How to Determine the Accuracy of an Alternative Diagnostic Test when It Is Actually Better than the Reference Tests: A Re-Evaluation of Diagnostic Tests for Scrub Typhus Using Bayesian LCMs

Cherry Lim[1], Daniel H. Paris[1,2], Stuart D. Blacksell[1,2], Achara Laongnualpanich[3], Pacharee Kantipong[3], Wirongrong Chierakul[1], Vanaporn Wuthiekanun[1], Nicholas P. J. Day[1,2], Ben S. Cooper[1,2], Direk Limmathurotsakul[1,4]*

1 Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, 2 Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom, 3 Chiang Rai Prachanukroh Hospital, Chiang Rai, Thailand, 4 Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

* direk@tropmedres.ac

Abstract

Background

The indirect immunofluorescence assay (IFA) is considered a reference test for scrub typhus. Recently, the Scrub Typhus Infection Criteria (STIC; a combination of culture, PCR assays and IFA IgM) were proposed as a reference standard for evaluating alternative diagnostic tests. Here, we use Bayesian latent class models (LCMs) to estimate the true accuracy of each diagnostic test, and of STIC, for diagnosing scrub typhus.

Methods/Principal Findings

Data from 161 patients with undifferentiated fever were re-evaluated using Bayesian LCMs. Every patient was evaluated for the presence of an eschar, and tested with blood culture for *Orientia tsutsugamushi*, three different PCR assays, IFA IgM, and the Panbio IgM immuno-chromatographic test (ICT). True sensitivity and specificity of culture (24.4% and 100%), 56kDa PCR assay (56.8% and 98.4%), 47kDa PCR assay (63.2% and 96.1%), *groEL* PCR assay (71.4% and 93.0%), IFA IgM (70.0% and 83.8%), PanBio IgM ICT (72.8% and 96.8%), presence of eschar (42.7% and 98.9%) and STIC (90.5% and 82.5%) estimated by Bayesian LCM were considerably different from those obtained when using STIC as a reference standard. The IgM ICT had comparable sensitivity and significantly higher specificity compared to IFA (p=0.34 and p<0.001, respectively).

Conclusions

The low specificity of STIC was caused by the low specificity of IFA IgM. Neither STIC nor IFA IgM can be used as reference standards against which to evaluate alternative diagnostic tests. Further evaluation of new diagnostic tests should be done with a carefully selected set of diagnostic tests and appropriate statistical models.

Introduction

Scrub typhus, a bacterial infection caused by *Orientia tsutsugamushi*, is endemic in South, East and Southeast Asia, as well as some areas in northern Australia [1,2]. The organism is transmitted to humans through skin inoculation by larval *Leptotrombidium* mites (chiggers). Scrub typhus can be severe and fatal when left untreated, with reported mortality ranging from 14% to 30% in Southeast Asia [3–5]. The diagnosis of scrub typhus is difficult. Patients with scrub typhus often come to hospital with undifferentiated fever and symptoms that are similar to other endemic infections such as leptospirosis, malaria and dengue. An eschar, a necrotic lesion formed at the site of inoculation, is the most characteristic sign of scrub typhus. However, an eschar is not observed in every scrub typhus patient, and similar lesions can also be observed in patients with other diseases such as spider bites, spotted fever group rickettsioses, and cutaneous lesions caused by tuberculosis, leishmaniasis and anthrax [6,7].

There are two main laboratory methods for diagnosing scrub typhus, namely bacterial and antibody detection. Bacterial detection methods include isolation of *O. tsutsugamushi* (culture) and polymerase chain reaction (PCR) assays targeting the 56kDa, 47kDa, and *groEL O. tsutsugamushi* genes [8–12]. Antibody detection methods include the indirect immunofluorescence antibody assay (IFA), the indirect immunoperoxidase assay (IIP), the Weil-Felix test, and various commercially available immunochromatographic tests (ICT) [8,13–15]. IFA uses fluorescent anti-human antibody to detect the presence of antibody specific to *O. tsutsugamushi* in patient serum, and is regularly used as a reference test against which alternative diagnostic tests for scrub typhus are evaluated [8,16]. However, IFA has several limitations [16]. The cut-off antibody titre of IFA for acute serum samples remains controversial, the determination of IFA results is subjective, and the true accuracy of IFA is suspected to be imperfect [8].

We recently proposed the Scrub Typhus Infection Criteria (STIC), a combination of culture, PCR assays, and IFA IgM, as a reference standard for scrub typhus diagnosis [17,18]. STIC is considered positive if either (a) *O. tsutsugamushi* is isolated, (b) at least two out of three PCR assays targeting the 56kDa, 47kDa and *groEL* genes are positive, (c) an admission IFA IgM titre is $\geq$ 1:12,800 or (d) there is at least a four-fold rise in convalescence IFA IgM titre compared to the admission IFA IgM titre [17,18]. The development and details of STIC, including selection of the cut-off titres of IFA IgM for STIC, are described elsewhere [17,18]. In short, STIC were designed based on available diagnostic tests to provide a robust set of criteria for a final diagnosis of acute scrub typhus infection with a high level of confidence in a research setting, and STIC have already been used as a comparator to evaluate the accuracy of several alternative diagnostic tests [17,18].

