Photodynamic diagnosis of bladder cancer compared with white light cystoscopy: systematic review and meta-analysis

Short title: Photodynamic diagnosis of bladder cancer

Graham Mowatt,[1] James N'Dow,[2] Luke Vale,[1,3] Ghulam Nabi,[4] Charles Boachie,[1] Jonathan A Cook,[1] Cynthia Fraser,[1] TR Leyshon Griffiths,[5] for the Aberdeen Technology Assessment Review (TAR) Group

[1]Health Services Research Unit, Institute of Applied Health Sciences, University of Aberdeen, AB25 2ZD

[2]Academic Urology Unit, Institute of Applied Health Sciences, University of Aberdeen, AB25 2ZD

[3]Health Economics Research Unit, Institute of Applied Health Sciences, University of Aberdeen, AB25 2ZD

[4]Centre for Academic Clinical Practice and Department of Urology, University of Dundee, DD1 4HN

[5]Department of Cancer Studies and Molecular Medicine, University of Leicester, LE5 4PW

Correspondence to:

Mr Graham Mowatt, Health Services Research Unit

3rd Floor, University of Aberdeen

Health Sciences Building, Foresterhill

Aberdeen AB25 2ZD, UK

Tel: 44 1224 552494, Fax: 44 1224 554580, Email: g.mowatt@abdn.ac.uk

ABSTRACT

Objectives

The aim of this study was to assess the test performance and clinical effectiveness of photodynamic diagnosis (PDD) compared with white light cystoscopy (WLC) in people suspected of new or recurrent bladder cancer.

Methods

A systematic review of randomised controlled trials (RCTs), non-randomised comparative studies or diagnostic cross-sectional studies comparing PDD with WLC. Fifteen electronic databases and websites were searched (last searches April 2008). For clinical effectiveness only RCTs were considered.

Results

Twenty-seven studies (2949 participants) assessed test performance. PDD had higher sensitivity than WLC (92%, 95% CI 80 to 100% versus 71%, 95% CI 49 to 93%) but lower specificity (57%, 95% CI 36 to 79% versus 72%, 95% CI 47 to 96%). For detecting higher risk tumours, median sensitivity of PDD (89% (6 to 100%)) was higher than WLC (56% (0 to 100%)) whereas for lower risk tumours it was broadly similar (92% (20 to 95%) versus 95% (8 to 100%)). Four RCTs (709 participants) using 5-aminolaevulinic acid (5-ALA) as the photosensitising agent reported clinical effectiveness. Using PDD at transurethral resection of bladder tumour (TURBT) resulted in fewer residual tumours at check cystoscopy (relative risk (RR) 0.37, 95% CI 0.20 to 0.69) and longer recurrence-free survival (RR 1.37, 95% CI 1.18 to 1.59), compared with WLC.

Conclusions

PDD detects more bladder tumours than WLC, including more high risk tumours. Based on four RCTs reporting clinical effectiveness, 5-ALA mediated PDD at TURBT facilitates a more complete resection and prolongs recurrence-free survival.

Keywords: Systematic review, Meta-analysis, Diagnostic tests, Bladder cancer

Acknowledgements

This paper was developed from a Health Technology Assessment on the clinical and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer that was funded by the National Institute for Health Research Health Technology Assessment Programme (project number 07/02/01). See the HTA Programme website for further project information. JC holds a Medical Research Council UK Special Research Training Fellowship in Health Services and Health of the Public. The Health Services Research Unit and Health Economics Research Unit are core funded by the Chief Scientist Office of the Scottish Government Health Directorates. We thank Clare Robertson and Susan Wong for assistance with assessing full text studies for inclusion and Clare Robertson for assistance with data extraction and quality assessment. The views expressed in this report are those of the authors and do not necessarily reflect those of the Chief Scientist Office, Medical Research Council or the Department of Health.

