Diagnostic Microbiology and Infectious Disease 78 (2014) 398–403

Contents lists available at ScienceDirect

Diagnostic Microbiology and Infectious Disease

journal homepage: www.elsevier.com/locate/diagmicrobio

Commercial nucleic acid amplification tests in tuberculous meningitis—a meta-analysis☆,☆☆

Regan S. Solomons [a,*], Sabine L. van Elsland [a,b], Douwe H. Visser [b], Kim G.P. Hoek [c], Ben J. Marais [d], Johan F. Schoeman [a], Anne M. van Furth [b]

[a] Department of Pediatrics and Child Health, Faculty of Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg 7505, Cape, South Africa
[b] Department of Pediatric Infectious Diseases and Immunology, Vrije Universiteit Medical Center, PO Box 7057, 1007 MB, Amsterdam, The Netherlands
[c] Division of Medical Microbiology, Faculty of Health Sciences, National Health Laboratory Service and University of Stellenbosch, Tygerberg 7505, Western Cape, South Africa
[d] Sydney Emerging Infectious Diseases and Biosecurity Institute (SEIB) and the Children's Hospital, Westmead. C29 - Children's Hospital Westmead, The University of Sydney, Westmead, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 15 September 2013
Received in revised form 29 December 2013
Accepted 2 January 2014
Available online 10 January 2014

Keywords:
Central nervous system
Tuberculosis
Nucleic acid amplification tests
Diagnostic accuracy

ABSTRACT

Although nucleic acid amplification tests (NAATs) promise a rapid, definitive diagnosis of tuberculous meningitis, the performance of first-generation NAATs was suboptimal and variable. We conducted a meta-analysis of studies published between 2003 and 2013, using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool to evaluate methodological quality. The diagnostic accuracy of newer commercial NAATs was assessed. Pooled estimates of diagnostic accuracy for commercial NAATs measured against a cerebrospinal fluid *Mycobacterium tuberculosis* culture-positive gold standard were sensitivity 0.64, specificity 0.98, and diagnostic odds ratio 64.0. Heterogeneity was limited; P value = 0.147 and I^2 = 33.85%. The Xpert MTB/RIF® test was evaluated in 1 retrospective study and 4 prospective studies, with pooled sensitivity 0.70 and specificity 0.97. The QUADAS-2 tool revealed low risk of bias, as well as low concerns regarding applicability. Heterogeneity was pronounced among studies of in-house tests. Commercial NAATs proved to be highly specific with greatly reduced heterogeneity compared to in-house tests. Sub-optimal sensitivity remains a limitation.

1. Introduction

In 1993, the World Health Organization (WHO) declared tuberculosis (TB) a global public health emergency, with an estimated 7–8 million cases and 1.3–1.6 million TB deaths per year. By 2012, the situation has improved in many areas, but absolute numbers remain virtually unchanged with an estimated 8.7 million new cases and 1.4 million TB deaths (WHO, 2012). Central nervous system involvement, mostly tuberculous meningitis (TBM), accounted for approximately 1% of all TB cases. (Rock et al., 2008). In fact, TBM has been reported as the most common form of meningitis diagnosed in children from TB endemic areas with access to expanded program of vaccination vaccines, including *Haemophilus influenza* type-B and pneumococcal vaccination (Wolzak et al., 2012). Delayed diagnosis of TBM is universally associated with poor treatment outcome (Thwaites et al., 2004).

The early clinical presentation of TBM is often non-specific with symptoms such as cough, loss of weight, fever, vomiting, and malaise. As the disease progresses, more specific features such as meningism, focal neurological signs, convulsions, and depressed level of consciousness occur (van Well et al., 2009). TBM outcome is often poor despite adequate anti-mycobacterial therapy, due to irreversible damage preceding delayed diagnosis and ongoing immune-mediated pathology on treatment. Early treatment initiation is critical to reduce TBM-associated morbidity, mortality, and healthcare costs, emphasizing the importance of early and accurate diagnosis (Garg, 1999; Schoeman et al., 2002).

Culture of *Mycobacterium tuberculosis* (*M.tb*) from cerebrospinal fluid (CSF) is regarded as the most definitive diagnosis, although this is rarely attained. TBM is a paucibacillary disease. This could explain that direct microscopy for acid-fast bacilli in CSF is rarely positive (Thwaites et al., 2000), while mycobacterial culture may take up to 42 days and has limited sensitivity (<50%) compared to clinical criteria (Hosoglu et al., 2002; Jönsson and Ridell, 2003; van Well et al., 2009). In clinical practice, the diagnosis of TBM is usually based on a combination of clinical, laboratory, and radiological findings. The use of uniform case definition categories has been proposed for research purposes (Marais et al., 2010) with “definite TBM” defined as a positive CSF *M.tb* culture and/or commercial nucleic acid amplification test (NAAT).