Nonetheless, we hypothesized that STIC, when used as a comparator, might have been falsely assumed to be perfect (100% sensitivity and 100% specificity), and as a consequence the accuracy of the alternative diagnostic tests might have been inaccurately estimated. Bayesian latent class models (LCM) are increasingly used to estimate accuracy of diagnostic tests since it does not need to assume that the accuracy of reference tests is perfect [19–24]. In this study, we

re-analyzed our existing data set from a previously published prospective study [17,18] and estimated the true accuracy of each diagnostic test for scrub typhus using Bayesian LCM.

Materials and Methods

Study patients

The data set analysed in this study was from a prospective study of acute febrile illness in a scrub typhus endemic area, Chiang Rai, Northern Thailand [17,18]. In brief, a total of 161 patients were recruited into the study from August 2007 to August 2008 [17,18]. All participants were over 15 years old, presented with acute febrile illness for less than 2 weeks, had three negative malaria blood smears, and had no evidence of a primary focus of infection. Every patient was examined for the presence of an eschar. Admission and convalescent blood samples were collected for a set of diagnostic tests [17,18].

Ethical statement

Ethical approval for the prospective study was obtained from the ethical committees of Chiang Rai Hospital, the Ministry of Public Health, Thailand, and from the Oxford Tropical Research Ethics Committee, United Kingdom. Signed written inform consent was obtained from every subject enrolled into the study [17,18].

Diagnostic tests

All patient specimens were tested for evidence of acute scrub typhus infection using a set of bacterial and antibody detection methods [17,18]. All of the tests were done and reported in the previous publications [17,18]. In short, the bacterial detection tests were performed on admission samples and included *in vitro* isolation of *O. tsutsugamushi* (culture), nested 56kDa PCR assay, 47kDa-based real-time PCR assay, and *GroEL*-based real-time PCR assay as described previously [12,17,25–27]. Antibody detection tests performed included on-admission ICT (PanBio ICT, Australia) and paired IFA, both targeting specific IgM antibodies against *O. tsutsugamushi*.

IgM-based IFA was used to measure IgM-specific titers against pooled whole-cell antigens from three strains of *O. tsutsugamushi* (the Karp, Kato and Gilliam type strains) as previously described [16]. The scrub typhus IgM ICT was a prototype rapid diagnostic test developed by Panbio Company, Australia, and was performed according to the manufacturer's instructions. The antigen used in the Panbio IgM ICT was a truncated recombinant 56-kDa protein of *O. tsutsugamushi* Karp strain [28].

Statistical analysis

STIC as gold standard in Gold Standard Model. The accuracy of each diagnosis test was estimated by comparing with STIC, which were considered as a perfect reference test (100% sensitivity and 100% specificity), as previously described [17,18]. STIC was considered positive if either (a) *O. tsutsugamushi* was isolated, (b) at least two out of three PCR assays targeting the 56kDa, 47kDa and *groEL* genes were positive, (c) an admission IFA IgM titre was ≥ 1:12,800 or (d) there was at least a four-fold rise in convalescence IFA IgM titre compared to the admission IFA IgM titre [17,18]. A combination of two diagnostic tests was considered positive if either one of those tests was positive. Prevalence, sensitivity, specificity, positive, and negative predictive values (PPV and NPV) of the tests with their respective 95% confidence intervals (CI) were calculated using STATA 11.1 (Stata Corp., College Station, Tex.). The results obtained were comparable to those previously published [17,18].

Bayesian Latent Class Models.   The true accuracy of each diagnostic test was then estimated using Bayesian LCMs. Detailed explanation on how Bayesian LCMs determine the unbiased accuracy of diagnostic tests has been given elsewhere [20–24,29,30]. In brief, Bayesian LCMs estimate accuracies of diagnostic tests based on the true disease status of each patient (infected or non-infected). Bayesian LCMs do not assume that any diagnostic test or combination of diagnostic tests is perfect. Diagnostic tests in the model included culture, a combination of PCR assays, paired IFA IgM, PanBio IgM ICT and presence of eschar. The combination of PCR assays and paired IFA IgM was defined according to the definition of STIC, which considered PCR assays positive when at least two out of three PCR assays, targeting 56kDa, 47kDa and *groEL* genes, were positive, and IFA IgM positive when admission IFA IgM titre was ≥ 1:12,800 or there was at least a four-fold rise in convalescence IFA IgM titre compared to admission IFA IgM titre.

To estimate the accuracy of a diagnostic test using LCMs, the best-fitting model, as determined by the presence or absence of correlation between diagnostic tests, was used. We evaluated possible correlations based on existing knowledge and external evidence. Therefore, correlations between bacterial detection tests (culture and a combination of PCR assays) and between antibody detection tests (IFA IgM and PanBio IgM ICT) were evaluated. As the Deviance Information Criterion (DIC) is not applicable to the random effect model used, the Akaike Information Criterion (AIC) was used to evaluate goodness of fit and to compare the models. A difference in AIC of more than 10 was taken as indicating definite support for the model with the lower value, while a difference between 5 and 10 was considered substantial support, and less than 5 inconclusive support.