INTRODUCTION

Bladder cancer is the seventh most common cancer in the UK, affecting more than 10,000 people each year (16). The majority of diagnosed patients (75 to 85%) present with non-muscle invasive disease, which is characterised by a probability of recurrence at five years of 31 to 78% (1). Flexible cystoscopy and voided urine cytology are currently the initial investigations of choice for patients with symptoms suggestive of bladder cancer. If flexible cystoscopy confirms a bladder tumour or urine cytology shows malignant cells in the absence of an upper urinary tract urothelial tumour, a rigid white light cystoscopy (WLC) under general or regional anaesthesia is performed with transurethral resection of bladder tumour (TURBT) where applicable.

The ultimate goal in the management of non-muscle invasive transitional cell carcinoma (TCC) of the bladder is the prevention of disease recurrence and progression. Early cancer detection is an essential prerequisite of successful therapy. Unfortunately, small papillary bladder tumours and flat urothelial tumours such as carcinoma *in situ* (CIS) can easily be overlooked during conventional WLC. Indeed, many of the recurrent tumours may be due to the persistence of residual tumour in the bladder after an incomplete TURBT. Moreover, progression to muscle invasive or metastatic TCC is more likely to occur in those with concomitant CIS (1). Non-muscle invasive TCC of the bladder is one of the most expensive cancers to manage on a per patient basis, because of its high prevalence, high recurrence rate and the need for long-term cystoscopic surveillance. The total cost of treatment and five year follow-up of patients with non-muscle invasive bladder cancer diagnosed during 2001-2002 in the UK was over £35 million (12).

Photodynamic diagnosis (PDD) is a technique that has been proposed to enhance tumour detection and resection. The principle of PDD is based on the interaction between a photosensitising agent with a high uptake by tumour cells and light with an appropriate wavelength, which is absorbed by the agent and re-emitted

with a different wavelength (18). We carried out a systematic review of the literature to assess the diagnostic performance of PDD compared with rigid WLC and its effects on patient outcomes.

METHODS

Search strategy

Highly sensitive electronic searches, using both controlled vocabulary and free text terms, were undertaken. The search strategies were originally developed for a systematic review (11) with a wider scope than this review and were designed to include retrieval of studies that assessed selected biomarker tests as well as PDD. We searched Medline (1966 - March Wk 3 2008), Medline In-Process (1st April 2008), Embase (1980 – Wk 13 2008), Biosis (1985- 27th March 2008), Science Citation Index (1970 – 1st April 2008), Health Management Information Consortium (March 2008) Cochrane Controlled Trials Register (The Cochrane Library, Issue 1 2008) as well as current research registers (National Research Register Archive (September 2007), Current Controlled Trials (March 2008), Clinical Trials (March 2008) and WHO International Clinical Trials Registry (March 2008)). Additional databases searched included the Cochrane Database of Systematic Reviews (The Cochrane Library, Issue 1, 2008), Database of Abstracts of Reviews of Effectiveness (March 2008), HTA Database (March 2008) and Medion (March 2008). Searches were restricted to English language publications. Details of the full strategies used for each database are available from the authors. Reference lists of all included studies were scanned to identify additional potentially relevant studies.

Study selection

We included studies that assessed the test performance or clinical effectiveness of PDD compared with WLC in people suspected of having bladder cancer or previously diagnosed with non-muscle invasive bladder cancer and on follow-up

cystoscopic examination. For test performance both randomised and observational (diagnostic cross-sectional or case-control) studies were included. However case-control studies in which the controls were healthy volunteers were excluded. The reference standard was histopathological examination of biopsied tissue and studies had to report or allow the calculation of true and false positives and negatives. For assessment of clinical effectiveness we included only RCTs and the outcomes considered were residual tumour at check cystoscopy, recurrence of bladder cancer over time following initial resection, and progression to muscle invasive disease.

Data abstraction and quality assessment

One reviewer screened the titles (and abstracts if available) of all reports identified by the search strategy. Full-text copies of all studies deemed to be potentially relevant were obtained and two reviewers independently assessed them for inclusion. One of three reviewers extracted details of study design, participants, index, comparator and reference standard tests and outcome data, and another checked the data extraction. Disagreements were resolved by consensus or arbitration by another reviewer.

