The accuracy and efficacy of screening tests for Chlamydia trachomatis: a systematic review

EMMA J. WATSON, ALLAN TEMPLETON, IAN RUSSELL*, JORMA PAAVONEN†, PER-ANDERS MARDH‡, ANGELIKA STARY§ and BABIL STRAY PEDERSON

Department of Obstetrics and Gynaecology, Aberdeen University, Aberdeen, Scotland, *Department of Health Sciences and Clinical Evaluation, University of York, York, UK, †Department of Obstetrics and Gynaecology, Helsinki University, Helsinki, Finland, Department of Clinical Bacteriology, University of Uppsala, Uppsala, Sweden, §Ambulatorium fur Pilzinfektionen, Vienna, Austria and ‡Department of Obstetrics and Gynaecology, Oslo University Hospital, Oslo, Norway

Screening women for lower genital tract infection with Chlamydia trachomatis is important in the prevention of pelvic inflammatory disease, ectopic pregnancy and infertility. This systematic review aims to state clearly which of the available diagnostic tests for the detection of C. trachomatis would be most effective in terms of clinical effectiveness. The review included all studies published from 1990 onward that evaluated diagnostic tests in asymptomatic, young, sexually active populations. Medline and Embase were searched electronically and key journals were hand-searched. Further studies were identified through the Internet and contact with experts in the field. All studies were reviewed by two reviewers and were scored by Irwig’s assessment criteria. Additional quality assessment criteria included a documented sexual history and recording of previous chlamydial infection. The reviews were subjected to meta-analysis and meta-regression. The 30 studies that were included examined three types of DNA-based test – ligase chain reaction (LCR), PCR and gene probe – as well as enzyme immuno-assay (EIA). The results showed that while specificities were high, sensitivities varied widely across the tests and were also dependent on the specimen tested. Pooled sensitivities for LCR, PCR, gene probe and EIA on urine were 96.5%, 85.6%, 92% and 38%, respectively, while on cervical swabs the corresponding sensitivities of PCR, gene probe and EIA were 88.6%, 84% and 65%. Meta-analysis demonstrated that DNA amplification techniques performed best for both urine and swabs in low prevalence populations. We conclude that nucleic acid amplification tests used on non-invasive samples such as urine are more effective at detecting asymptomatic chlamydial infection than conventional tests, but there are few data to relate a positive result with clinical outcome.

Introduction

Chlamydia trachomatis is the most common bacterial sexually transmitted infection in Western Europe and women carry the main burden of this disease. The case for national screening programmes has been made but there is a need for more data to show how this can be done most effectively. This review focuses on the best test for the detection of C. trachomatis when used in a screening context.

C. trachomatis is an obligate intracellular gram-negative bacterium. Infection with this agent can be asymptomatic in up to 80% of women [1], which can make diagnosis and detection difficult. Chlamydia has its highest prevalence amongst young men and women. More than 13.5% of women <25 years old have lower genital tract infection, reducing to <4.9% in women over 25 [2].

Left undetected and untreated chlamydia can ascend the upper genital tract, causing inflammation and scarring in both the female and the male reproductive tract [3]. Many reports now indicate that it is the major causative agent in the development of pelvic inflammatory disease (PID) in women. The major sequelae of
The asymptomatic nature of chlamydial infection makes screening essential if control of this infection is to be achieved. In Sweden, policies to reduce the prevalence of infection have been in place since the 1980s and rates of chlamydial infection and its complications have fallen. Because of the severity of the complications of infection with chlamydia and their implications in health economic terms, several other countries including the UK, France, Holland and Finland have now taken action to reduce the prevalence of this infection.

To be effective, a national screening programme must use the most accurate diagnostic test available. Currently there is little or no consensus on which diagnostic tool to use as a screening device or which sampling method to use [6]. The ‘gold standard’ for detection of chlamydia is still considered by many to be cell culture [7]. Culture is 100% specific, but estimates of sensitivity are as low as 50%. The majority of laboratories have moved away from culture, as it is expensive, time-consuming and technically difficult. The use of an expanded gold standard, commonly consistent results with two non-culture techniques, is considered to be more useful as a research tool, but most laboratories use only one non-culture method as their routine test for detection of chlamydia [8].

There is considerable variation in health-care professionals’ knowledge of chlamydial infection. Furthermore, there is a degree of confusion as to which diagnostic tests and sampling methods should be employed [9]. Within the last decade, tests that are based on nucleic acid amplification have become available. They appear to be highly sensitive and specific and these tests have the added advantage that they are effective for use with non-invasive specimens such as urine and vulval swabs [10].

It is important that any test adopted in a national screening programme can be used in the primary care setting by practitioners without the need for expensive training. Furthermore, the ideal test for a screening programme should have the capacity to be used in both sexes. The reason for this is that any screening programme, while initially being directed at young women, should have the potential to involve men, both for contact tracing and for possible expansion of the programme. The reasons for eventual expansion may be three-fold. Firstly, so as not to stigmatise women’s sexuality [11], secondly to involve men in the health-care system and not to exclude them, and finally to screen only half the population affected by a condition would be ineffective. Clad et al. [12] demonstrated that screening women detected only 54% of infected couples, but screening men alone detected 81% of infected couples.

The objective of this systematic review is to state clearly which available diagnostic test is the most accurate and effective when used in young, asymptomatic, sexually active populations. For the results to be valid and transferable it is important that the populations examined in the review are similar to the population in which the test would eventually be used in a screening programme. Diagnostic tests perform differently in high and low risk populations. As prevalence varies, so does the positive predictive value [13]. Irwig et al. [14] have set out guidelines to assess studies that examine the usefulness of diagnostic tests. These guidelines have been used to evaluate the screening tests included in this systematic review (Table 1).

