The specificity of prostate-specific antigen (PSA) for early intervention in repeat biopsy is unsatisfactory. PCA3 may be more accurate in outcome prediction than other methods for the early detection of prostate cancer. However, the results were inconsistent in repeated biopsies. Therefore, we performed a systematic review and meta-analysis to evaluate the role of PCA3 in outcome prediction. A systematic bibliographic search was conducted for articles published before April 2013, using PubMed, Medline, Web of Science, Embase and other databases from health technology assessment agencies. The quality of the studies was assessed on the basis of QUADAS criteria. Eleven studies of diagnostic tests with moderate to high quality were selected. A meta-analysis was carried out to synthesise the results. The results of the
meta-analyses were heterogeneous among studies. We performed a subgroup analysis (with or without inclusion of high-grade prostatic intraepithelial neoplasia (HGPIN) and Atypical small acinar proliferation (ASAP)). Using a PCA3 cut-off of 20 or 35, in the two sub-groups, the global sensitivity values were 0.93 or 0.80 and 0.79 or 0.75, specificities were 0.65 or 0.44 and 0.78 or 0.70, positive likelihood ratios were 1.86 or 1.58 and 2.49 or 1.78, negative likelihood ratios were 0.81 or 0.43 and 0.91 or 0.82, and diagnostic ORs were 5.73 or 3.45 and 7.13 or 4.11. The AUCs of the SROC curve were 0.85 or 0.72 and 0.81 or 0.69. PCA3 can be used for repeat biopsy of the prostate to improve accuracy of PCA detection. Innecessary biopsies can be avoided by using a PCA cut-off score of 20.

**Keywords:** PCA3; prostate cancer; repeat biopsy; systematic review; Meta-analysis

Asian Journal of Andrology advance online publication: ***; doi: ***

Received: 22 June 2013; Revised: 17 July; Accepted: 26 September 2013

**Introduction**

Prostate cancer (PCA) is recognised as one of the most common cancers in men in the Western world\(^1\). Early detection of PCA relies primarily on an elevated prostate-specific antigen (PSA) level and an abnormal digital rectal examination (DRE), which signal the need for prostate biopsy.
However, 75% of men with PSA values between 2.5 and 10 ng ml\(^{-1}\) and/or a suspicious DRE have a negative first biopsy, even though 10%–35% of these men are diagnosed with PCA upon repeat biopsies\(^2,3\). The European Association of Urology guidelines recommend a repeat biopsy in men who have a negative first biopsy but a persistent suspicion of PCA\(^4\). However, the repeat biopsies are negative in 80% of examined men. Discomfort, anxiety, and severe complications can be associated with prostate biopsies. Repeated biopsies also result in a greater economic cost\(^2,3\). To avoid unnecessary biopsies and increase the probability of detecting PCA during a repeat biopsy, additional tests are needed. In this regard, the prostate cancer gene 3 (PCA3) assay, a new PCA gene-based marker, appears to be promising. PCA3 expression has been found to be 66-fold higher than that in benign and normal prostate tissue in > 95% of malignant prostate tissue tested\(^5-7\). Numerous studies have shown a high level of PCA3 during the first biopsy. The sensitivity and specificity have been reported to be as high as 82.3% and 89.0%, respectively, with small differences\(^8-11\). However, these results differed in repeated biopsies. To clarify the discrepancy, we performed a meta-analysis.

**Methods**

**Data collection**

A systematic bibliographic search was conducted for articles published
before April 2013, using PubMed, Medline, Web of Science, Embase and databases from health technology assessment agencies. Additionally, manual searches were performed in journals specialising in cancer and urology. The search strategy consisted of consecutively entering the following key words: “prostate”; “prostatic neoplasms”; “prostate” and “cancer”; “carcinoma” or “tumour”; “PCa”; “upm3”; “dd3”; “pca3”; “prostate cancer antigen3”; and “aptimapca3”. Abstracts or unpublished reports were not included. No language restrictions were applied. All non-English articles were translated if necessary.

