was limited published available data for this study. Raw data on TP, TN, FP, and FN were not reported. The sensitivity and specificity estimates are those reported by the study authors.

APPENDICES

Appendix 1. Search strategies

Cochrane Central Register of Controlled Trials (CENTRAL)

#1 (plague):ti,ab,kw
#2 yersinia pestis
#3 #1 or #2
#4 diagnosis or diagnostic* or detect*
#5 RDT*
#6 "Enzyme-Linked Immunosorbent Assay" or ELISA
#7 lateral flow
#8 chromatograph*
#9 immunochromatograph*
#10 #4 or #5 or #6 or #7 or #8 or #9
#11 #3 and #10

MEDLINE (PubMed)

Search	Query
#1	Search plague Field: Title/Abstract
#2	Search yersinia pestis Field: Title/Abstract
#3	Search "y pestis" Field: Title/Abstract
#4	Search "Plague"[Mesh]
#5	Search "Yersinia pestis"[Mesh]
#6	Search (((#5) OR #4 OR #3) OR #2 OR #1
#7	Search diagnosis or diagnostic* or detect* Field: Title/Abstract
#8	Search RDT* Field: Title/Abstract
#9	Search "rapid diagnos* " Field: Title/Abstract
#10	Search "Enzyme-Linked Immunosorbent Assay" or ELISA Field: Title/Abstract
#11	Search "lateral flow" Field: Title/Abstract
#12	Search chromatograph* Field: Title/Abstract
#13	Search immunochromatograph* Field: Title/Abstract
#14	Search dipstick* Field: Title/Abstract
#15	Search F1 antigen* Field: Title/Abstract
#16	Search F1RDT* Field: Title/Abstract
#17	Search "rapid identification" Field: Title/Abstract

(Continued)

#18	Search ((test kit [Title/Abstract] OR test kits [Title/Abstract] OR testing kit [Title/Abstract] OR testing kits [Title/Abstract]))
#19	Search "Immunoassay"[Mesh]
#20	Search "Reagent Kits, Diagnostic"[Mesh]
#21	Search "Reagent Strips"[Mesh]
#22	Search "Sensitivity and Specificity"[Mesh]
#23	Search ((((((((((((#22) OR #21) OR #20) OR #19) OR #18) OR #17) OR #16) OR #15) OR #14) OR #13) OR #12) OR #11 OR #10) OR #9 OR #8 OR #7
#24	Search #23) AND #6

Embase (Ovid)

1 plague.mp. or *plague/
2 yersinia pestis.mp. or Yersinia pestis/
3 *diagnosis/
4 diagnostic test/
5 RDT*.mp.
6 "rapid diagnos$ test$ ".ab. or "rapid diagnos$ test$".ti.
7 ELISA.mp. or enzyme linked immunosorbent assay/
8 immunochromatography/ or immunoaffinity chromatography/ or chromatography/ or immunochromatograph*.mp.
9 lateral flow.mp.
10 1 or 2
11 3 or 4 or 5 or 6 or 7 or 8 or 9
12 10 and 11

Science Citation Index-Expanded (Web of Science)

# 3#2 AND #1
Indexes=SCI-EXPANDED, CPCI-S, Timespan=All years
# 2TS=(RDT* or "Enzyme-Linked Immunosorbent Assay" or ELISA) OR TS=("lateral flow" or chromatography or immunochromatography)
Indexes=SCI-EXPANDED, CPCI-S, Timespan=All years
# 1TOPIC: (plague or yersinia pestis)
Indexes=SCI-EXPANDED, CPCI-S, Timespan=All years

Google Scholar = plague and diagnosis, plague and RDT

Clinicaltrials.gov, WHO ICTRP: plague and diagnosis, plague and RDT

Appendix 2. Data extraction form

Study ID	First author Year of publication Journal of publication
Setting	Country Plague prevalence and endemicity in study setting Study start and end dates

(Continued)