### Methods

#### Search strategy

Studies from 1990 onwards that assessed the effectiveness of tests used to diagnose *C. trachomatis* infection were located on the electronic databases Medline, CINAHL and Embase. Relevant journals were hand-searched. The Internet was explored with Lycos, Alta Vista and Excite as search engines and Medscape was also used to detect information and conference proceedings. The bibliographies of included studies were also searched for relevant articles. Experts in the field were contacted by electronic mail for study information. The search included a filter and headings such as chlamydia, exp. diagnosis and mass screening. One reviewer (E.J.W.) examined the titles and abstracts on three occasions. The subset of articles that focused on asymptomatic populations and therefore, was, relevant to the meta-analysis was further evaluated by two reviewers (E.J.W. and J.S.W.) with the inclusion and exclusion criteria. Thirty-two studies that evaluated diagnostic tests for *C. trachomatis* infection were included in the systematic review.

#### Selection criteria

Any trial from 1990 onwards that evaluated methods of detecting urogenital infection with chlamydia was

### Table 1. Irwig’s [14] criteria for assessing studies examining diagnostic tests

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Was a recognised gold standard used?</td>
</tr>
<tr>
<td>2.</td>
<td>Were the results read blind and independently of the reference?</td>
</tr>
<tr>
<td>3.</td>
<td>Were the patients that were examined by the test consecutive patients?</td>
</tr>
<tr>
<td>4.</td>
<td>Was there verification of negative results?</td>
</tr>
<tr>
<td>5.</td>
<td>Was the test technique well described?</td>
</tr>
</tbody>
</table>
included. The lack of randomised controlled studies available in this field meant that only comparative studies were examined.

Population

The patients included in the study had to be sexually active young men or women with no symptoms of chlamydia infection. It was important that they were asymptomatic, as there is evidence that diagnostic tests for chlamydia perform better among patients who are asymptomatic, perhaps because of an increased elementary body load. The age range of the populations in the included studies was 14–40 years.

Setting

Papers that described populations with a low prevalence, taken as <5% [15], were included regardless of the setting and studies that were set in primary care or a family planning clinic were included regardless of the prevalence of chlamydia infection.

Intervention

The diagnostic tests examined for detection of C. trachomatis were nucleic acid amplification techniques (PCR and LCR), gene probes (GP), enzyme immunoassay (EIA) and direct immunofluorescence (DFA). The leucocyte esterase test (LET) was also examined to determine if it would be a useful screening tool. All were compared to culture or an expanded gold standard. The sensitivity of culture was calculated by comparing it with two non-culture techniques. All methods of sample collection were reviewed.

Outcome

Detection of chlamydia in the lower genital tract of men and women.

Study quality

The review process was not blind to study authorship, as there is no proof that this adds any quality measure to the review. The quality of the studies will affect the validity of the result and, therefore, study quality was assessed by the criteria suggested by Irwig et al. [14]. Studies were excluded if study design was considered to be poor, as judged by an Irwig score of <5 out of 10.

Statistical methods

The statistical packages used to evaluate the diagnostic tests included meta-test software kindly supplied by J. Lau (JLau1@Lifespan.org) and meta-analysis software in Rev man 4.01 (Update Software, Oxford, UK). SPSS (Chicago, IL, USA) was also used.

Results

This systematic review and the subsequent meta-analysis included all methods for the diagnosis of urogenital chlamydia shown in Table 2, compared to a gold standard. The gold standard is culture for chlamydia performed as described by Mardh et al. [16] or chlamydia diagnosed by two non-culture tests, now known as the expanded gold standard [17]. The focus of the meta-analysis was to provide an overall summary of diagnostic test accuracy for detection of asymptomatic chlamydial infection. The leucocyte esterase test (LET), while not a diagnostic test for C. trachomatis, was included in this review as it has been evaluated for screening purposes. Several hundred study abstracts were examined for potential inclusion; 74 studies were identified for further evaluation.

Study characteristics

Each study was evaluated by the guidelines set out by Irwig et al. [14] as indicated previously and the results are shown in Table 3. Thirty-two studies [17–48] were identified for possible inclusion in the meta-analysis, but two were excluded because of study quality [47, 48]. Forty-two studies did not meet the inclusion criteria when the papers were examined (Table 4) [49–90]. The gold standard was tissue culture of a cervical or urethral swab in 28 of the studies and 2 non-culture techniques used either on urine or cervical swabs in the remainder. When culture was the modality under examination, the sensitivity was calculated by comparing the results with two non-culture techniques. Six (19%) of the studies indicated that results were read blind, i.e., without knowledge of the result of the gold

Table 2. Description of methods available for the detection of C. trachomatis

<table>
<thead>
<tr>
<th>Test method</th>
<th>Specimen type</th>
<th>Skill and time required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleic acid amplification (PCR and LCR)</td>
<td>Vulval, cervical or urethral swab; first-void urine</td>
<td>Needs specialised equipment and qualified staff: 4–24 h</td>
</tr>
<tr>
<td>Gene probe</td>
<td>Vulval, cervical or urethral swab</td>
<td>Easier to perform than culture; easy to read: 3 h</td>
</tr>
<tr>
<td>Enzyme immunoassay (EIA)</td>
<td>Urethral swab; cervical swab</td>
<td>Simple to process; subjective results; skilled staff: 30 min</td>
</tr>
<tr>
<td>Direct immunofluorescence (DFA)</td>
<td>Urethral swab; cervical swab</td>
<td>Specialised culture medium and culture conditions; skilled staff: 72 h</td>
</tr>
<tr>
<td>Culture</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
standard. However, many of the tests evaluated were assessed as positive or negative with automated equipment, which by its own nature is blinded. Unless it was stated in the methodology that the results were read blind this was recorded as unknown. One-third of the studies did not recruit consecutive patients. This may be due in part to the nature of the disease being tested, as it is often difficult to gain consent from patients for an invasive test for a sexually transmitted disease. Thirty of the studies performed verification of negative results in addition to the positive results. In all these studies the technique and study methodology were well described.