The inclusion criteria included studies whose population consisted of adult men who had undergone a repeat biopsy for prostate cancer. The intervention must have consisted of a quantitative determination of PCA3 gene expression in urine samples by molecular biology methods. The prostate biopsy was the gold standard with which to assess the technique. The results had to include the specific values of the diagnostic tests, such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and ROC curves, which must have been calculated using true positives, false positives, false negatives and true negatives. We also collected the following characteristics: the name of the first author of the study, the year of publication, the population studied, the mean age of the subjects, the mean PSA level, and the cut-off point. The bibliographic references were selected individually by two
researchers. All references were full articles. Quality assessment was based on the QUADAS questionnaire.

Statistical analysis of the included studies

The data from each study were organised systematically and extracted to obtain the true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). Meta-DiSc software was used to calculate the indices of diagnostic validity, including the sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio negative, likelihood ratio positive, and diagnostic OR. This allowed us to assess the discriminative power of the PCA3 test. Each value was determined together with a 95% CI.

We conducted the meta-analysis in accordance with evidence-based data we extracted. We evaluated the quality of the articles according to the QUADAS questionnaire. Meta-DiSc software (version 1.4) was used to aggregate the results. First, we determined the possible existence of a threshold effect by calculating the Spearman’s correlation coefficient and by using a graphic representation of “sensitivity” or “1-specificity” on a ROC space. Second, the possible heterogeneity of the studies was assessed by a chi-square test for sensitivity, specificity, PPV and NPV. The Q value was used to determine the probability coefficients and the OR. The results were represented in a forest plot. If there was evidence of a threshold effect, the studies were combined to create a summarised
SROC, and the area under the curve (AUC) was calculated. The analysis was performed following the random effects model, as well as subgroup analysis if heterogeneity was found.

**Results**

*Descriptive analysis of the included studies*

The systematic search for original articles yielded 900 bibliographic references. After reading the full text of all articles, 11 studies on repeat biopsy were included (Figure 1)\(^{12-22}\). All studies had adequate sensitivity, specificity, and positive and negative predictive values. According to the QUADAS questionnaire (Table 1), the quality of the studies on diagnostic testing was moderate to high.

The studies retrieved data from a total of 3,373 patients with a mean age between 62.5 and 67.0 years and mean PSA levels ranging from 4.8 to 16.0 ng ml\(^{-1}\) (Table 2). All patients underwent a repeat biopsy for comparison with the antigen determination. The intervention consisted of a quantitative determination of the PCA3 gene in urine samples of the patients. The studies presented the results as sensitivity, specificity, PPV, NPV and ROC curves (Tables 3 and 4). Using a PCA3 cut-off of 20 or 35, the sensitivities were between 67.0\% and 92.0\% or 38.0\% and 78.6\%, respectively, whereas the specificities were between 16.7\% and 64.0\% or 23.6\% and 78.6\%, respectively; the PPV and NPV values ranged from 26.1\% to 52.0\% or 15.7\% to 52.0\% and from 77.8\% to 89.9\% or 66.0\%
to 90.5%, respectively; and the AUCs were between 0.577 and 0.730 or 0.605 and 0.715, respectively. All studies reported 95% CIs. Cutoff points were established based on PCA3 scores. We divided the 11 studies into two groups: group A included HGPIN and ASAP, whereas group B did not.\textsuperscript{13,14,18,20-22}

Meta-analysis

The analysis was conducted using the 11 articles above. With a PCA3 cut-off of 20 or 35, Spearman’s correlation coefficient was 0.841 ($P = 0.002$) and 0.726 ($P = 0.011$), respectively, and the ROC space showed a curvilinear trend. The results suggest the existence of a threshold effect (Figure 2 and 3). Group A, which contained subjects with HGPIN and ASAP, was not similar to Group B. We then performed a separate meta-analysis on each group. The pooled sensitivities of using a PCA3 cut-off of 20 or 35 in group A and group B were 72% or 49% and 90% or 75%, respectively (Figure 4–7), and the specificities were 53% or 35% and 74% or 57%, respectively (Figure 8–11). Using a PCA3 cut-off of 20, the positive LR, negative LR, and diagnostic OR, the AUCs were 1.37, 0.49, 3.18 and 0.8462, respectively (Figure 12–15).