NAATs have been introduced to provide rapid TB diagnosis and enhanced sensitivity compared to smear microscopy (Caws et al., 2000; Pfyffer et al., 1996; Rafi and Naghily, 2003; Reischl et al., 1998;

☆ This manuscript, or parts thereof, has not been published previously, in any form.
☆☆ Conflict of interests: None of the authors have any conflict of interests.
* Corresponding author. Tel.: +27-21-938-9891; fax: +27-21 938-4858.
E-mail address: regan@sun.ac.za (R.S. Solomons).

0732-8893/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.diagmicrobio.2014.01.002

Thwaites et al., 2004). Although primarily developed for the analysis of respiratory specimens, these methods are often used in non-respiratory specimens as well (Bonington et al., 1998; Caws et al., 2000; Chedore and Jamieson, 2002; Pai et al., 2003; Pfyffer et al., 1996). They are presumed to be highly specific (Brisson and Aznar, 1991; Marais et al., 2010), since they detect *M.tb*-specific DNA sequences such as the IS6110 insertion element, MBP64, 65-kDa antigen, and the rpoB region (Blakemore et al., 2010; Rafi et al., 2006).

In 2003, a systematic review evaluated the test accuracy of NAATs in the diagnosis of TBM (Pai et al., 2003). The authors included 49 studies published between 1990 and 2002; both commercial and in-house NAATs were evaluated. The 14 studies with commercial NAATs revealed a pooled sensitivity and specificity of 56% and 98%, respectively. Summary accuracy measures of 35 studies with in-house NAATs could not be determined due to heterogeneity of the tests. Reasons for heterogeneity included: 1) inadequate standardization of laboratory techniques, 2) use of highly variable reference standards, 3) and small patient numbers with limited statistical power (Thwaites et al., 2004). The review concluded that commercial NAATs provided valuable information when positive, but due to poor sensitivity, a negative test did not exclude TBM (Pai et al., 2003). This finding motivated the inclusion of a positive commercial NAAT as a marker of "definite TBM" in a proposed uniform TBM case definition for use in clinical research (Marais et al., 2010).

Since then, many additional studies evaluated the use of commercial NAATs in the diagnosis of TBM, but no updated meta-analysis has been performed. We performed a systematic review of all recent studies (published since 2003) that evaluated the use of NAATs to diagnose TBM, with particular emphasis on commercial tests including the Xpert MTB/RIF® test.

2. Methods

We identified all studies published between January 2003 and April 2013 from the following online databases: PubMed (MedLine), Web of Knowledge, Scopus, and LILACS. Search terms used were: "Tuberculosis, Central Nervous System", "Tuberculoma, Intracranial", "Tuberculosis", "Mycobacterium tuberculosis", "Extrapulmonary tuberculosis", "Tuberculous meningitis", "Tuberculous pachymeningitis", "Central nervous system" and/or "Kochs disease" and "Polymerase Chain Reaction", "Ligase chain reaction", "GeneXpert" and/or "Nucleic acid amplification testing". Only articles written in English were included. Case reports and review articles were excluded. Studies with less than 10 subjects were also excluded. References of selected articles were reviewed to identify additional eligible studies. Three reviewers (RS, SLvE, and AMvF) independently evaluated study inclusion; differences were resolved by consensus.

2.1. Data extraction

Two reviewers (RS and SLvE) independently extracted data including number of cases, number of controls, reference standard used, and type of NAAT evaluated. Diagnostic odds ratios were extracted or calculated from the data provided. Differences were resolved by consensus. Methodological quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (Fontela et al., 2009; Whiting et al., 2003; Whiting et al., 2011).

2.2. Statistical analysis

Data analysis was performed using Statistical Package for the Social Sciences version 19 (SPSS Inc, Chicago, IL, USA), Comprehensive Meta Analysis version 2 (Biostat, Eaglewood, NJ, USA), and Meta-DiSc (Unit of Clinical Biostatistics, Ramón y Cajal Hospital, Madrid, Spain). Sensitivity, specificity, and diagnostic odds ratio (DOR) were computed for each of the included studies. Pooled summary effect estimates were calculated, using a random effects model. Where both CSF culture and clinical criteria were analyzed separately as reference standards, only the studies with CSF culture as the reference standard were included. When articles evaluated more than 1 NAAT or more than 1 quality measure, these were analyzed separately.