The investigators had taken and documented a sexual history in only seven of the studies. The validity of a sexual history has been called into question and patients will often say what they feel is expected, so while this element is interesting it was not a basis for exclusion.

### Study outcomes

Studies that reported test accuracy commonly did so in terms of sensitivity and specificity. Sensitivity is the ability of the test to correctly identify those with the disease; the specificity is the ability of the test to identify those who do not have the disease. Some of the tests examined performed better than the gold standard and this makes evaluation difficult. In this situation, tests are compared to an expanded gold standard. Data from each study were re-examined to establish sensitivity and specificity, but missing data in some studies made re-analysis difficult. A scatter plot of the tests’ sensitivity to inverse specificity was plotted for each test to illustrate the test’s accuracy and is a useful visual guide to the variability of performance (Fig. 1). An ideal test would have a sensitivity of 100% and an inverse specificity of zero. The number of false-negative results in each study was summed to calculate a pooled sensitivity for the test types by specimen type.

<table>
<thead>
<tr>
<th>Study</th>
<th>Test and sample</th>
<th>Prevalence (%)</th>
<th>Sex</th>
<th>Sample size</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Irwig score (max. 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjei 1994 [18]</td>
<td>EIA, cervix</td>
<td>15</td>
<td>M, F</td>
<td>300</td>
<td>96.5</td>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td>Bass 1993 [19]</td>
<td>PCR, cervix</td>
<td>7.8</td>
<td>F</td>
<td>1153</td>
<td>8.5</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Bassiri 1997 [20]</td>
<td>LCR, urine</td>
<td>2.4</td>
<td>F</td>
<td>3340</td>
<td>...</td>
<td>...</td>
<td>7</td>
</tr>
<tr>
<td>Bassiri 1995 [21]</td>
<td>LCR, urine</td>
<td>3.1</td>
<td>F</td>
<td>447</td>
<td>18.8</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Biro 1994 [22]</td>
<td>EIA, cervix</td>
<td>11</td>
<td>F</td>
<td>228</td>
<td>52</td>
<td>100</td>
<td>6</td>
</tr>
<tr>
<td>Biro 1994</td>
<td>EIA, cervix</td>
<td>11</td>
<td>F</td>
<td>228</td>
<td>80</td>
<td>99</td>
<td>6</td>
</tr>
<tr>
<td>Biro 1994 [22]</td>
<td>DNA probe, cervix</td>
<td>11</td>
<td>F</td>
<td>228</td>
<td>72</td>
<td>96</td>
<td>...</td>
</tr>
<tr>
<td>Bowden 1998 [23]</td>
<td>LET, urine</td>
<td>4.9</td>
<td>M</td>
<td>245</td>
<td>77.8</td>
<td>80.8</td>
<td>3</td>
</tr>
<tr>
<td>Braverman 1990 [24]</td>
<td>EIA, cervix</td>
<td>12.3</td>
<td>M</td>
<td>97</td>
<td>75</td>
<td>99</td>
<td>6</td>
</tr>
<tr>
<td>Brokenshire 1997 [25]</td>
<td>EIA, cervix</td>
<td>4.5</td>
<td>F</td>
<td>622</td>
<td>84.4</td>
<td>100</td>
<td>...</td>
</tr>
<tr>
<td>Catry 1995 [26]</td>
<td>PCR, cervix</td>
<td>1.03</td>
<td>F</td>
<td>193</td>
<td>100</td>
<td>100</td>
<td>7</td>
</tr>
<tr>
<td>Cherniakov 1997 [27]</td>
<td>LCR, urine</td>
<td>6</td>
<td>F</td>
<td>447</td>
<td>96</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Cherniakov 1994 [27]</td>
<td>EIA, urine</td>
<td>6</td>
<td>F</td>
<td>447</td>
<td>37</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Cherniakov 1994 [27]</td>
<td>EIA, cervix</td>
<td>6</td>
<td>F</td>
<td>447</td>
<td>78.3</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Chout 1995 [28]</td>
<td>PCR, cervix</td>
<td>11</td>
<td>F</td>
<td>485</td>
<td>100</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Cook 1999 [29]</td>
<td>PCR, cervix</td>
<td>2.3</td>
<td>F</td>
<td>1149</td>
<td>85</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>De Barbeyrac 1994 [30]</td>
<td>PCR, cervix</td>
<td>4.8</td>
<td>F</td>
<td>549</td>
<td>95.3</td>
<td>100</td>
<td>7</td>
</tr>
<tr>