Discussion

In this review, we analysed the available literature regarding the use of urine PCA3 as a guiding marker for repeat prostate biopsy for detecting PCA. Although the levels of PCA3 in the urine are lower than in the
prostate tissue, PCA3 is readily detectable in urine samples. Clearly, PCA3 in the first biopsy shows excellent value. Some studies showed that during the first biopsy, when a PCA3 cut-off score of 35 was used, the sensitivity and specificity were up to 82.3% and 89.0%, with little differences between these studies. The results were much better than those using PSA. The best PSA cut-off value showed only 57.4% and 53.8% sensitivity and specificity, respectively.\textsuperscript{11-13,23-26} In an American study, the diagnostic accuracy of the score was evaluated in men undergoing an initial biopsy (277) and a repeat biopsy (280).\textsuperscript{27} In a European study, the AUC of PCA3 was 0.761 in the initial biopsy and 0.658 in the repeat biopsy.\textsuperscript{22} This finding suggests that PCA3 is more accurate than PSA at guiding both repeat biopsy and initial biopsy. The diagnostic accuracy was not affected by prostate volume, age or tPSA ranges.\textsuperscript{22,27}

For repeat biopsy cases, there was some variability among the studies in terms of PCA3. PCA3 has great value as a diagnostic tool. However, the problem is the optimal cut-off value. Although the specificity of a score of 20 is lower than that of 35, the values of other parameters are superior at a score of 20 than at a score of 35. The sensitivity results indicate that 75% of patients can be diagnosed by assessing PCA3 and using a cut-off score of 20. Thus, the results suggest that 20 is an appropriate cut-off score. The negative LR results indicate that PCA3 detection will lead to a
significant reduction in unnecessary biopsies, by more than half. The positive LR results indicate that the probability of a patient with positive PCA3 is almost 1.5 times higher than that of a patient with negative PCA3 to have PCA. The AUC can be interpreted as the performance acceptability of the diagnostic test. The AUC of a score of 20 is higher than that of a score of 35, which indicates greater diagnostic value. According to the analysed data and the meta-analysis, a PCA3 score cut-off of 20 is better than a score cut-off of 35. Although there were differences in these studies, the PCA3 results indicate that the detection of this biomarker has acceptable diagnostic validity indices and adequate sensitivity and can be used for guiding repeat biopsies of the prostate for PCA testing.

Using a PCA3 score cut-off of 20, group A showed better results than group B. Although group A had a slightly lower sensitivity than that of group B (72% vs 90%), the specificity of group A was higher (53% vs 35%). The specificity of group B was too low for diagnosis. Group A had more balanced sensitivity and specificity values, possibly because group A subjects had a higher PCA3 score. Most patients were still diagnosed with HGPIN and ASAP on repeat biopsy. Some studies showed that subjects diagnosed with HGPIN and ASAP had higher scores than healthy controls. Further studies are needed to determine why HGPIN and ASAP higher than normal. On a repeat biopsy, a PCA3 cut-off score
of 20 with HGPIN and ASAP is a valuable diagnostic tool and can be clinically applied.

There are several limitations of our meta-analysis. Some studies were not performed blinded, whereas some lacked explanation of the loss of the patients. But most have given explanations. These do not affect the results. We have tried to avoid these biases by expanding our search to several databases and conducting a rigorous screening for articles. We evaluated the quality of the articles according to the QUADAS questionnaire evaluation. The quality of the studies on diagnostic testing was moderate to high. We eliminated poor quality papers and avoided language restrictions. However, there were potential publication biases, such as unpublished studies and reports from commercial enterprises, which were excluded. It should be noted that the PCA3 score is inconclusive. Some studies used a cut-off score of 25, but most of the studies that we searched used a cut-off score of 20. Moreover, several studies showed that cut-off scores of 20 and 25 yielded similar results\textsuperscript{25,27}. Whereas other genes and proteins such as AMACR, HPG-1, STAMP1, STAMP2, DPIP, Trp-p8, GSTM1, GSTT1, CYP1A1, CYP1A2, CYP2E1, MDM2 T309G and NPY\textsuperscript{28-37} have also been considered as prostate-specific markers and their expression is altered in pathologic conditions, PCA3 is the only gene with that can be used with high specificity as a diagnostic tool\textsuperscript{38}. Additionally, PCA3 detection is a
minimally invasive test. Furthermore, PCA3 detection has good diagnostic performance because the sample is collected by urinary sediment after prostate massage$^{39}$.