Receiver operating characteristic (ROC) curves based on either the regression of logit sensitivity on specificity, the regression of logit specificity on sensitivity, or an orthogonal regression line by minimizing the perpendicular distances were derived. These lines were transformed back to the original ROC scale to obtain a summary ROC (SROC) curve. Derived logit estimates of sensitivity, specificity, and respective variances were used to construct a hierarchical SROC curve with these summary estimates. The area under the curve serves as a global measure of test performance; a value of 1 indicates perfect accuracy (Dwamena, 2007). Heterogeneity was assessed by applying the χ^2 homogeneity test to calculated odds ratios (as a single measure) and determining I^2, with values of more than 50% indicating heterogeneity (Abroug et al., 2011; Dwamena, 2007; Greco et al., 2003). Statistical significance was set at 0.05 for heterogeneity testing.

3. Results

The study selection process is summarized in Fig. 1. The literature search revealed 1125 potential articles, which was narrowed down to 69 articles after title screening. This was narrowed down further to 62 articles after abstract screening. Thirty-six articles were excluded after screening the text, and 4 articles added after cross referencing. Ten studies in 8 articles, describing commercial tests, were selected; 40 studies in 22 articles describing in-house NAATs were tabulated separately (Abroug et al., 2011; Bhigjee et al., 2007; Blakemore et al., 2010; Brisson and Aznar, 1991; Causse et al., 2011; Chaidir et al., 2012; Desai et al., 2006; Deshpande et al., 2007; Dora et al., 2008; Haldar et al., 2009; Huang et al., 2009; Iacob and Banica, 2009; Johansen et al., 2004; Juan et al., 2006; Kulkarni et al., 2005; Kusum et al., 2011; Malbruny et al., 2011; Nagdev et al., 2010a; Nagdev et al., 2010b; Nagdev et al., 2011; Patel et al., 2013; Pfyffer et al., 1996; Quan et al., 2006; Rafi and Naghily, 2003; Rafi et al., 2007; Rana et al., 2010; Sastry et al., 2013; Sharma et al., 2010; Takahashi et al., 2008; Thwaites et al., 2000; Thwaites et al., 2004; Tortoli et al., 2012; Vadwai et al., 2011) (Supplementary Table 1). Reference standards used in the 10 studies evaluating commercial NAATs included a positive CSF *M.tb* culture in 9 (90%) and clinical criteria in 1 (10%). To avoid misleading results, only the 9 commercial studies with positive CSF *M.tb* culture as the reference standard were analyzed. A variety of DNA extraction techniques and target sequences were used. Table 1 summarizes key characteristics of the commercial NAAT studies. Fig. 2 reflects formal assessment of the 4 study domains evaluated by the QUADAS-2 tool; inter-reviewer variability using the tool was 10.6% (Whiting et al., 2011).

Summary test accuracy estimates for the 9 commercial NAATS evaluated were sensitivity 0.64 (95% confidence interval [CI] 0.56–0.72), specificity 0.98 (95% CI 0.96–0.99), positive likelihood ratio 20.36 (95% CI 11.29–36.73), negative likelihood ratio 0.39 (95% CI 0.30–0.53), and DOR 64.0 (95% CI 26.9–152.1). Heterogeneity was limited; P value = 0.147 and I^2 = 33.85%. Table 2 shows heterogeneity testing after stratification of the commercial NAATs based on study design, prospective nature, and Xpert MTB/RIF testing. Fig. 3 provides an overview of sensitivities and specificities of commercial NAATs in forest plot format. Fig. 4 presents the SROC curve for the commercial NAAT studies combined, with the respective studies presented as circles. The area under the curve (AUC) for all commercial tests combined was 0.92.

Summary test accuracy estimates for the 40 in-house tests revealed sensitivity of 0.73 (95% CI 0.71–0.75), specificity of 0.92 (95% CI 0.90–0.93), positive likelihood ratio of 9.56 (95% CI 6.61–13.84), negative likelihood ratio of 0.27 (95% CI 0.20–0.35), and DOR of 40.6 (95% CI 26.6–61.9). Heterogeneity was pronounced; P value =