The best-fitting model was used to determine the accuracy of each diagnostic test evaluated and combinations thereof. The Bayesian p values were used to compare accuracies between diagnostic tests. All Bayesian LCMs assumed that no prior information (non-informative priors) about unknown parameters (i.e. prevalence, sensitivities and specificities) was available, except that the specificity of culture was fixed at 100%. Sensitivity analysis was performed in which different prior information were used. All parameters with respective 95% credible intervals (CrI) were estimated using a Markov chain Monte Carlo method and WinBUGS 1.4 [31]. Interquartile ranges (IQRs) are presented as 25th and 75th percentiles. S1 and S2 Text provide full data sets and all of the models used, respectively. S1 and S2 Tables provides details about the method and the results for the best-fitting model selection.

Post-hoc model evaluation.   We searched for other complete individual-level data sets in PubMed using MeSH terms ("Scrub Typhus" and "Sensitivity and Specificity") for analysis with Bayesian LCMs. To validate the low sensitivity of IFA estimated by the Bayesian LCMs, we also estimated naïve sensitivities of those tests in only those who were culture and/or PCR assay positive in our data set [23]. To validate the low specificity of IFA estimated by the Bayesian LCMs, we evaluated the IFA results in patients with a firm diagnosis of other diseases in the data set, as previously described [18]. Murine typhus was diagnosed by IFA detecting antibody to the *Rickettsia typhi* Wilmington strain in paired specimens using slides purchased from the Australian Rickettsial Reference Laboratory[18]. Dengue was diagnosed using Panbio ELISA tests (Panbio, Brisbane Australia) to detect NS1 antigen and IgM antibodies in paired specimens [18].

Results

One hundred and sixty one patients with acute febrile illness were recruited into the study. The median age was 41 years (IQR 29–51, range 15–81), and 99/158 (63%) were male. The median fever days before admission to hospital was 5 days (IQR 3–7 days), and convalescent samples

were available for 138 patients (85.7%). Fifty-five were diagnosed with acute scrub typhus infection by STIC, giving a prevalence of 34.2% (95%CI 26.9–42.0). There were 9, 27, 46 and 31 patients with positive scrub typhus test results on culture, using a combination of PCR assays, with the IFA IgM and with the PanBio ICT IgM, respectively. Seventeen patients had an eschar.

STIC as gold standard

Table 1 shows the accuracy of each diagnostic test estimated by comparing it with STIC and assuming that STIC are a perfect reference. The sensitivity of IFA IgM was estimated to be 83.6%, and its specificity was assumed to be perfect (100%).

Bayesian LCMs

Bayesian LCMs were then applied to obtain an unbiased estimate of the accuracy of each diagnostic test. There was no substantial difference between all Bayesian LCMs used (S2 Table). The model which took account of correlation between serological tests (IFA IgM and ICT IgM) had the lowest AIC and was selected as the best-fitting model.

True accuracies of diagnostic tests for scrub typhus estimated by the Bayesian LCM were significantly different from those estimated when assuming that STIC is perfect (Table 1). The Bayesian LCM estimated that STIC was an imperfect diagnostic test with a sensitivity of 90.5% and a specificity of 82.5%. The Bayesian LCM indicated that the prevalence of scrub typhus was actually much lower than estimated by relying on STIC, as there were a number of false positive results due to the low specificity of the IFA IgM (83.8%). In contrast, the true sensitivity of culture (24.4%), nested 56kDa PCR assay (56.7%), 47kDa-based real-time PCR assay (63.2%), *GroEL*-based real-time PCR assay (71.4%), PanBio ICT IgM (72.8%), and presence of eschar (42.7%) estimated by the Bayesian LCM were much higher than those estimated by assuming that STIC were perfect (Table 1). The specificity of Panbio ICT IgM was significantly higher than that of IFA (96.8% vs. 83.8%, p<0.001), but there was no significant difference in their sensitivities (72.8% vs. 70.0%, p = 0.34).

Because of the high diagnostic accuracy of the ICT IgM and the presence of an eschar, and the potential use of this combination at the bedside, we used the model to estimate the accuracy of a combination of both diagnostic tests. Using positive results from either test as indicating a positive diagnosis, the sensitivity and specificity of this combination were respectively 75.6% (95%CrI: 65.1–85.7%) and 95.9% (95%CrI: 93.0–99.2%, Table 1). Due to the high sensitivity of the *GroEL* PCR assay, we also estimated the accuracy of a combination of this PCR assay with either ICT IgM or with the presence of an eschar. Using positive results from either *GroEL* PCR assay or ICT IgM as indicating a positive diagnosis, the sensitivity and specificity of this combination were 88.6% (95%CrI: 79.0–94.4%) and 90.8% (95%CrI: 93.0–94.8%, Table 1), respectively.