Taking the above findings together, early use of the non-invasive method of PCA3 detection may lead to a significant reduction in the number of repeat biopsies that is conducted. Several studies showed that the PCA3 score was closely linked to the Gleason score and clinical stage. However, some studies showed conflicting result and questioned the relationship between the PCA3 score and PCA aggressiveness.$^{27,40}$ The PCA3 score decrease in patients who had been diagnosed with PCA but was still higher than normal$^{13,18-20,22}$. This finding does not affect the value of PCA3 as a diagnostic tool. Whether PCA3 can be used for clinical staging is not conclusive, and the association between PCA3 score and Gleason score requires further evaluation in controlled studies. Based on the results, we conclude that a PCA3 cut-off score of 20 is better than a cut-off score of 35 and that PCA3 is a much better diagnostic marker than PSA. This finding will be clinically useful for improving diagnostic accuracy and avoiding unnecessary biopsies in patients. However, more studies are needed to determine the costs and efficacy of this approach.

**Author contributions**

YL, XG conceived and designed the experiments. YL, PH extract and
analyzed the data. YL, CM checked the data. YL, XG drafted the paper.

**Competing interests**

The authors declare no competing interests.

**References**

8. Crawford ED, Rove KO, Trabulsi EJ, Qian JQ, Drewnowska KP et


biopsy: does an elevated PCA3 score indeed predict the presence of prostate cancer? *BJU Int* 2010; 106: 1138-42.


### Table 1: The QUADAS questionnaire evaluation of the quality of the 11 articles

<table>
<thead>
<tr>
<th>Reference</th>
<th>Selection</th>
<th>Interval</th>
<th>GS</th>
<th>New test</th>
<th>Blinded</th>
<th>Same data</th>
<th>Interpretation</th>
<th>Losses explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepe, 2013(^{12})</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Goode, 2013(^{13})</td>
<td>yes</td>
<td>nd</td>
<td>yes</td>
<td>yes</td>
<td>nd</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Wu, 2012(^{14})</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Pepe, 2012(^{15})</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Bollito, 2012(^{16})</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Barbera, 2012(^{17})</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Auprich, 2012(^{18})</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>nd</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Pepe, 2011(^{19})</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Remzi, 2010(^{20})</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Aubin, 2010(^{21})</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Haese, 2008(^{22})</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

