![Cochrane Library logo]

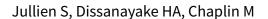**Cochrane** Database of Systematic Reviews

Rapid diagnostic tests for plague (Protocol)

Jullien S, Dissanayake HA, Chaplin M

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

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, Barcelona Institute for Global Health, University of Barcelona, Barcelona, Spain. sophjullien@gmail.com.

A B S T R A C T

This is a protocol for a Cochrane Review (Diagnostic test accuracy). The objectives are as follows:

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

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 few decades. Between 1989 and 2003, 38,310 human cases of plague were reported from 25 countries, 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 are the countries with 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 large 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 (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 persons 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. The bubonic plague is the most common form and is characterized by enlarged lymph nodes with necrotic areas called buboes. Pneumonic plague is 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 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 to 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. RDTs are therefore useful diagnostic tools for use at the community level and in low-resource settings.

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 lower concentration of the F1 antigen that the test can detect. The F1RDT can be used in bubo aspirate, urine and sputum; it is not usually used in blood as the pink result line would be difficult to see. Currently, the F1RDT that is mainly being used in the field is produced in Madagascar. The updated version was developed in 2001 (Chanteau 2003). Other F1RDTs for plague are produced by New Horizons in the USA, although it is not licensed for use in humans (New Horizons 2019), and in Taiwan (Hsu 2018). 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 associated with lymph node swelling in case of bubonic plague, and/or cough, haemoptysis and chest pain for pneumonic plague. While a first diagnosis of plague is suspected 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. *Y pestis* grows easily in standard culture media, and while bacteriological isolation is highly specific, it is also highly sensitive un-

Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

der 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 obtention 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, targeting several genes including the *pla* gene, encoding plasminogen activator, and the *caf1* gene, encoding F1 capsule antigen.

These diagnostic tests (culture, paired serology, and polymerase chain reaction) 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 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 and/or bubo aspirate) 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 patients 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 patients 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 fun-

damental 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 patients, for example asymptomatic persons that have been in contact with people suffering from 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 persons.

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 (mainly for the pneumonic and septicaemic forms 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. 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. 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.

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 take into account while evaluating the F1RDT is the use and importance of specificity of the test in diagnosing patients 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 patients 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. The F1RDT therefore 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 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). It is therefore 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 polymerase chain reaction.

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.

OBJECTIVES

To determine the diagnostic accuracy of the rapid diagnostic test (RDT) based on the antigen F1 (F1RDT) for detecting plague in persons 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 will include 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, polymerase chain reaction, or serology). We will exclude 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 will include patients (including children and pregnant women) living in or visiting areas where plague is 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 is eligible for inclusion in the review, we will include the study provided that there are disaggregated data that we can extract for that subgroup.

Index tests

The index test we will assess is the F1RDT to detect plague.

Target conditions

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

Reference standards

We will include studies that use one of the following reference standards to diagnose plague.

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

Search methods for identification of studies

We will attempt to identify all relevant studies regardless of language, publication status, or publication date.

Electronic searches

We will search 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), MEDLINE (PubMed), Embase (accessed via Ovid), Science Citation Index (Web of Science). We will also search the WHO International Clinical Trials Registry Platform (ICTRP) and ClinicalTrials.gov for trials in progress.

Searching other resources

We will search the proceedings and abstracts of relevant conferences from the past five years. We will handsearch the reference lists of relevant papers and contact researchers working in the field. We will also search for related articles to the included studies using the PubMed "similar articles" function.

Data collection and analysis

Selection of studies

Two review authors will independently screen all the abstracts retrieved by the search strategy, using the predefined eligibility criteria. We will exclude studies that are clearly irrelevant based on the titles and abstracts. We will retrieve full-text copies of the remaining studies and apply the predefined criteria for inclusion in the review. We will resolve any disagreements in assessment through discussion, or by referral to the third review author when required. We will list all studies excluded after full-text assessment in a 'Characteristics of excluded studies' table. We will illustrate the study selection process in a PRISMA diagram.

Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

Data extraction and management

Two review authors will independently conduct data extraction and management, using data extraction forms. We will compare these data and resolve any disagreement through discussion. For each included study, we will gather information into 'Characteristics of included studies' tables, comprising data on the following (Appendix 2).

- Setting, design and duration of the study.
- Baseline characteristics of the study population and the 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 will independently assess the methodological quality of each included study, using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, which is based on four domains: patient 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 will answer each of the signalling questions as either 'yes', 'no', or 'unclear', and give the reason for our judgement. We will resolve any disagreement through discussion and through consultation with a third review author in case of persisting disagreement.

Statistical analysis and data synthesis

We will stratify all analyses by the reference standard used. Within each stratum, we will construct a two-by-two table (containing the number of true positive, true negative, false positive, and false negative results) for each study. Where only sensitivity and specificity estimates are reported, we will attempt to derive the two-by-two table from the reported data. We will enter the two-by-two data into Review Manager 5 (RevMan 2014). Estimates of sensitivity and specificity from each individual study will be summarized on forest plots and plotted using summary receiver operating characteristic (SROC) plots.

We will also calculate positive and negative predictive value estimates for various scenarios to help interpret the impact of F1RDT findings.

If meta-analysis is appropriate given the number of studies and extent of clinical heterogeneity, we will pool results from the included studies. If there is little variation in threshold between studies, we will use the bivariate model to obtain pooled estimates of sensitivity and specificity at common thresholds. The bivariate model will be fitted using the metandi and xtmelogit commands in Stata version 14 (Stata 2015). We will plot the pooled estimates of sensitiv-

ity and specificity using SROC plots in Review Manager 5 (RevMan 2014). If thresholds vary considerably between the included studies, we will plot the sensitivity and specificity for each study in ROC space and use the hierarchical summary receiver operating characteristic (HSROC) model to estimate a SROC curve. The HSROC model will be fitted using PROC NLMIXED in the SAS software (SAS 2011). We may estimate both summary sensitivity and specificity points, and SROC curves, if both methods will produce clinically useful information.

Investigations of heterogeneity

If data are available, we will assess the impact of the following variables on accuracy of F1RDTs by performing subgroup analyses or meta-regression (by inclusion as a covariate in the bivariate model — using xtmelogit — or the HSROC model).

- Forms of plague (bubonic, septicaemic, and pneumonic).
- Specimen tested (bubonic aspirate, urine, and sputum).
- Prior antibiotic treatment.
- Location of performance of the F1RDT (field and laboratory).
- Threshold for detecting the disease (as set by the manufacturer).

We will stratify the findings by type of reference standard used: bacterial isolation by culture, polymerase chain reaction, and serology showing a four-fold difference in F1 antibody titres between two samples from acute and convalescent phases.

Sensitivity analyses

We may perform a sensitivity analysis in which we only include studies that have a low risk of bias for the four domains (patient selection, index test, reference standard, and flow and timing). We may also conduct a sensitivity analysis restricted to those studies at low risk of bias for patient selection only. We will compare these results with those including all included studies, to investigate the robustness of the diagnostic accuracy estimates.

Assessment of reporting bias

Little is known on how to assess and detect reporting bias for diagnostic test accuracy studies (Macaskill 2010). We may test for association between the natural-logarithm of the diagnostic odds ratio (lnDOR) and the "effective sample size", a simple function of the number of diseased and non-diseased individuals (Deeks 2005). However, this test has low power for detecting funnel plot asymmetry when there is heterogeneity in the diagnostic odds ratio (DOR). If funnel plot asymmetry is detected, we will explore potential reasons for this association between study size and test accuracy, as reporting bias may not necessarily be the cause.

Assessment of the certainty of the evidence

We will assess the certainty of the evidence using the GRADE principles and GRADEpro GDT software (GRADE Handbook 2013; GRADEpro GDT 2015). We will rate the certainty of the evidence as either high, moderate, low or very low by assessing four domains (risk of bias, indirectness, inconsistency and imprecision), as follows.