Post-hoc model evaluation

We could not identify any other data sets that were available and applicable for re-analysis using Bayesian LCMs. The naïve sensitivity of each diagnostic test estimated in only those who were culture, PCR assays or eschar positive is shown in Table 2. The naïve sensitivities of the IFA IgM in all groups were comparable to that estimated by the Bayesian LCM, and none were as high as the 83.6% estimated under the assumption that STIC is a perfect reference. This indicated that the low sensitivity of IFA IgM estimated by the Bayesian LCM was credible. We also evaluated the positivity of IFA IgM for scrub typhus in those who had alternative diagnoses. Of 10 murine typhus patients, 4 (40%) were IFA IgM positive, while culture, PCR assays and IgM

Table 1. Prevalence of scrub typhus and accuracy of diagnostic tests, estimated using the Scrub Ty-phus Infection Criteria (STIC) as the gold standard or using Bayesian latent class models (LCMs).

Parameters	STIC as the gold standard (95% CI)†	Final Bayesian LCM (95% CrI)††
Prevalence	34.2 (26.9–42.0)	23.0 (15.9–31.5)
STIC		
Sensitivity	100	90.5 (79.6–100)
Specificity	100	82.5 (79.4–85.6)
PPV	100	60.0 (51.0–69.1)
NPV	100	97.2 (91.5–100)
Blood culture for *O. tsutsugamushi*		
Sensitivity	16.4 (7.8–28.8)	24.4 (12.2–41.3)
Specificity	100	100
PPV	100	100
NPV	69.7 (61.8–76.9)	81.6 (76.3–86.2)
A combination of PCR assays*		
Sensitivity	49.1 (35.4–62.9)	65.8 (47.4–82.4)
Specificity	100	97.9 (92.7–100)
PPV	100	88.9 (74.1–100)
NPV	79.1 (71.2–85.6)	91.0 (85.8–94.0)
Nested 56kDa-based PCR assay		
Sensitivity	40.0 (27.0–54.1)	56.8 (48.8–65.6)
Specificity	99.1 (94.9–100)	98.4 (96.7–100)
PPV	95.7 (78.1–99.9)	91.3 (78.3–100)
NPV	76.1 (68.1–82.9)	88.4 (83.3–92.0)
47kDa-based real-time PCR assay		
Sensitivity	47.3 (33.7–61.2)	63.2 (54.1–72.2)
Specificity	98.1 (93.4–99.8)	96.1 (93.2–98.4)
PPV	92.9 (76.5–99.1)	82.1 (67.9–92.9)
NPV	78.2 (70.2–84.9)	90.2 (85.0–93.2)
***GroEL*-based real-time PCR assay**		
Sensitivity	56.4 (42.3–69.7)	71.4 (62.5–80.0)
Specificity	96.2 (90.6–99.0)	93.0 (90.0–95.8)
PPV	88.6 (73.3–96.8)	74.3 (62.9–85.7)
NPV	81.0 (73.0–87.4)	92.1 (87.3–94.4)
IFA IgM**		
Sensitivity	83.6 (71.2–92.2)	70.0 (55.8–83.8)
Specificity	100	83.8 (76.4–90.0)
PPV	100	56.5 (47.8–65.2)
NPV	92.2 (85.7–96.4)	90.4 (85.2–94.8)
Panbio ICT IgM		
Sensitivity	47.3 (33.7–61.2)	72.8 (57.8–86.6)
Specificity	95.3 (89.3–98.5)	96.8 (91.7–99.7)
PPV	83.9 (66.3–94.5)	87.1 (77.4–100)
NPV	77.7 (69.6–84.5)	92.3 (86.9–96.2)
Presence of eschar		
Sensitivity	25.5 (14.7–39.0)	42.7 (26.4–61.1)
Specificity	97.2 (92.0–99.4)	98.9 (95.5–100)
PPV	82.4 (56.6–96.2)	94.1 (82.4–100)

(Continued)

Table 1. (*Continued*)

Parameters	STIC as the gold standard (95% CI)†	Final Bayesian LCM (95% CrI)††
NPV	71.5 (63.4–78.7)	85.4 (79.9–89.6)
A combination of PanBio ICT IgM and presence of eschar***		
Sensitivity	47.3 (33.7–61.2)	75.6 (65.1–85.7)
Specificity	93.4 (86.9–97.3)	95.9 (93.0–99.2)
PPV	78.8 (61.1–91.0)	84.9 (72.7–97.0)
NPV	77.3 (69.1–84.3)	93.0 (88.3–96.3)
A combination of *GroEL* PCR assay and PanBio ICT IgM ***		
Sensitivity	65.5 (51.4–77.8)	88.6 (79.0–94.4)
Specificity	92.5 (85.7–96.7)	90.8 (87.0–94.8)
PPV	81.8 (67.3–91.8)	75.0 (61.4–86.4)
NPV	83.8 (75.8–89.9)	96.6 (92.3–98.3)
A combination of *GroEL* PCR assay and presence of eschar***		
Sensitivity	61.8 (47.7–74.6)	84.6 (75.0–91.7)
Specificity	93.4 (86.9–97.3)	92.0 (88.6–95.1)
PPV	82.9 (67.9–92.8)	75.6 (63.4–85.4)
NPV	82.5 (74.5–88.8)	95.0 (90.8–97.5)

STIC is considered positive if either (a) *O. tsutsugamushi* is isolated, (b) at least two out of three PCR assays targeting the 56kDa, 47kDa and *groEL* genes are positive, (c) an admission IFA IgM titre is ≥ 1:12,800 or (d) there is at least a four-fold rise in convalescence IFA IgM titre compared to the admission IFA IgM titre [17,18].