GS: Gold standard; nd: not described

### Table 2: Main characteristics of the 11 included studies

<table>
<thead>
<tr>
<th>Reference</th>
<th>Population</th>
<th>Mean age</th>
<th>Mean PSA</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepe, 2013(^{12})</td>
<td>100</td>
<td>66.0</td>
<td>7.9</td>
<td>urine</td>
</tr>
<tr>
<td>Goode, 2013(^{13})</td>
<td>167</td>
<td>66.0</td>
<td>4.8</td>
<td>urine</td>
</tr>
<tr>
<td>Wu, 2012(^{14})</td>
<td>103</td>
<td>63.5</td>
<td>11.0</td>
<td>urine</td>
</tr>
<tr>
<td>Pepe, 2012(^{16})</td>
<td>118</td>
<td>62.5</td>
<td>8.5</td>
<td>urine</td>
</tr>
<tr>
<td>Bollito, 2012(^{16})</td>
<td>509</td>
<td>67.0</td>
<td>6.7</td>
<td>urine</td>
</tr>
<tr>
<td>Barbera, 2012(^{17})</td>
<td>177</td>
<td>64.0</td>
<td>9.5</td>
<td>urine</td>
</tr>
<tr>
<td>Auprich, 2012(^{18})</td>
<td>127</td>
<td>63.0</td>
<td>16.0</td>
<td>urine</td>
</tr>
<tr>
<td>Pepe, 2011(^{19})</td>
<td>74</td>
<td>64.0</td>
<td>8.9</td>
<td>urine</td>
</tr>
<tr>
<td>Remzi, 2010(^{20})</td>
<td>463</td>
<td>64.9</td>
<td>7.2</td>
<td>urine</td>
</tr>
<tr>
<td>Aubin, 2010(^{21})</td>
<td>1072</td>
<td>nd</td>
<td>nd</td>
<td>urine</td>
</tr>
<tr>
<td>Haese, 2008(^{22})</td>
<td>463</td>
<td>65.0</td>
<td>7.0</td>
<td>urine</td>
</tr>
</tbody>
</table>
Table 3: Diagnostic results based on the data retrieved from the articles included (score 20)

<table>
<thead>
<tr>
<th>References</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>S</th>
<th>E</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepe, 2013</td>
<td>26</td>
<td>60</td>
<td>2</td>
<td>12</td>
<td>92.9</td>
<td>16.7</td>
<td>30.2</td>
<td>85.7</td>
</tr>
<tr>
<td>Wu, 2012</td>
<td>24</td>
<td>23</td>
<td>13</td>
<td>43</td>
<td>67.0</td>
<td>64.0</td>
<td>52.0</td>
<td>78.0</td>
</tr>
<tr>
<td>Pepe, 2012</td>
<td>29</td>
<td>62</td>
<td>3</td>
<td>24</td>
<td>90.6</td>
<td>27.9</td>
<td>31.9</td>
<td>88.9</td>
</tr>
<tr>
<td>Bollito, 2012</td>
<td>108</td>
<td>215</td>
<td>15</td>
<td>171</td>
<td>88.2</td>
<td>44.3</td>
<td>40.7</td>
<td>89.6</td>
</tr>
<tr>
<td>Barbera, 2012</td>
<td>44</td>
<td>96</td>
<td>4</td>
<td>33</td>
<td>91.7</td>
<td>25.6</td>
<td>31.5</td>
<td>89.5</td>
</tr>
<tr>
<td>Auprich, 2012</td>
<td>35</td>
<td>62</td>
<td>9</td>
<td>21</td>
<td>85.0</td>
<td>25.3</td>
<td>38.0</td>
<td>77.8</td>
</tr>
<tr>
<td>Pepe, 2011</td>
<td>25</td>
<td>33</td>
<td>2</td>
<td>14</td>
<td>92.6</td>
<td>21.6</td>
<td>43.1</td>
<td>88.9</td>
</tr>
<tr>
<td>Remzi, 2010</td>
<td>94</td>
<td>165</td>
<td>34</td>
<td>170</td>
<td>73.4</td>
<td>50.7</td>
<td>36.3</td>
<td>83.3</td>
</tr>
<tr>
<td>Aubin, 2010</td>
<td>134</td>
<td>380</td>
<td>56</td>
<td>502</td>
<td>70.5</td>
<td>56.9</td>
<td>26.1</td>
<td>89.9</td>
</tr>
<tr>
<td>Haese, 2008</td>
<td>94</td>
<td>165</td>
<td>34</td>
<td>170</td>
<td>73.4</td>
<td>50.7</td>
<td>36.3</td>
<td>83.3</td>
</tr>
</tbody>
</table>

E: specificity; FP: false positives; FN: false negatives; S: sensitivity; TN: true negatives; TP: True positives; NPV: negative predictive value; PPV: positive predictive value.