- Risk of bias: we will assess risk of bias by using the QUADAS-2 tool.
- Indirectness: we will use the QUADAS-2 tool to assess applicability concerns and look for important differences between the populations studied, the setting, and the review question.

- Inconsistency: we will explore inconsistency by investigating potential sources of heterogeneity, and we will downgrade the certainty of the evidence when we cannot explain inconsistency in the accuracy estimates.
- Imprecision: we will consider the width of the confidence intervals (CIs) and question whether the truth set at the lower or upper limit of the 95% CI would change our decision. We will also calculate 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 will make judgements on imprecision using these calculations.

We will construct a 'Summary of findings' table, which will show the main review findings along with the certainty of the evidence.

ACKNOWLEDGEMENTS

REFERENCES

Additional references

CDC 2019

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

Chanteau 2003

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**:211-6.

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).

Deeks 2005

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005;**58**:882-93.

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). https://gdt.gradepro.org/app/handbook/handbook.html.

GRADEpro GDT 2015 [Computer program]

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

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 Cochane Collaboration, 2010. Available from: methods.cochrane.org/sdt/handbook-dta-reviews. The Cochrane Collaboration.

Mavalankar 1995

Mavalankar DV. Indian 'plague' epidemic: unanswered questions and key lessons. *Journal of the Royal Society of Medicine* 1995;**88**(10):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. Yersinia 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.

RevMan 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.

SAS 2011 [Computer program]

SAS Institute Inc. SAS/STAT®. Version 9.3. Cary, North Carolina: SAS Institute Inc, 2011.

Stata 2015 [Computer program]

StataCorp. Stata. Version 14. College Station, TX: StataCorp, 2015.

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, Regional Office for South-East Asia (SEARO). Operational guidelines on plague surveillance, diagnosis, prevention and control. 2009. https://apps.who.int/iris/bitstream/handle/10665/205593/B4534.pdf?sequence=1&isAllowed=y (accessed prior to 21 October 2019).

WHO 2016

World Health Organization. Plague around the world, 2010-2015. Weekly Epidemiological Record 2016; Vol. 91, issue 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).

A P P E N D I C E S

Appendix 1. MEDLINE (PubMed) search strategy

#1Search "Plague"[Mesh]
#2Search "Yersinia pestis"[Mesh]
#3Search plague Field: Title/Abstract
#4Search ((#3) OR "Yersinia pestis"[Mesh]) OR "Plague"[Mesh]
#5Search diagnosis [sh]
#6Search diagnosis or diagnostic* or detect* Field: Title/Abstract
#7Search "Diagnostic Techniques and Procedures"[Mesh]
#8Search RDT* Field: Title/Abstract
#9Search "Enzyme-Linked Immunosorbent Assay"[Mesh]
#10Search ELISA Field: Title/Abstract
#11Search lateral flow Field: Title/Abstract
#12Search "Chromatography, Affinity"[Mesh]
#13Search immunochromatograp* Field: Title/Abstract
#14 Search (((#13) OR #12) OR #11 OR #10 OR #9 OR #8 OR #7 or #6 OR #5
#15Search (#14) AND #4

This is the preliminary search strategy for MEDLINE (PubMed). It will be adapted for other electronic databases. We will report all search strategies in full in the final version of the review.

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
Study design	Whether patients 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)?