<td>Ferrero 1998 [33]</td>
<td>PCR, cervix</td>
<td>5.4</td>
<td>F</td>
<td>717</td>
<td>100</td>
<td>96</td>
<td>8</td>
</tr>
<tr>
<td>Ferrero 1998 [31]</td>
<td>PCR, cervix</td>
<td>5.4</td>
<td>F</td>
<td>407</td>
<td>92.3</td>
<td>98.6</td>
<td>8</td>
</tr>
<tr>
<td>Gaydos 1998 [32]</td>
<td>LCR, cervix</td>
<td>5–7.3</td>
<td>F</td>
<td>465</td>
<td>88.6</td>
<td>99.7</td>
<td>...</td>
</tr>
<tr>
<td>Kay 1997 [33]</td>
<td>PCR, cervix</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Loeffelho 1992 (site B) [34]</td>
<td>EIA, cervix</td>
<td>5</td>
<td>F</td>
<td>375</td>
<td>97</td>
<td>99.7</td>
<td>9</td>
</tr>
<tr>
<td>McNagny 1992 [35]</td>
<td>LET</td>
<td>4</td>
<td>M</td>
<td>1095</td>
<td>41</td>
<td>91</td>
<td>6</td>
</tr>
<tr>
<td>Mattarazzo 1997 [36]</td>
<td>LCR and LET, urine</td>
<td>5.4–8.6</td>
<td>M, F</td>
<td>10 118</td>
<td>58.9</td>
<td>94.9</td>
<td>4</td>
</tr>
<tr>
<td>Pasternack 1996 [37]</td>
<td>PCR, cervix</td>
<td>5.9</td>
<td>F</td>
<td>666</td>
<td>82</td>
<td>99.7</td>
<td>7</td>
</tr>
<tr>
<td>Pasternack 1996 [37]</td>
<td>PCR, cervix</td>
<td>5.9</td>
<td>F</td>
<td>666</td>
<td>82</td>
<td>99.8</td>
<td>...</td>
</tr>
<tr>
<td>Pasternack 1996 [37]</td>
<td>PACE2, cervix</td>
<td>5.9</td>
<td>F</td>
<td>666</td>
<td>79.5</td>
<td>100</td>
<td>...</td>
</tr>
<tr>
<td>Paukku 1997 [38]</td>
<td>PCR, urine</td>
<td>5.6</td>
<td>F</td>
<td>1,090</td>
<td>85</td>
<td>99</td>
<td>9</td>
</tr>
<tr>
<td>Schachtler 1995 [40]</td>
<td>LCR, urine</td>
<td>4.3</td>
<td>F</td>
<td>4853</td>
<td>88.2</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Schubiner 1990 [41]</td>
<td>DFA,GEIA</td>
<td>7</td>
<td>F</td>
<td>200</td>
<td>68.80, 98.100,99</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Shafir 1993 [42]</td>
<td>EIA</td>
<td>7</td>
<td>M</td>
<td>618</td>
<td>84</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Shulmack 1994 [43]</td>
<td>PCR, cervix</td>
<td>1.3</td>
<td>F</td>
<td>993</td>
<td>99.2</td>
<td>99.8</td>
<td>9</td>
</tr>
<tr>
<td>Stary 1996 [17]</td>
<td>LCR, PCR</td>
<td>4</td>
<td>M</td>
<td>705</td>
<td>93</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Thejs 1994 [44]</td>
<td>DFA, cervix</td>
<td>4.3</td>
<td>F</td>
<td>419</td>
<td>77.8</td>
<td>99.5</td>
<td>10</td>
</tr>
<tr>
<td>Thejs 1994 [44]</td>
<td>PCR, cervix</td>
<td>4.3</td>
<td>F</td>
<td>419</td>
<td>71.4</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>Thejs 1994 [44]</td>
<td>EIA, cervix</td>
<td>4.3</td>
<td>F</td>
<td>419</td>
<td>64.7</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>Toye 1996 [45]</td>
<td>PCR, cervix</td>
<td>7.9</td>
<td>F</td>
<td>242</td>
<td>89.5</td>
<td>99.3</td>
<td>10</td>
</tr>
<tr>
<td>Vincelotte 1999 [46]</td>
<td>PCR, cervix</td>
<td>3.2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Anestad 1995 [47] and Moncada 1992 [48] excluded from Table 3 because of missing data.
The studies were analysed in subgroups according to specimen type and test, as shown in Fig. 2. Only studies involving asymptomatic populations in a primary care or low prevalence setting were examined. This enabled evaluation of the test under more searching conditions. Certain tests for the detection of chlamydia may perform best when the elementary body load is high or if, as with all tests, the prevalence of the condition is high. Studies were performed in a wide range of locations in different types of population. The mean prevalence of chlamydia infection among the populations studied was 4.5%. This reflects the expected prevalence in all age groups tested in primary care. However, the range was from <1% to 15%. This variation illustrates the difference in prevalence among primary care settings depending on risk factors and location. It may also represent differences in prevalence between countries that offer screening compared with those that do not.