Table 4: Diagnostic results based on the data retrieved from the articles included (score 35)

<table>
<thead>
<tr>
<th>References</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>S</th>
<th>E</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepe, 2013</td>
<td>22</td>
<td>55</td>
<td>6</td>
<td>17</td>
<td>78.6</td>
<td>23.6</td>
<td>28.6</td>
<td>73.9</td>
</tr>
<tr>
<td>Goode, 2013</td>
<td>8</td>
<td>43</td>
<td>11</td>
<td>105</td>
<td>42.0</td>
<td>71.0</td>
<td>15.7</td>
<td>90.5</td>
</tr>
<tr>
<td>Wu, 2012</td>
<td>15</td>
<td>17</td>
<td>22</td>
<td>49</td>
<td>38.0</td>
<td>77.0</td>
<td>50.0</td>
<td>66.0</td>
</tr>
<tr>
<td>Pepe, 2012</td>
<td>23</td>
<td>50</td>
<td>9</td>
<td>36</td>
<td>71.9</td>
<td>41.8</td>
<td>31.5</td>
<td>80.0</td>
</tr>
<tr>
<td>Bollito, 2012</td>
<td>93</td>
<td>117</td>
<td>30</td>
<td>269</td>
<td>75.2</td>
<td>69.8</td>
<td>52.0</td>
<td>86.7</td>
</tr>
<tr>
<td>Barbera, 2012</td>
<td>35</td>
<td>65</td>
<td>13</td>
<td>64</td>
<td>73.0</td>
<td>41.8</td>
<td>35.0</td>
<td>80.6</td>
</tr>
<tr>
<td>Auprich, 2012</td>
<td>33</td>
<td>35</td>
<td>11</td>
<td>48</td>
<td>75.0</td>
<td>57.8</td>
<td>48.5</td>
<td>81.4</td>
</tr>
<tr>
<td>Pepe, 2011</td>
<td>21</td>
<td>25</td>
<td>6</td>
<td>28</td>
<td>70.4</td>
<td>43.5</td>
<td>42.2</td>
<td>71.5</td>
</tr>
<tr>
<td>Remzi, 2010</td>
<td>60</td>
<td>94</td>
<td>68</td>
<td>241</td>
<td>46.9</td>
<td>71.9</td>
<td>38.9</td>
<td>77.9</td>
</tr>
<tr>
<td>Aubin, 2010</td>
<td>92</td>
<td>190</td>
<td>98</td>
<td>692</td>
<td>48.4</td>
<td>78.6</td>
<td>32.6</td>
<td>87.6</td>
</tr>
<tr>
<td>Haese, 2008</td>
<td>60</td>
<td>94</td>
<td>68</td>
<td>241</td>
<td>47.0</td>
<td>72.0</td>
<td>39.0</td>
<td>78.0</td>
</tr>
</tbody>
</table>

E: specificity; FP: false positives; FN: false negatives; S: sensitivity; TN: true negatives; TP: True positives; NPV: negative predictive value; PPV: positive predictive value.

Figure 1. General outline for the selection of the studies included.
Figure 2. Analysis of the threshold effect: Spearman’s correlation coefficient.
Figure 3. Analysis of the threshold effect: ROC space.
Figure 4. Forest plots of the meta-analysis values for: sensitivity (score 35 group a).
Figure 5. Forest plots of the meta-analysis values for: sensitivity (score 35 group b).
Figure 6. Forest plots of the meta-analysis values for: specificity (score 20 group a).
Figure 7. Forest plots of the meta-analysis values for: sensitivity (score 20 group b).
Figure 8. Forest plots of the meta-analysis values for: sensitivity (score 20 group b).
Figure 9. Forest plots of the meta-analysis values for: specificity (score 20 group b).
Figure 10. Forest plots of the meta-analysis values for: specificity (score 35 group a).
Figure 11. Forest plots of the meta-analysis values for: specificity (score 35 group b).
Figure 12. Forest plots of the meta-analysis values for: positive likelihood ratio (score 20).
Figure 13. Forest plots of the meta-analysis values for: negative likelihood ratio.
Figure 14. Forest plot of the meta-analysis values for: diagnostic odds ratio.
Figure 15. Forest plot of the meta-analysis values for: SROC curve (score 20 group a).
Articles screened by title and abstract (900)