†Values are means with 95% confidence interval.

††Values are medians with 95% credible interval.

*A combination of PCR assays was defined as positive when at least two out of the three PCR assays (56kDa PCR assay, 47kDa-based real-time PCR assay and *groEL*-based real-time PCR assays) were positive.

** IFA IgM was defined as positive in those with either admission IFA IgM titre of ≥1: 12,800 or at least a four-fold rise in convalescence IFA IgM titre compared to the admission IFA IgM titre.

*** A combination of two diagnostic tests was considered positive if either one of those tests was positive.

doi:10.1371/journal.pone.0114930.t001

ICT were all negative. Of 24 dengue patients, 5 (21%) were IFA IgM positive, while culture, PCR assays, and IgM ICT were all negative. This indicated that the low specificity of the IFA IgM in our setting was also credible.

Sensitivity analysis

Although sensitivities of diagnostic tests estimated by the Bayesian LCMs were minimally different when different priors were used, there was no substantial difference in specificities of diagnostic tests estimated and in the main outcomes of the analysis (S3 Table). All models estimated that IFA IgM had low sensitivity and low specificity, and that specificity of ICT IgM was significantly higher than that of IFA.

Discussion

The key finding of this study is that the sensitivity and specificity of current reference tests for scrub typhus (IFA IgM and STIC (which includes the IFA IgM as one of the criteria)) are low.

Table 2. Naïve sensitivity of diagnostic tests estimated in those who had blood culture positive for _O. tsutsugamushi_, had a combination of PCR assays positive or had an eschar.

Population (n total)	56kDa-based PCR assays	47kDa-based PCR assay	_GroEL_-based PCR assay	IFA IgM *	Panbio ICT IgM	Presence of an eschar
Patients who had blood culture positive (n = 9)	66.7% (n = 6)	66.7% (n = 6)	77.8% (n = 7)	66.7% (n = 6)	55.6% (n = 5)	44.4% (n = 4)
Patients who had a combination of PCR assays positive (n = 27)**	81.5% (n = 22)	96.3% (n = 26)	100% (n = 27)	77.8% (n = 21)	74.1% (n = 20)	40.7% (n = 11)
Patients who had an eschar (n = 17)	64.7% (n = 11)	58.8% (n = 10)	64.7% (n = 11)	76.5% (n = 13)	88.3% (n = 15)	Not applicable

Data are sensitivity (n positive).

* IFA IgM was defined as positive in those with either admission IFA IgM titre of $\geq$1: 12,800 or at least a four-fold rise in convalescence IFA IgM titre compared to the admission IFA IgM titre.

** A combination of PCR assays was defined as positive when at least two out of the three PCR assays (56kDa PCR assay, 47kDa-based real-time PCR assay and _groEL_-based real-time PCR assays) were positive.

doi:10.1371/journal.pone.0114930.t002

We confirmed this finding by performing a sophisticated statistical analysis, which did not a priori assume that either IFA IgM or STIC was a gold standard test with 100% sensitivity and 100% specificity. In addition, we also observed that a prototype IgM ICT had a better performance than IFA IgM. We demonstrated that the gold standard model inaccurately estimated the accuracy of the alternative diagnostic tests, and suggest that further evaluation of new diagnostic tests for scrub typhus should be performed with appropriate statistical models. Our study also has an implication for clinical practice. Our new finding supports the usage of a combination of IgM ICT assay and presence of eschar as a point-of-care diagnostic test for scrub typhus in resource-limited settings where PCR for scrub typhus is not part of standard of care. Various combinations of PCR, IgM ICT and presence of eschar could also be considered as a point-of-care diagnostic test for scrub typhus in developed countries.

To the best of our knowledge, this is the first study to systematically evaluate the accuracy of each diagnostic test and of reference tests for scrub typhus using Bayesian LCMs. We validated the low sensitivity of IFA by determining naïve sensitivities of each diagnostic test within different strata of the disease; the method that is also recommended to be used in the absence of a gold standard [23]. The low sensitivity of IFA IgM observed in all strata of the disease evaluated supports that the true sensitivity of IFA IgM is low. We also validated the low specificity of IFA IgM by checking the potential for false positivity of IFA IgM in patients who had a diagnosis of murine typhus or dengue infection in our study. We found that IFA IgM was commonly positive alone in those patients, and this supports the estimates obtained by our Bayesian LCM.

The finding that IFA IgM has a low sensitivity is neither novel nor surprising since IFA IgM have been found to be negative in scrub typhus patients who are blood culture positive for _O. tsutsugamushi_ or PCR assay positive [8]. This is presumably because antibodies to _O. tsutsugamushi_ may take several weeks to become detectable by IFA, and some patients might also fail to develop the antibodies against the organism even after a long period following infection [8]. Furthermore, significant variation in the antigenic composition of the IFA slides and the _O. tsutsugamushi_ strain causing the infection may also contribute to this lack of accuracy.