To compare different tests, false-negative results were selected as a suitable measure of poor outcome, because the subjects studied were asymptomatic and failure to detect the condition could lead to serious sequelae. The false-negative rate for each test was expressed as an odds ratio (OR). An OR is commonly reported in systematic reviews and describes the likely harm an intervention may have. The OR is calculated by dividing the likelihood of a false-negative result in the test under evaluation by the likelihood of a false-negative result in the gold standard test. An OR of 1 (the neutral value) indicates that there is no difference between the evaluated test and the gold standard. An OR of >1 means that the evaluated test has more false-negative results than the gold standard, while an OR of

### Table 4. Excluded studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altaie 1992 [49]</td>
<td>Symptomatic patients</td>
</tr>
<tr>
<td>Anderson 1998 [50]</td>
<td>High risk patients</td>
</tr>
<tr>
<td>Beck-Sage 1998 [51]</td>
<td>Symptomatic patients</td>
</tr>
<tr>
<td>Blanding 1993 [52]</td>
<td>Symptomatic patients included</td>
</tr>
<tr>
<td>Bygdeman 1994 [53]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Carroll 1998 [54]</td>
<td>Wrong setting</td>
</tr>
<tr>
<td>Chan 1998 [55]</td>
<td>Wrong setting</td>
</tr>
<tr>
<td>Croftchelt 1998 [56]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Chan 1996 [57]</td>
<td>High prevalence</td>
</tr>
<tr>
<td>Davis 1998 [58]</td>
<td>Symptomatic patients</td>
</tr>
<tr>
<td>de Barbeyrac 1995 [59]</td>
<td>Unable to extract low prevalence data</td>
</tr>
<tr>
<td>Dunn 1998 [60]</td>
<td>Symptomatic patients</td>
</tr>
<tr>
<td>Dille 1993 [61]</td>
<td>Laboratory-based</td>
</tr>
<tr>
<td>Dong 1999 [62]</td>
<td>Not gold standard</td>
</tr>
<tr>
<td>Ehret 1993 [63]</td>
<td>Symptomatic patients</td>
</tr>
<tr>
<td>Gaydos 1996 [64]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Hook 1997 [65]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Jaschek 1993 [66]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Kellog 1994 [67]</td>
<td>High risk women</td>
</tr>
<tr>
<td>Lee 1995 [68]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Leonard 1992 [69]</td>
<td>Not gold standard</td>
</tr>
<tr>
<td>Miyashita 1996 [70]</td>
<td>Laboratory-based</td>
</tr>
<tr>
<td>Newhall 1999 [71]</td>
<td>Symptomatic patients</td>
</tr>
<tr>
<td>Ostergaard 1998 [72]</td>
<td>Wrong population</td>
</tr>
<tr>
<td>Pasternack 1997 [73]</td>
<td>Wrong setting</td>
</tr>
<tr>
<td>Pate 1998 [74]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Puolakkainen 1998 [75]</td>
<td>Wrong setting, moderate prevalence</td>
</tr>
<tr>
<td>Quinn 1996 [76]</td>
<td>Wrong setting</td>
</tr>
<tr>
<td>Ridgway 1996 [77]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Roymans 1996 [78]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Rampianes 1993 [79]</td>
<td>Symptomatic patients</td>
</tr>
<tr>
<td>Sanders 1994 [80]</td>
<td>Wrong setting, moderate prevalence</td>
</tr>
<tr>
<td>Schachter 1994 [81]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Sellors 1991 [82]</td>
<td>Symptomatic</td>
</tr>
<tr>
<td>Smith 1996 [83]</td>
<td>High risk</td>
</tr>
<tr>
<td>Stary 1998 [84]</td>
<td>High risk, high prevalence</td>
</tr>
<tr>
<td>Stary 1997 [85]</td>
<td>High prevalence</td>
</tr>
<tr>
<td>Tahrizi 1998 [86]</td>
<td>Symptomatic women, only compared to one other non-culture test</td>
</tr>
<tr>
<td>Tong 1996 [87]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Wiltkin 1996 [88]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Wylie 1998 [89]</td>
<td>Wrong setting, high prevalence</td>
</tr>
<tr>
<td>Woods 1994 [90]</td>
<td>Not clinically based</td>
</tr>
</tbody>
</table>

The studies were analysed in subgroups according to specimen type and test, as shown in Fig. 2. Only studies involving asymptomatic populations in a primary care or low prevalence setting were examined. This enabled evaluation of the test under more searching conditions. Certain tests for the detection of chlamydia may perform best when the elementary body load is high or if, as with all tests, the prevalence of the condition is high. Studies were performed in a wide range of locations in different types of population. The mean prevalence of chlamydia infection among the populations studied was 4.5%. This reflects the expected prevalence in all age groups tested in primary care. However, the range was from <1% to 15%. This variation illustrates the difference in prevalence among primary care settings depending on risk factors and location. It may also represent differences in prevalence between countries that offer screening compared with those that do not.

To compare different tests, false-negative results were selected as a suitable measure of poor outcome, because the subjects studied were asymptomatic and failure to detect the condition could lead to serious sequelae. The false-negative rate for each test was expressed as an odds ratio (OR). An OR is commonly reported in systematic reviews and describes the likely harm an intervention may have. The OR is calculated by dividing the likelihood of a false-negative result in the test under evaluation by the likelihood of a false-negative result in the gold standard test. An OR of 1 (the neutral value) indicates that there is no difference between the evaluated test and the gold standard. An OR of >1 means that the evaluated test has more false-negative results than the gold standard, while an OR of
<1 means that the evaluated test has less false-negative results than the gold standard.

The meta-analysis demonstrated that DNA-based tests detected more cases of asymptomatic chlamydial infections than the conventional non-culture tests. Summing the results of all included studies, 3506 chlamydia positive results were diagnosed. The OR of a false-negative result for each test examined is shown in Fig. 3. LCR on urine had the lowest number of false-negative results, with an OR of a false-negative result of 0.33 (0.13–0.8). In comparison, EIA on cervix, currently the more commonly used test, was 10 times likely to give a false-negative result at 4.10 (1.15–14.59).

Nine of the studies were commercially funded. These comprised 4 of the 13 studies that assessed PCR, 3 of the 7 that assessed LCR, one of the 3 that assessed the DNA probe and one of the 9 that assessed EIA. The meta-analysis was conducted with these studies both included and excluded. When these studies were removed from the analysis, the ORs were slightly altered for urine samples tested with LCR and cervical samples tested with PCR, as shown in Table 5.

Discussion

Many studies and literature reviews have suggested that nucleic acid amplification tests are better than other diagnostic methods for detecting chlamydia, but this has not usually been based on a rigorous comparison. A meta-analysis of papers that study test performance in a pragmatic screening setting was needed to examine this. This meta-analysis demonstrates that nucleic acid amplification techniques are superior to other methods for detecting asymptomatic chlamydial infection in a young, sexually active population. The test most often used to detect this sexually transmitted infection, the EIA, has sensitivities in the range 70–80% even when used by experts. The implications of this are that, if this were to be used in a national screening programme, 30% of infections could be missed.