Two reviewers independently assessed the quality of the included studies using a version of the quality assessment of diagnostic accuracy studies (QUADAS) tool adapted to make it more applicable for assessing reports of tests for bladder cancer. QUADAS is a quality assessment tool for use in systematic reviews of diagnostic studies (17) but it is designed to be adapted to make it applicable to a specific review topic. Disagreements were resolved by consensus or arbitration by a third reviewer.

Quantitative data synthesis

For studies of test performance, separate summary receiver operating characteristic (SROC) curves were derived for patient and biopsy level analysis. These meta-

analysis models were fitted using the hierarchical summary receiver operating characteristic (HSROC) model (9) in SAS 9.1. Summary sensitivity, specificity, positive and negative likelihood ratios and diagnostic odds ratios (DORs) for each model were reported as point estimate and 95% confidence interval (CI). Due to the clustering of biopsies within patients, the intervals from the biopsy level analyses were expected to be an underestimate of the true uncertainty.

For studies reporting clinical effectiveness, dichotomous outcome data were combined as relative risk (RR). In the absence of statistical heterogeneity, which was explored using chi-squared tests, I^2 statistics and visual inspection, a fixed effect model was used. Where there was evidence of heterogeneity, data were analysed using a random effects meta-analysis.

RESULTS

Trial flow

Figure 1 shows the flow of studies through the review. A list of the included diagnostic studies is shown in Supplementary Table 1 and a list of the included effectiveness studies is shown in Supplementary Table 2.

Study characteristics and methodological quality

The 27 diagnostic studies, published in 36 reports enrolled 2949 participants, with 2807 contributing to the analysis. Across 19 studies (2327 participants) reporting this information, 41% of the patients (n=946) were first time presenters with symptoms suspicious of bladder cancer while 59% (n=1381) had previously diagnosed bladder cancer. Further details of the diagnostic studies are shown in Supplementary Table 3.

In the four RCTs reporting effectiveness outcomes, published in eight reports, the groups were randomised to WLC or PDD, while in the other studies the groups were randomised to WLC or WLC and PDD. In Babjuk and colleagues (2), 33%

(20/60) of the PDD group and 45% (28/62) of the WLC group were newly presenting with symptoms suspicious of bladder cancer while 67% (40/60) of the PDD group and 55% (34/62) of the WLC group had previously diagnosed bladder cancer. The remaining studies did not report this information. All four studies used 5-aminolaevulinic acid (5-ALA) as the photosensitising agent. The follow-up periods for the studies were eight years, five years, two years and 10 to 14 days. Kriegmair and colleagues (7) only aimed to evaluate residual tumour following TURBT. Further details of the effectiveness studies are shown in Supplementary Table 4.

Figure 2 summarises the results of the quality assessment for the diagnostic studies. In all studies partial verification bias (all patients received a reference standard test) and test review bias (PDD and WLC were interpreted without knowledge of the results of the reference standard) were avoided. However all studies were judged to suffer from incorporation bias, in that PDD was considered not to be independent of the reference standard test, as biopsies used in the reference standard test were obtained via the PDD procedure. In all four studies reporting effectiveness outcomes it was unclear whether the sequence generation was really random or the treatment allocation was adequately concealed or whether outcomes assessors, care providers or patients were blinded.

Quantitative data synthesis

Diagnostic performance

In the pooled estimates for patient level analysis, based on evidence from five studies, PDD had higher sensitivity than WLC (92%, 95% CI 80 to 100% versus 71%, 95% CI 49 to 93%) but lower specificity (57%, 95% CI 36 to 79% versus 72%, 95% CI 47 to 96%). In the pooled estimates for biopsy level analysis, based on evidence from 14 studies, PDD also had higher sensitivity than WLC (93%, 95% CI 90 to 96% versus 65%, 95% CI 55 to 74%) but lower specificity (60%, 95% CI 49 to 71% versus 81%, 95%

CI 73 to 90%). Figure 3 shows the SROC plot for studies reporting biopsy-level analysis.

Across studies, the median sensitivity (range) of PDD compared with WLC for detecting lower risk, less aggressive tumours was broadly similar for patient level detection but higher for PDD for biopsy level detection (Table 1). However, for the detection of more aggressive, higher risk tumours the median sensitivity of PDD for both patient and biopsy level detection was higher than WLC. The higher sensitivity of PDD was also reflected in the detection of CIS alone, both for patient and biopsy level detection (Table 1).