Study design	Whether participants were enrolled prospectively or retrospectively
	Sampling strategy (consecutive or random)
	Inclusion and exclusion criteria
Target condition	Any form of plague or a particular form of plague (bubonic, septicaemic, pneumonic), with case definitions
Participants	Sample size
	Characteristics: age, gender, comorbidities
	Signs and symptoms presented
	Recent prior antibiotic treatment
Index test	Brand name, target antigen, batch numbers
	Which biological sample was tested (urine, sputum, bubo aspirate)?
	Transport and storage condition
	Need for sample preparation
	Who performed the test (including any special training provided)?
	Where was the test performed (field or laboratory)?
	Threshold considered for positive result?
Reference standard	Which reference standard was used (culture, PCR, serology, combination)?
	Which biological sample was tested (blood, urine, sputum, bubo aspirate)?
	Who performed the reference standard test(s) (including training level)?
	Where was the test performed?
	How many observers or repeats were used?
	Time between RDT and reference test?
	Blinding of operator to RDT result?
	Has the laboratory received quality accreditation by an external agency?
Index and reference standard test results	Numbers of true positives, false positives, true negatives, and false negatives
	Number of uninterpretable or doubtful results
Notes	Source of funding
	Anything else of relevance

PCR: polymerase chain reaction, RDT: rapid diagnostic test.

Appendix 3. Tailored QUADAS-2 tool

Item	Yes	No	Unclear
Domain 1: patient selection			
Was a consecutive or random sample of patients enrolled?	If the study reported consecutive enrolment or random sampling of participants presenting with suspicion of plague.	If participants were purposefully selected, for example based on previous test results.	If insufficient information to make a decision on how participants were selected.
Was a case-control design avoided?	This item will always be 'Yes' because we excluded case-control studies from this review.	Not applicable.	Not applicable.
Did the study avoid inappropriate exclusions?	If no participants were excluded after inclusion in the study, or if exclusions were clearly described and appropriate (e.g. exclusion of the participants with a known diagnosis).	If specific populations who would be representative of field conditions were excluded.	If unreported or insufficient information to make a decision.
Did the study considered prior administration of antibiotics?	If participants who received antibiotics prior to sample collection were excluded.	If participants who received antibiotics prior to sample collection were included.	If unreported or insufficient information to make a decision.
Risk of bias (high, low, or unclear)	Could the selection of participants have introduced bias?		
	'High' if ≥ 1 of the above signalling questions was 'No,' indicating that there was a concern.	'Low' if the answer to all 3 signalling questions was 'Yes.'	'Unclear' if the answer to ≥ 1 signalling question was 'Unclear' and 0 are answered 'No.'
Applicability concerns (high, low, or unclear)	Are there concerns that the included participants did not match the review question?		
	'High' if the included participants were inherently different from the participants who would be expected to receive the RDT.	'Low' if the included participants were suspected to have plague and matched those who would be expected to receive the test.	'Unclear' if insufficient information on participant characteristics to make a decision.
Domain 2: index test			
Were the index test results interpreted without knowledge of the results of the reference standard?	If RDT was performed fully blinded to the reference standard result.	If reference standard result was known prior to interpretation of RDT result.	If blinding to reference standard result was not explicitly stated.
If a threshold was used, was it prespecified?	If a threshold was prespecified.	If a threshold was not prespecified.	If unreported.
Risk of bias (high, low, or unclear)	Could the conduct or interpretation of the index test have introduced bias?		
	'High' if the answer to either of the above signalling questions was 'No,' indicating that there was a concern.	'Low' if the answer to both signalling questions was 'Yes.'	'Unclear' if the answer to ≥ 1 signalling question was 'Unclear' and 0 were answered 'No.'
Applicability concerns	Are there concerns that the index test, its conduct, or interpretation differed from the review question?		