- 834 excluded
  - Duplication (495)
  - Not relevant (339)

Full-text articles reviewed (66)

- 55 excluded
  - Only first biopsy (31)
  - Only detects diagnosed (13)
  - Reviews or commentaries (8)
  - Mixes with other methods (1)
  - Incomplete data (2)

Included in analysis (11)
Spearman correlation coefficient: 0.726

Logit (TPR) vs Logit (FPR) \( P = 0.011 \)

<table>
<thead>
<tr>
<th>Var</th>
<th>Coeff</th>
<th>Std. Error</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1.023</td>
<td>0.163</td>
<td>6.290</td>
<td>0.0001</td>
</tr>
<tr>
<td>b (1)</td>
<td>-0.001</td>
<td>0.134</td>
<td>0.008</td>
<td>0.9940</td>
</tr>
</tbody>
</table>
Symmetric SROC
AUC = 0.6647
SE (AUC) = 0.0244
Q* = 0.6251
Sensitivity (95% CI)

- Wu. A. K: 0.65 (0.47–0.80)
- Auprich. M: 0.80 (0.65–0.90)
- Remzi. M: 0.73 (0.65–0.81)
- Aubin. S. M. J: 0.71 (0.63–0.77)
- Haese. A: 0.73 (0.65–0.81)
- Pooled: 0.72 (0.68–0.76)

Chi-square = 2.66, df = 4 (P = 0.6159)
Inconsistency (I-square) = 0.0%
Sensitivity (95% CI)

- Pepe. P: 0.93 (0.76–0.99)
- Bollito. E: 0.88 (0.81–0.93)
- Barbera. M: 0.92 (0.80–0.98)
- Pepe. P: 0.93 (0.76–0.99)
- Pooled: 0.90 (0.86–0.93)

chi-square = 1.29, df = 4 (P = 0.8635)
Inconsistency (I-square) = 0.0%
Sensitivity (95% CI)

Goode    0.42 (0.20–0.67)
Wu. A    0.41 (0.25–0.58)
Auprich. M 0.75 (0.60–0.87)
Remzi. M  0.47 (0.38–0.56)
Aubin. S  0.48 (0.41–0.56)
Haese. A  0.47 (0.38–0.56)

Pooled    0.49 (0.45–0.53)

chi-square = 14.33  df = 5  (P = 0.0137)
Inconsistency (I-square) = 65.1%
Sensitivity (95% CI)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepe. P</td>
<td>0.79 (0.59–0.92)</td>
</tr>
<tr>
<td>Pepe. P</td>
<td>0.72 (0.53–0.86)</td>
</tr>
<tr>
<td>Bollito. E</td>
<td>0.76 (0.67–0.83)</td>
</tr>
<tr>
<td>Barbera. M</td>
<td>0.73 (0.58–0.85)</td>
</tr>
<tr>
<td>Pepe. P</td>
<td>0.78 (0.58–0.91)</td>
</tr>
<tr>
<td>Pooled</td>
<td>0.75 (0.69–0.80)</td>
</tr>
</tbody>
</table>

chi-square = 0.60  df = 4 (P = 0.9628)

Inconsistency (I-square) = 0.0%
Specificity (95% CI)

Wu. A. K  0.65 (0.52–0.76)
Auprich. M  0.25 (0.16–0.36)
Remzi. M  0.51 (0.45–0.56)
Aubin. S. M. J  0.57 (0.54–0.60)
Haese. A  0.51 (0.45–0.56)

Pooled  0.53 (0.51–0.56)

chi-square = 37.16  df = 4  (P = 0.0000)
Inconsistency (I-square) = 89.2%
Specificity (95% CI)

Pepe. P 0.17 (0.09–0.27)
Pepe. P 0.28 (0.19–0.39)
Bollito. E 0.44 (0.39–0.49)
Barbera. M 0.26 (0.18–0.34)
Pepe. P 0.30 (0.17–0.45)