The confirmation that IFA IgM has a low specificity is alarming. A high rate of false positivity provided by IFA IgM has long been suspected [8,16,32,33]. The major weaknesses of IFA include the reading of the fluorescent-stained slides, and the arbitrary cut-off limits for scrub typhus diagnosis [8,16,34]. When reading IFA slides determination of the fluorescence endpoint is based on the judgement of microscopists; therefore, the test is inherently subjective

[16,35]. Dual infection with scrub typhus and other infectious diseases are found by serological tests to be common [33], but reports of dual infection by culture or molecular detection of both pathogens are rare [36]. These suggest that a proportion of the dual infections reported might actually be caused by false positivity or cross-reactivity of the IFA IgM. The plausible explanation is that infections caused by other organisms in our setting, such as murine typhus, may lead to production of IgM that reacts with the *O. tsutsugamushi* used in the IFA IgM assay [36]. The low specificity of the IFA IgM leads to the poor specificity of STIC. This is because STIC is considered positive if either culture is positive, at least two out of three PCR assays are positive, or IFA IgM is positive. Thus any false positivity due to IFA IgM will also lead to a false positivity of STIC.

A major effect of the low accuracy of STIC is that the accuracies of alternative diagnostic tests have been inaccurately estimated. The true sensitivities of blood culture, individual PCR assays, Panbio ICT IgM and presence of an eschar are much higher than those estimated by assuming that STIC is a perfect reference. The Panbio ICT IgM was erroneously discarded and not further developed into the market because of its apparent low sensitivity (Personal communication, SDB). Sensitivity of the Panbio ICT IgM was reported as 47.3% by comparing with STIC as a gold standard [17,18]. In this study, we show that the prototype ICT IgM developed by the Panbio Company actually has a high sensitivity and a higher specificity than the IFA IgM in our setting. Our study also suggests that ICT IgM might now be considered for use as a point-of-care test. For example, a combination of the Panbio ICT IgM and presence of an eschar has a PPV and NPV of 84.9% and 93.0% respectively, representing a test that could be used to diagnose scrub typhus with a moderately high degree of accuracy in our setting.

The high diagnostic accuracy of the ICT IgM could be due to the different antigen used in the test [13]. The ICT IgM used a truncated recombinant 56-kDa protein from a single *O. tsutsugamushi* Karp strain, which was purified and may explain the high specificity of PanBio ICT [28]. The IFA IgM used whole-cell antigen from three strains of *O. tsutsugamushi* (the Karp, Kato and Gilliam strains), and these whole-cell antigens may cross react to antibodies from other diseases. We found that although ICT IgM used the 56-kDa protein from a single Karp strain, the sensitivity of ICT IgM is comparable to the IFA IgM which used whole-cell antigen from the Karp, Kato and Gilliam strains. It is possible that the ICT IgM is also positive for scrub typhus patients infected with strains other than Karp. This is supported by the previous study showing that the 56-kDa protein from a Karp strain reacts to immune serum of mice infected with Kato, Gilliam and other strains [37].

The high diagnostic accuracy of presence of an eschar in our study could be explained because other diseases that can cause eschars, such as anthrax and spotted fever group rickettsioses are not common in Thailand [38]. Brown spider bites may cause wounds with identical appearance as an eschar; however, the scrub typhus-related eschar is painless. Therefore, it is still recommended that clinicians in scrub typhus endemic areas conduct an intensive search for painless eschars in patients with febrile illness.

This study has some limitations. The sample size in our study was small. We could not identify other data sets of suspected scrub typhus to prove whether the result observed in our study is generalisable in different settings. We could not evaluate the effects of other factors such as duration of symptoms, antimicrobial received prior to presentation, and different cut-off values of IFA IgM in our model.

We conclude that the current reference tests (STIC and IFA IgM) are imperfect in our setting, and there is an urgent need for improvement of serological reference tests for scrub typhus. Future evaluation of alternative diagnostic tests for acute scrub typhus should be done using a carefully selected set of diagnostic tests and appropriate statistical models.

Supporting Information

S1 Table. Description and model selection criteria.
(DOCX)

S2 Table. Prevalence, sensitivities and specificities for diagnostic tests for scrub typhus estimated by Bayesian latent class models.
(DOCX)

S3 Table. Prevalence, sensitivities and specificities for a diagnosis of scrub typhus using Bayesian latent class models with non-informative, sceptical and enthusiastic priors.
(DOCX)

S1 Text. The data set for Bayesian latent class models.
(DOCX)

S2 Text. WinBUGS models.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: CL DHP SDB AL PK WC VW NPJD BSC DL. Analyzed the data: CL DL. Wrote the paper: CL DHP SDB VW NPJD BSC DL.