(Continued)

(high, low, or unclear)	'High' if the index test was not performed in field conditions, or if the study described inappropriate storage conditions for the index test.	'Low' if the study described suitable storage conditions for the index test and that the index test was designed for testing biological samples for plague and was used in field conditions.	'Unclear' if insufficient information to make a decision.
Domain 3: reference standard			
Was the reference standard likely to correctly classify the target condition?	This item will always be 'Yes' because a correct reference standard was part of the inclusion criteria of this review.	Not applicable.	Not applicable.
Were the reference standard results interpreted without knowledge of the results of the index tests?	This item will always be 'Yes' because all the reference standards (culture, PCR, and serology) are objective tests with no room for subjective interpretation of test results.	Not applicable.	Not applicable.
Risk of bias (high, low, or unclear)	Could the reference standard, its conduct, or its interpretation have introduced bias?		
	'High' if the answer to either of the above signalling questions was 'No,' indicating that there was a concern.	'Low' if the answer to both signalling questions was 'Yes.'	'Unclear' if the answer to ≥ 1 signalling question was 'Unclear' and 0 were answered 'No.'
Applicability concerns (high, low, or unclear)	Are there concerns that the target condition as defined by the reference standard did not match the review question?		
	We answered this question as 'Low' concerns for all studies because diagnosis of plague by culture, PCR, or paired serology does match the review question.		
Domain 4: flow and timing			
Was there an appropriate interval between index test and reference standard?	If no antibiotic was administered between sample collection for index test and reference standard, and if transportation of samples was < 7 days. We considered that the introduction of antibiotics was more relevant than time between collection of samples for both tests, as people with suspicion of plague will be started on antibiotics as early as possible, and might affect results (of culture mainly).	If antibiotic therapy was started between sample collection for RDT and reference standard for a significant proportion of participants, or if transportation of samples was > 7 days on average.	If there was insufficient information to make a decision.
Did all participants receive a reference standard?	If all participants received a reference standard.	If participants did not receive a reference standard.	If there was insufficient information to determine whether or not all participants received a reference standard.
Did all participants receive the same reference standard?	If all the participants received the same reference standard.	If participants did not receive the same reference standard.	If there was insufficient information to determine whether or not all partici-

(Continued)

			pants received the same reference standard.
Were all participants included in the analysis?	If there were no withdrawals or exclusions (number of participants in the 2×2 table matched the number of participants recruited into the study) or if sufficient explanation was given for any discrepancy.	If withdrawals or exclusions were not explained or accounted for.	In unreported or insufficient information to make a decision.
Risk of bias (high, low, or unclear)	Could the patient flow have introduced bias?		
	'High' if ≥ 1 of the above signalling questions was 'No,' indicating that there was a concern.	'Low' if the answer to all above signalling questions was 'Yes.'	'Unclear' if the answer to ≥ 1 signalling question was 'Unclear' and none were answered 'No.'
Applicability concerns	Not applicable.		

PCR: polymerase chain reaction, RDT: rapid diagnostic test.

HISTORY

Protocol first published: Issue 10, 2019
Review first published: Issue 6, 2020

CONTRIBUTIONS OF AUTHORS

SJ and HAD assessed the eligibility of the studies, extracted the data, and assessed the methodological quality of the included studies.

MC conducted the statistical analyses and contributed to the assessment of the methodological quality of the included studies.

SJ interpreted the findings and assessed the certainty of the evidence using the GRADE approach.

SJ drafted the text with considerable input from MC.

All review authors read and approved the final manuscript draft.

DECLARATIONS OF INTEREST

SJ: worked for the CIDG at the Liverpool School of Tropical Medicine from September 2015 to April 2016. SJ has a contract with the World Health Organization (WHO) for the development of evidence synthesis in the process of updating the WHO plague guideline.

HAD: none.

MC: none.

SOURCES OF SUPPORT

Internal sources

- Liverpool School of Tropical Medicine, UK

External sources

- Department for International Development (DFID), UK

  Project number 300342-104
- World Health Organization, Switzerland

  Evidence synthesis commissioned to inform the WHO guideline panel for developing recommendations.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We conducted the review according to the published protocol (Jullien 2019).

INDEX TERMS

Medical Subject Headings (MeSH)

Antigens, Bacterial [*analysis]; Confidence Intervals; Cross-Sectional Studies; False Negative Reactions; False Positive Reactions; Plague [*diagnosis] [immunology]; Sensitivity and Specificity; Time Factors; Yersinia pestis [*immunology]

MeSH check words

Adult; Child; Humans