By focusing on studies that evaluated diagnostic tests in asymptomatic populations this meta-analysis is relevant to screening programmes. Thirty-two studies that examined a combination of non-culture tests were included. Whereas many studies have evaluated diagnostic test performance in high risk, high prevalence populations, relatively few have studied the asympto-
<table>
<thead>
<tr>
<th>Study</th>
<th>Test</th>
<th>Gold Standard</th>
<th>OR (95% CI random)</th>
<th>Weight OR (95% Cl random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 Urine LCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chernesky 1994 [27]</td>
<td>1/27</td>
<td>15/27</td>
<td>13.2</td>
<td>0.03 [0.00,0.26]</td>
</tr>
<tr>
<td>Schachter 1995 [40]</td>
<td>30/277</td>
<td>78/277</td>
<td>35.6</td>
<td>0.36 [0.25,0.50]</td>
</tr>
<tr>
<td>Bassi 1995 [1]</td>
<td>2/16</td>
<td>7/16</td>
<td>16.5</td>
<td>0.18 [0.03,0.90]</td>
</tr>
<tr>
<td>Stary 1996 [17]</td>
<td>2/29</td>
<td>0/29</td>
<td>7.7</td>
<td>0.56 [0.25,116.77]</td>
</tr>
<tr>
<td>Gaydosh 1998 [32]</td>
<td>7/64</td>
<td>13/64</td>
<td>27.0</td>
<td>0.53 [0.19,1.45]</td>
</tr>
<tr>
<td>Subtotal (95% Cl)</td>
<td>48/413</td>
<td>112/413</td>
<td>100.0</td>
<td>0.33 [0.13,0.80]</td>
</tr>
<tr>
<td>02 Urine PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pasternak 1996 [37]</td>
<td>7/39</td>
<td>6/39</td>
<td>17.2</td>
<td>1.20 [0.36,3.97]</td>
</tr>
<tr>
<td>Toye 1996 [45]</td>
<td>5/20</td>
<td>7/20</td>
<td>15.4</td>
<td>0.62 [0.16,4.23]</td>
</tr>
<tr>
<td>Pasku 1997 [38]</td>
<td>9/61</td>
<td>6/61</td>
<td>18.4</td>
<td>1.58 [0.43,4.77]</td>
</tr>
<tr>
<td>Kay 1997 [33]</td>
<td>4/71</td>
<td>9/71</td>
<td>16.9</td>
<td>0.41 [0.12,1.40]</td>
</tr>
<tr>
<td>Bassi 1997 [20]</td>
<td>6/81</td>
<td>7/81</td>
<td>9.5</td>
<td>0.13 [0.02,1.10]</td>
</tr>
<tr>
<td>Gaydosh 1998 [32]</td>
<td>4/35</td>
<td>10/53</td>
<td>16.6</td>
<td>0.55 [0.16,1.93]</td>
</tr>
<tr>
<td>Subtotal (95% Cl)</td>
<td>41/336</td>
<td>45/354</td>
<td>100.0</td>
<td>0.84 [0.37,1.89]</td>
</tr>
<tr>
<td>03 Cervix LCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loeve 1992 [34]</td>
<td>0/34</td>
<td>14/34</td>
<td>4.9</td>
<td>0.02 [0.00,0.36]</td>
</tr>
<tr>
<td>Bass 1993 [19]</td>
<td>6/101</td>
<td>15/101</td>
<td>15.4</td>
<td>0.36 [0.13,0.98]</td>
</tr>
<tr>
<td>De Barbeyrac 1994 [30]</td>
<td>0/20</td>
<td>12/20</td>
<td>4.7</td>
<td>0.02 [0.00,0.31]</td>
</tr>
<tr>
<td>Skurnik 1994 [43]</td>
<td>0/1</td>
<td>0/1</td>
<td>2.3</td>
<td>1.00 [0.01,2.43]</td>
</tr>
<tr>
<td>Thejls 1994 [44]</td>
<td>0/18</td>
<td>6/10</td>
<td>4.4</td>
<td>0.02 [0.00,0.40]</td>
</tr>
<tr>
<td>Cahy 1995 [26]</td>
<td>0/2</td>
<td>0/2</td>
<td>2.5</td>
<td>1.00 [0.01,17.27]</td>
</tr>
<tr>
<td>Chout 1995 [26]</td>
<td>3/54</td>
<td>0/54</td>
<td>4.6</td>
<td>7.41 [0.37,146.96]</td>
</tr>
<tr>
<td>Pasternak 1996 [37]</td>
<td>7/39</td>
<td>6/39</td>
<td>13.6</td>
<td>1.20 [0.36,3.97]</td>
</tr>
<tr>
<td>Toye 1996 [45]</td>
<td>2/0</td>
<td>11/19</td>
<td>9.7</td>
<td>0.09 [0.02,0.48]</td>
</tr>
<tr>
<td>Gaydosh 1998 [32]</td>
<td>4/35</td>
<td>10/35</td>
<td>12.9</td>
<td>0.32 [0.06,0.70]</td>
</tr>
<tr>
<td>Cook 1999 [29]</td>
<td>4/26</td>
<td>12/26</td>
<td>12.6</td>
<td>0.21 [0.06,0.72]</td>