References

1. Phongmany S, Rolain JM, Phetsouvanh R, Blacksell SD, Soukkhaseum V, et al. (2006) Rickettsial infections and fever, Vientiane, Laos. Emerg Infect Dis 12: 256–262. PMID: 16494751

2. Choi YJ, Jang WJ, Ryu JS, Lee SH, Park KH, et al. (2005) Spotted fever group and typhus group rickettsioses in humans, South Korea. Emerg Infect Dis 11: 237–244. PMID: 15752441

3. Yasunaga H, Horiguchi H, Kuwabara K, Hashimoto H, Matsuda S (2011) Delay in tetracycline treatment increases the risk of complications in Tsutsugamushi disease: data from the Japanese Diagnosis Procedure Combination database. Internal medicine 50: 37–42. PMID: 21212571

4. Wang CC, Liu SF, Liu JW, Chung YH, Su MC, et al. (2007) Acute respiratory distress syndrome in scrub typhus. Am J Trop Med Hyg 76: 1148–1152. PMID: 17556627

5. Thap LC, Supanaranond W, Treeprasertsuk S, Kitvatanachai S, Chinprasatsak S, et al. (2002) Septic shock secondary to scrub typhus: characteristics and complications. The Southeast Asian journal of tropical medicine and public health 33: 780–786. PMID: 12757226

6. Kim DM, Won KJ, Park CY, Yu KD, Kim HS, et al. (2007) Distribution of eschars on the body of scrub typhus patients: a prospective study. Am J Trop Med Hyg 76: 806–809. PMID: 17488895

7. Shiao CC, Lin SY (2011) Eschar: a clue to scrub typhus. CMAJ 183: E1152. doi: 10.1503/cmaj.101929 PMID: 21911554

8. Koh GC, Maude RJ, Paris DH, Newton PN, Blacksell SD (2010) Diagnosis of scrub typhus. Am J Trop Med Hyg 82: 368–370. doi: 10.4269/ajtmh.2010.09-0233 PMID: 20207857

9. Prakash JA, Kavitha ML, Mathai E (2011) Nested polymerase chain reaction on blood clots for gene encoding 56 kDa antigen and serology for the diagnosis of scrub typhus. Indian journal of medical microbiology 29: 47–50. doi: 10.4103/0255-0857.76524 PMID: 21304195

10. Kim DM, Park G, Kim HS, Lee JY, Neupane GP, et al. (2011) Comparison of conventional, nested, and real-time quantitative PCR for diagnosis of scrub typhus. J Clin Microbiol 49: 607–612. doi: 10.1128/JCM.01216-09 PMID: 21068287

11. Paris DH, Aukkanit N, Jenjaroen K, Blacksell SD, Day NP (2009) A highly sensitive quantitative real-time PCR assay based on the groEL gene of contemporary Thai strains of Orientia tsutsugamushi. Clinical microbiology and infection 15: 488–495. doi: 10.1111/j.1469-0691.2008.02671.x PMID: 19416296

12. Paris DH, Blacksell SD, Stenos J, Graves SR, Unsworth NB, et al. (2008) Real-time multiplex PCR assay for detection and differentiation of rickettsiae and orientiae. Transactions of the Royal Society of Tropical Medicine and Hygiene 102: 186–193. PMID: 18093627

13. Blacksell SD, Jenjaroen K, Phetsouvanh R, Tanganuchitcharnchai A, Phouminh P, et al. (2010) Accuracy of rapid IgM-based immunochromatographic and immunoblot assays for diagnosis of acute scrub typhus and murine typhus infections in Laos. Am J Trop Med Hyg 83: 365–369. doi: 10.4269/ajtmh.2010.09-0534 PMID: 20682883

14. Tay ST, Kamalanathan M, Rohani MY (2003) Detection of rickettsial antibodies using Weil-Felix (OXK and OX19) antigens and the indirect immunoperoxidase assay. Southeast Asian J Trop Med Public Health 34: 171–174. PMID: 12971531

15. Coleman RE, Sangkasuwan V, Suwanabun N, Eamsila C, Mungviriya S, et al. (2002) Comparative evaluation of selected diagnostic assays for the detection of IgG and IgM antibody to Orientia tsutsugamushi in Thailand. Am J Trop Med Hyg 67: 497–503. PMID: 12479551

16. Blacksell SD, Bryant NJ, Paris DH, Doust JA, Sakoda Y, et al. (2007) Scrub typhus serologic testing with the indirect immunofluorescence method as a diagnostic gold standard: a lack of consensus leads to a lot of confusion. Clin Infect Dis 44: 391–401. PMID: 17205447

17. Paris DH, Blacksell SD, Nawtaisong P, Jenjaroen K, Teeraratkul A, et al. (2011) Diagnostic accuracy of a loop-mediated isothermal PCR assay for detection of Orientia tsutsugamushi during acute Scrub Typhus infection. PLoS Negl Trop Dis 5: e1307. doi: 10.1371/journal.pntd.0001307 PMID: 21931873

18. Blacksell SD, Paris DH, Chierakul W, Wuthiekanun V, Teeratakul A, et al. (2012) Prospective evaluation of commercial antibody-based rapid tests in combination with a loop-mediated isothermal amplification PCR assay for detection of Orientia tsutsugamushi during the acute phase of scrub typhus infection. Clin Vaccine Immunol 19: 391–395. doi: 10.1128/CVI.05478-11 PMID: 22219313

19. Dendukuri N, Joseph L (2001) Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 57: 158–167. PMID: 11252592

20. Limmathurotsakul D, Turner EL, Wuthiekanun V, Thaipadungpanit J, Suputtamongkol Y, et al. (2012) Fool's Gold: Why Imperfect Reference Tests Are Undermining the Evaluation of Novel Diagnostics: A Reevaluation of 5 Diagnostic Tests for Leptospirosis. Clin Infect Dis 55(3):322–31. doi: 10.1093/cid/cis403 PMID: 22523263