Pooled 0.35 (0.32–0.39)

chi-square = 34.02  df = 4  (P = 0.0000)
Inconsistency (I-square) = 88.20%
Goode 0.71 (0.63–0.78)
Wu. A 0.74 (0.62–0.84)
Auprich. M 0.58 (0.46–0.69)
Remzi. M 0.72 (0.67–0.77)
Aubin 0.78 (0.76–0.81)
Haese. A 0.72 (0.67–0.77)

Pooled 0.74 (0.72–0.76)

chi-square = 21.66  df = 5  (P = 0.0006)
Inconsistency (I-square) = 76.9%
Specificity (95% CI)

- Pepe. P: 0.24 (0.14–0.35)
- Pepe. P: 0.42 (0.31–0.53)
- Bollito. E: 0.70 (0.65–0.74)
- Barbera. M: 0.50 (0.41–0.59)
- Pepe. P: 0.53 (0.39–0.67)
- Pooled: 0.57 (0.53–0.61)

chi-square = 70.70 df = 4 (P = 0.0000)

Inconsistency (I-square) = 94.3%
Positive LR (95% CI)

<table>
<thead>
<tr>
<th>Author</th>
<th>Positive LR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepe. P</td>
<td>1.11</td>
<td>(0.96–1.29)</td>
</tr>
<tr>
<td>Wu. A</td>
<td>1.86</td>
<td>(1.24–2.79)</td>
</tr>
<tr>
<td>Pepe. P</td>
<td>1.26</td>
<td>(1.06–1.49)</td>
</tr>
<tr>
<td>Bollito. E</td>
<td>1.58</td>
<td>(1.41–1.76)</td>
</tr>
<tr>
<td>Barbera. M</td>
<td>1.23</td>
<td>(1.08–1.41)</td>
</tr>
<tr>
<td>Auprich. M</td>
<td>1.06</td>
<td>(0.88–1.29)</td>
</tr>
<tr>
<td>Pepe. P</td>
<td>1.32</td>
<td>(1.06–1.63)</td>
</tr>
<tr>
<td>Remzi. M</td>
<td>1.49</td>
<td>(1.28–1.73)</td>
</tr>
<tr>
<td>Aubin</td>
<td>1.64</td>
<td>(1.45–1.84)</td>
</tr>
<tr>
<td>Haese. A</td>
<td>1.49</td>
<td>(1.28–1.73)</td>
</tr>
</tbody>
</table>

Pooled Positive LR 1.37 (1.23–1.52)

Cochran-Q = 39.84  df = 9 (P = 0.0000)
Inconsistency (I-square) = 77.4%
Tau-squared=0.0212
Pepe. P 0.43 (0.10–1.79)
Wu. A 0.54 (0.34–0.86)
Pepe. P 0.34 (0.11–1.04)
Bollito. E 0.28 (0.17–0.45)
Barbera. M 0.33 (0.12–0.87)
Auprich. M 0.81 (0.41–1.61)
Pepe. P 0.25 (0.06–1.01)
Remzi. M 0.52 (0.39–0.71)
Aubin 0.52 (0.41–0.65)
Haese. A 0.52 (0.39–0.71)

Pooled Negative LR 0.49 (0.42–0.57)
Cochran-Q = 10.46  df = 9 (P = 0.0000)
Inconsistency (I-square) = 14.0%
Tau-squared = 0.0090
Pepe. P  2.60 (0.54–12.45)
Wu. A   3.45 (1.48–8.03)
Pepe. P  3.74 (1.04–13.44)
Bollito. E 5.73 (3.22–10.19)
Barbera. M 3.78 (1.26–11.33)
Auprich. M 1.32 (0.54–3.19)
Pepe. P  5.30 (1.10–25.49)
Remzi. M  2.85 (1.82–4.45)
Aubin  3.16 (2.25–4.44)
Haese. A  2.85 (1.82–4.45)

Pooled Diagnostic OR 3.18 (2.62–3.86)
Cochran-Q = 8.99  df = 9 (P = 0.4380)
Inconsistency (I-square) = 0.0%
Tau-squared = 0.0000