</tr>
<tr>
<td>Vinelette 1999 [46]</td>
<td>4/26</td>
<td>10/26</td>
<td>12.5</td>
<td>0.28 [0.08,1.10]</td>
</tr>
<tr>
<td>Subtotal (95% Cl)</td>
<td>30/375</td>
<td>96/367</td>
<td>100.0</td>
<td>0.26 [0.12,0.56]</td>
</tr>
<tr>
<td>04 Urine PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferrero 1998 [31]</td>
<td>7/39</td>
<td>13/39</td>
<td>100.0</td>
<td>0.44 [0.15,1.26]</td>
</tr>
<tr>
<td>Subtotal (95% Cl)</td>
<td>7/39</td>
<td>13/39</td>
<td>100.0</td>
<td>0.44 [0.15,1.26]</td>
</tr>
<tr>
<td>05 Cervix PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biro 1994 [22]</td>
<td>7/25</td>
<td>5/25</td>
<td>35.1</td>
<td>1.56 [0.42,5.78]</td>
</tr>
<tr>
<td>Pasternak 1996 [37]</td>
<td>8/39</td>
<td>6/39</td>
<td>38.5</td>
<td>1.42 [0.44,4.56]</td>
</tr>
<tr>
<td>Ferrero 1998 [31]</td>
<td>0/39</td>
<td>9/39</td>
<td>13.5</td>
<td>0.04 [0.00,0.73]</td>
</tr>
<tr>
<td>Schubiner 1990 [41]</td>
<td>5/25</td>
<td>0/25</td>
<td>13.0</td>
<td>13.68 [0.71,262.19]</td>
</tr>
<tr>
<td>Subtotal (95% Cl)</td>
<td>20/128</td>
<td>20/128</td>
<td>100.0</td>
<td>1.16 [0.25,5.47]</td>
</tr>
<tr>
<td>06 ESA urine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brave 1990 [24]</td>
<td>3/12</td>
<td>0/12</td>
<td>9.8</td>
<td>9.21 [0.24,200.61]</td>
</tr>
<tr>
<td>Shafe 1993 [42]</td>
<td>7/43</td>
<td>17/43</td>
<td>34.1</td>
<td>0.30 [0.11,0.82]</td>
</tr>
<tr>
<td>Chernsak 1994 [27]</td>
<td>12/27</td>
<td>12/27</td>
<td>32.6</td>
<td>0.21 [0.07,0.62]</td>
</tr>
<tr>
<td>Bassi 1995 [21]</td>
<td>13/16</td>
<td>7/16</td>
<td>23.6</td>
<td>5.57 [1.13,27.52]</td>
</tr>
<tr>
<td>Subtotal (95% Cl)</td>
<td>40/98</td>
<td>36/98</td>
<td>100.0</td>
<td>1.86 [0.39,8.75]</td>
</tr>
<tr>
<td>07 ESA cervix</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monsaca 1992 [49]</td>
<td>1/17</td>
<td>0/17</td>
<td>6.1</td>
<td>3.18 [0.12,83.77]</td>
</tr>
<tr>
<td>Loeve 1992 [34]</td>
<td>14/34</td>
<td>0/34</td>
<td>7.5</td>
<td>48.80 [2.76,866.27]</td>
</tr>
<tr>
<td>Chernsak 1994 [27]</td>
<td>7/27</td>
<td>12/27</td>
<td>20.4</td>
<td>1.36 [0.08,216.07]</td>
</tr>
<tr>
<td>Biro 1994 [22]</td>
<td>7/20</td>
<td>0/20</td>
<td>7.2</td>
<td>22.79 [1.20,432.61]</td>
</tr>
<tr>
<td>Thejls 1994 [44]</td>
<td>7/18</td>
<td>6/18</td>
<td>18.8</td>
<td>1.27 [0.33,4.97]</td>
</tr>
<tr>
<td>Schubiner 1994 [43]</td>
<td>7/18</td>
<td>5/18</td>
<td>18.6</td>
<td>1.65 [0.43,6.71]</td>
</tr>
<tr>
<td>Toye 1996 [45]</td>
<td>19/19</td>
<td>11/19</td>
<td>7.2</td>
<td>28.83 [1.52,547.37]</td>
</tr>
<tr>
<td>Pasku 1997 [38]</td>
<td>6/81</td>
<td>0/81</td>
<td>7.4</td>
<td>14.41 [0.79,261.62]</td>
</tr>
<tr>
<td>Schubiner 1990 [41]</td>
<td>3/25</td>
<td>0/25</td>
<td>6.9</td>
<td>7.93 [0.38,162.07]</td>
</tr>
<tr>
<td>Subtotal (95% Cl)</td>
<td>69/239</td>
<td>34/239</td>
<td>100.0</td>
<td>4.10 [1.15,14.89]</td>
</tr>
<tr>
<td>08 DFA cervix</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rahm 1992 [39]</td>
<td>2/22</td>
<td>0/22</td>
<td>29.0</td>
<td>5.49 [0.25,121.19]</td>
</tr>
<tr>
<td>Thejls 1994 [44]</td>
<td>3/18</td>
<td>6/18</td>
<td>71.0</td>
<td>0.40 [0.08,1.94]</td>
</tr>
<tr>
<td>Subtotal (95% Cl)</td>
<td>5/40</td>
<td>6/40</td>
<td>100.0</td>
<td>1.05 [0.08,12.93]</td>
</tr>
</tbody>
</table>