21. Limmathurotsakul D, Jamsen K, Arayawichanont A, Simpson JA, White LJ, et al. (2010) Defining the true sensitivity of culture for the diagnosis of melioidosis using Bayesian latent class models. PLoS One 5(8): e12485. doi: 10.1371/journal.pone.0012485 PMID: 20830194

22. Limmathurotsakul D, Chantratita N, Teerawattanasook N, Piriyagitpaiboon K, Thanwisai A, et al. (2011) Enzyme-linked immunosorbent assay for the diagnosis of melioidosis: better than we thought. Clin Infect Dis 52: 1024–1028. doi: 10.1093/cid/cir080 PMID: 21460318

23. Trikalinos TA, Balion CM (2012) Options for Summarizing Medical Test Performance in the Absence of a "Gold Standard". Rockville MD.

24. Lim C, Wannapinij P, White L, Day NP, Cooper BS, et al. (2013) Using a web-based application to define the accuracy of diagnostic tests when the gold standard is imperfect. PLoS One 8: e79489. doi: 10.1371/journal.pone.0079489 PMID: 24265775

25. Luksameetanasan R, Blacksell SD, Kalambaheti T, Wuthiekanun V, Chierakul W, et al. (2007) Patient and sample-related factors that effect the success of in vitro isolation of Orientia tsutsugamushi. Southeast Asian J Trop Med Public Health 38: 91–96. PMID: 17539252

26. Horinouchi H, Murai K, Okayama A, Nagatomo Y, Tachibana N, et al. (1996) Genotypic identification of Rickettsia tsutsugamushi by restriction fragment length polymorphism analysis of DNA amplified by the polymerase chain reaction. Am J Trop Med Hyg 54: 647–651. PMID: 8686786

27. Jiang J, Chan TC, Temenak JJ, Dasch GA, Ching WM, et al. (2004) Development of a quantitative real-time polymerase chain reaction assay specific for Orientia tsutsugamushi. Am J Trop Med Hyg 70: 351–356. PMID: 15100446

28. Ching WM, Rowland D, Zhang Z, Bourgeois AL, Kelly D, et al. (2001) Early diagnosis of scrub typhus with a rapid flow assay using recombinant major outer membrane protein antigen (r56) of Orientia tsutsugamushi. Clin Diagn Lab Immunol 8: 409–414. PMID: 11238230

29.   Joseph L, Gyorkos TW, Coupal L (1995) Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am J Epidemiol 141: 263–272. PMID: 7840100

30.   Zhuo X, Obuchowski N, McClish D (2008) Statistical Methods in Diagnostic Medicine: Wiley-Blackwell. pp. 359–395.

31.   Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS- A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing 10: 325–337.

32.   Sonthayanon P, Chierakul W, Wuthiekanun V, Limmathurotsakul D, Amornchai P, et al. (2013) Molecular Confirmation of Co-Infection by Pathogenic *Leptospira* spp. and *Orientia tsutsugamushi* in Patients with Acute Febrile Illness in Thailand. Am J Trop Med Hyg 89: 797–799. doi: 10.4269/ajtmh.13-0402 PMID: 24002486

33.   Liu YX, Feng D, Suo JJ, Xing YB, Liu G, et al. (2009) Clinical characteristics of the autumn-winter type scrub typhus cases in south of Shandong province, northern China. BMC infectious diseases 9: 82. doi: 10.1186/1471-2334-9-82 PMID: 19493361

34.   Reisner BS, DiBlasi J, Goel N (1999) Comparison of an enzyme immunoassay to an indirect fluorescent immunoassay for the detection of antinuclear antibodies. American journal of clinical pathology 111: 503–506. PMID: 10191770

35.   Phetsouvanh R, Thojaikong T, Phoumin P, Sibounheuang B, Phommasone K, et al. (2013) Inter- and intra-operator variability in the reading of indirect immunofluorescence assays for the serological diagnosis of scrub typhus and murine typhus. Am J Trop Med Hyg 88: 932–936. doi: 10.4269/ajtmh.12-0325 PMID: 23478577

36.   Phommasone K, Paris DH, Anantatat T, Castonguay-Vanier J, Keomany S, et al. (2013) Concurrent Infection with Murine Typhus and Scrub Typhus in Southern Laos-the Mixed and the Unmixed. PLoS Negl Trop Dis 7: e2163. doi: 10.1371/journal.pntd.0002163 PMID: 24009783

37.   Ching WM, Wang H, Eamsila C, Kelly DJ, Dasch GA (1998) Expression and refolding of truncated recombinant major outer membrane protein antigen (r56) of Orientia tsutsugamushi and its use in enzyme-linked immunosorbent assays. Clin Diagn Lab Immunol 5: 519–526. PMID: 9665960

38.   Parola P, Miller RS, McDaniel P, et al. (2003) Emerging rickettsioses of the Thai-Myanmar border. Emerg Infect Dis 9(5): 592–5. PMID: 12737744