(Continued)

	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

Abbreviations: 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 patients presenting with suspicion of plague.	If patients were purposefully selected, for example based on previous test results.	If insufficient information to make a decision on how patients were selected.
Was a case-control design avoided?	This item will always be 'Yes' because we will exclude case-control studies from this review.	Not applicable.	Not applicable.
Did the study avoid inappropriate exclusions?	If no patients were excluded after inclusion in the study, or if exclusions are clearly described and appropriate (for example exclusion of the patients 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 patients who received antibiotics prior to sample collection were excluded.	If patients 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 patients have introduced bias?		
	'High' if at least one of the above signalling questions is 'No', indicating that there is a concern.	'Low' if the answer to all three signalling questions is 'Yes'.	'Unclear' if the answer to at least one signalling question is 'Unclear' and none are answered 'No'.
Applicability concerns (high, low, or unclear)	Are there concerns that the included patients do not match the review question?		
	'High' if the included participants are inherently different from the patients who would be expected to receive the RDT.	'Low' if the included participants are suspected to have plague and match those who would be expected to receive the test.	'Unclear' if there is insufficient information on patient characteristics to make a decision.

(Continued)

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 pre-specified?	If a threshold is pre-specified.	If a threshold is not pre-specified.	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 is 'No', indicating that there is a concern.	'Low' if the answer to both signalling questions is 'Yes'.	'Unclear' if the answer to at least one signalling question is 'Unclear' and none are answered 'No'.
Applicability concerns (high, low, or unclear)	Are there concerns that the index test, its conduct, or interpretation differs from the review question?		
	'High' if the index test is not performed in field conditions, or if the study describes inappropriate storage conditions for the index test as described by the manufacturer.	'Low' if the study describes suitable storage conditions for the index test as described by the manufacturer and that the index test is designed for testing biological samples for plague and is used in field conditions.	'Unclear' if there is insufficient information to make a decision.

Domain 3: Reference standard

Is the reference standard likely to correctly classify the target condition?	This item will always be 'Yes' because a correct reference standard is 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 has introduced bias?		
	'High' if the answer to either of the above signalling questions is 'No', indicating that there is a concern.	'Low' if the answer to both signalling questions is 'Yes'.	'Unclear' if the answer to at least one signalling question is 'Unclear' and none are answered 'No'.
Applicability concerns (high, low, or unclear)	Are there concerns that the target condition as defined by the reference standard does not match the review question?		
	We will answer this question as 'low' for all studies because diagnosis of plague by culture, PCR or paired serology does match the review question.		

(Continued)

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 less than 7 days. We felt that the introduction of antibiotics was more relevant than time between collection of samples for both tests, as patients with suspicion of plague will be started on antibiotics as early as possible, and might affect results (of culture mainly).	If antibiotherapy was started between sample collection for RDT and reference standard for a significant proportion of patients, or if transportation of samples was more than 7 days on average.	If there is insufficient information to make a decision.
Did all patients receive a reference standard?	If all the participants received a reference standard.	If participants did not receive a reference standard.	If there is insufficient information to determine whether or not all patients received a reference standard.
Did all patients 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 is insufficient information to determine whether or not all patients received the same reference standard.
Were all patients included in the analysis?	If there were no withdrawals or exclusions (number of participants in the two-by-two table matches the number of participants recruited into the study) or if sufficient explanation was given for any discrepancy.	If withdrawals or exclusions are not explained or accounted for.	In unreported or there is insufficient information to make a decision.
Risk of bias (high, low, or unclear)	Could the patient flow have introduced bias?		
	'High' if at least one of the above signalling questions is 'No', indicating that there is a concern.	'Low' if the answer to all above signalling questions is 'Yes'.	'Unclear' if the answer to at least one signalling question is 'Unclear' and none are answered 'No'.
Applicability concerns	Not applicable.		

Abbreviations: PCR: polymerase chain reaction; RDT: rapid diagnostic test.

CONTRIBUTIONS OF AUTHORS

SJ drafted the protocol and addressed the comments from the referees. MR provided methodological input and helped draft the protocol. HAD assessed the protocol draft. All review authors read and approved the final protocol draft.

DECLARATIONS OF INTEREST

SJ worked for the Cochrane Infectious Diseases Group at the Liverpool School of Tropical Medicine from September 2015 to April 2016. SJ has a contract with the WHO for the development of evidence synthesis in the process of updating the WHO plague guideline.

HAD has no known conflicts of interest.

MR has no known conflicts of interest.

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 (WHO), Switzerland.

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