Fig. 3. Meta-analysis of pooled studies. For each study, the number of false-negative results (n) is expressed as a fraction of the total number of tests (N) for both the test under examination and the gold standard.
mastic low prevalence population that will be the focus of any targeted screening programmes. The tests under examination all had very high specificities and high sensitivities, and the high specificities of the tests made evaluation difficult. Diagnostic tests do not commonly perform as well as those studied in this meta-analysis. There are usually compromises to be made between sensitivity and specificity, but this does not appear to be the case for chlamydia tests. None of the studies had a clinical outcome recorded and this makes a thorough evaluation difficult. It was impossible to calculate positive or negative predictive values. Negative verification was in-built to all the study designs but there was no record that those women who were negative for the test and gold standard did not eventually experience adverse clinical outcomes.

Population-based testing for chlamydia is still uncommon even in research studies. Studies conducted amongst asymptomatic populations need to be large and there is often difficulty in recruitment. This factor may have led to publication bias in this review. A further limitation of this meta-analysis is the quality of the included studies. The ideal study of diagnostic test performance would involve comparison with a recognised gold standard. However, in the field of chlamydia research the gold standard, culture, does not perform as well as the tests against which it is compared. This is possibly the most important anomaly in this review.

The nucleic acid amplification tests were demonstrably better both for urine samples (LCR and PCR) and cervical swabs (PCR) than alternative methods of detecting chlamydia. The collection of cervical swabs is invasive, requiring speculum examination, and is uncomfortable and time-consuming. A further disadvantage of cervical swabs is that 30% of women have urethral infection only and a cervical swab would miss these cases [91]. Furthermore, skilled personnel are required to take swabs from the cervix. The equivalent test in men, a urethral swab, is not very acceptable in the screening context. An acceptability survey comparing urine sampling to urethral or cervical sampling would be needed to confirm this. Urine has an added advantage in women that first-void urine (FVU) may act as a genital lavage and samples will contain material from the urethra as well as the cervix. Testing urine offers an acceptable and efficient method of detecting chlamydia. The use of urine as a specimen would make any community-based screening programme easier to implement. There would be no need for expensive and time-consuming training of staff on how to take the samples. Urine samples are non-invasive, non-gender discriminatory, and quick to analyse. Further studies have also shown that, even when pooled, urine samples remain an accurate method of diagnosing chlamydia and this may make a screening programme more cost-effective, especially in low prevalence areas [92]. However, the presence of inhibitory substances in urine from women, both pregnant and non-pregnant, has been a cause for concern, with 3.9–11.0% of samples displaying inhibition depending on the DNA test used. Mahony et al. [93] demonstrated that cold storage, freezing or diluting samples overcame this inhibition in the majority of urine samples.

There is also evidence that self-collected urine or vulval swabs are highly sensitive. Unfortunately, these studies could not be included in this meta-analysis. None of these studies met the inclusion/exclusion criteria for this review as they focused on symptomatic women or high risk contacts of patients with known positive results [71, 83, 85, 94]. If tested, these options may increase the possibility of population-based screening for chlamydia.

When considering a screening programme it is important to ensure that, in the first instance, it is of benefit and, secondly, it does no harm. Chlamydia causes a sexually transmitted disease that can be treated easily and its sequelae are prevented by a short course of an inexpensive antibiotic. Nevertheless an incorrect diagnosis of a sexually transmitted disease can have many medical and personal implications. The cost of nucleic acid amplification tests has often been cited as a reason why they are not in common usage. However, the benefit of detecting 30% more cases of asymptomatic chlamydia, the cost saving of not having a doctor collecting specimens and the avoidance of a

<table>
<thead>
<tr>
<th>Test and sample</th>
<th>OR of a false-negative result (all studies)</th>
<th>OR of a false-negative result (excluding commercial studies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCR, urine</td>
<td>0.33 (0.13--0.8)</td>
<td>0.36 (0.12--1.06)</td>
</tr>
<tr>
<td>PCR, cervix</td>
<td>0.26 (0.12--0.54)</td>
<td>0.27 (0.07--1.09)</td>
</tr>
<tr>
<td>PCR, urine</td>
<td>0.84 (0.37--1.89)</td>
<td>0.84 (0.37--1.89)</td>
</tr>
<tr>
<td>Gene probe, cervix</td>
<td>0.84 (0.37--1.89)</td>
<td>0.84 (0.37--1.89)</td>
</tr>
<tr>
<td>Gene probe, urine (one study)</td>
<td>0.44 (0.15--1.26)</td>
<td>0.44 (0.15--1.26)</td>
</tr>
<tr>
<td>EIA, cervix</td>
<td>4.1 (1.15--14.59)</td>
<td>3.42 (0.64--18.26)</td>
</tr>
<tr>
<td>EIA, urine</td>
<td>1.36 (0.39--5.75)</td>
<td>1.41 (0.26--7.64)</td>
</tr>
<tr>
<td>DFA, cervix</td>
<td>1.05 (0.09--12.93)</td>
<td>1.05 (0.09--12.93)</td>
</tr>
<tr>
<td>LET, urine</td>
<td>47.02 (6.21--356.04)</td>
<td>47.02 (6.21--356.04)</td>
</tr>
</tbody>
</table>
pelvic examination may outweigh those extra costs. There has been debate as to whether pelvic examination of young women is beneficial in these programmes. However, the rate of detectable pathology in this group is so low that it may not be necessary [95].

Conclusions

A reduction in the incidence of pelvic inflammatory disease will have an important impact on health care generally, but targeted screening of asymptomatic individuals will be necessary to achieve this. Joyner et al. [96] have demonstrated persistence of genital chlamydia infections in 87% of untreated men and women. Ectopic pregnancy, tubal factor infertility and chronic pelvic pain were estimated to cost the health care in England at least £50 million in 1998 [97]. The incidence of pelvic inflammatory disease and its sequelae has fallen markedly in countries like Sweden and parts of the USA where there are screening programmes for chlamydia. More recently, Egger et al. [98] have shown that the rate of ectopic pregnancy reduced in the first year after the introduction of screening.

Screening for C. trachomatis in both sexes is important in many ways, not only for the clinical impact of the reduction of PID but, socially, to draw young men into the health-care system. The UK currently has >300 screening programmes in place [99], none of which is aimed at or specifically involves young men. With screening for chlamydia the opportunity arises to educate men and share the responsibility for their health care.

This review has demonstrated that nucleic acid amplification techniques, especially LCR for urine samples, are highly effective at detecting asymptomatic chlamydia infection. The review also demonstrated that EIA, which is used extensively to detect asymptomatic infection, will probably miss a large proportion of infections in a screening programme. If screening for chlamydia is to be efficient the programme must detect as many cases as possible and currently nucleic acid amplification methods with urine samples are the best option. New techniques are constantly being introduced and as peer reviewed papers are published it is important that further reviews are conducted.

This work was funded by an EU Biomed Grant BMH4-98-3059 based in the Department of Obstetrics and Gynaecology, Aberdeen University, Aberdeen, UK. The authors acknowledge the assistance of Jane Wilson (JW).

References


25. Brokenshure MK, Say PJ, van Vonn AI, Wong C. Evaluation of...