Type of photosensitising agent

Most studies (n = 18) used 5-ALA as the photosensitising agent, while five used hexaminolaevulinate (HAL), two used hypericin and two used either 5-ALA or HAL. In patient based detection of bladder cancer, across four studies using 5-ALA and three using HAL, the median (range) sensitivity and specificity for 5-ALA was 96% (64 to 100%) and 52% (33 to 67%) respectively, compared with 90% (53 to 96%) sensitivity and 81% (43 to 100%) specificity for HAL. In biopsy based detection of bladder cancer, across 15 studies using 5-ALA, the median (range) sensitivity and specificity for 5-ALA was 95% (87 to 98%) and 57% (32 to 67%), compared with 85% (76 to 94%) and 80% (58 to 100%) for HAL.

Clinical effectiveness

All four studies, involving 544 patients, reported residual tumour rate (pTa and pT1). The timing of cystoscopy following TURBT ranged from 10 to 14 days, to 10 to 15 weeks after the initial resection. Compared with WLC, the use of PDD was associated with statistically significantly fewer residual pTa and pT1 tumours (RR 0.32, 95% CI 0.15 to 0.70 and RR 0.26, 95% CI 0.12 to 0.57 respectively), with an overall RR of 0.37 (95% CI 0.20 to 0.69). Two studies involving 313 patients reported

recurrence-free survival at 12 and 24 months. In the pooled estimates there was a statistically significant difference in favour of PDD at 24 months (RR 1.37, 95% CI 1.18 to 1.59) but not at 12 months (RR 1.40, 95% CI 0.96 to 2.03). The benefits of using PDD at TURBT in reducing tumour recurrence (pooled estimate RR 0.64, 95% CI 0.39 to 1.06) and progression (pooled estimate RR 0.57, 95% CI 0.22 to 1.46) in the longer term were less clear, with the effect estimates favouring PDD without reaching statistical significance.

DISCUSSION

Statement of principal findings

The pooled estimates for both patient and biopsy level analysis showed that PDD had higher sensitivity than WLC for detecting bladder cancer, but lower specificity. PDD also had a much higher sensitivity than WLC in the detection of more aggressive, higher risk tumours, including the detection of CIS alone. With regard to effectiveness outcomes, compared with WLC the use of PDD during TURBT resulted in a statistically and clinically significant reduction in residual pTa and pT1 tumours, longer recurrence-free survival of patients at two years following surgery and a longer interval between TURBT and tumour recurrence. There was no clear evidence of a difference between PDD and WLC for the outcomes of tumour recurrence and progression in the longer term. These results should be interpreted with caution as they are based on only a small number of studies.

Adjuvant single-dose chemotherapy administered within the first 24 hours and ideally within the first six hours following TURBT is standard practice in the UK and much of Europe and was shown in a meta-analysis to reduce the relative risk of recurrence by 39% with a median follow-up of 3.4 years (15). The administration of adjuvant intravesical therapy varied across the four RCTs and this made it more difficult to assess what the true added value of PDD might be in reducing bladder tumour recurrence rates in routine practice. Although single-dose intravesical

chemotherapy can chemoresect small residual papillary marker lesions (10) it is known to be insufficient treatment for patients with intermediate and high-risk tumours including concomitant CIS, the types more likely to be detected by PDD (14).

Strengths and limitations of the study

In terms of strengths, a recently recommended HSROC model was employed which takes account of the trade off between true and false positives and models between study heterogeneity (8). Pooled estimates of both patient and biopsy level detection were undertaken. However biopsy level estimates were likely to be an underestimate of the true uncertainty due to clustering of biopsies within patients. For reports of clinical effectiveness we focused on RCTs. In terms of limitations, non-English language studies were excluded. Based on screening English language titles or abstracts our searches identified 33 non-English language studies relating to PDD, some of which may have otherwise met the inclusion criteria.

Implications for practice and research

Our results suggest that the appropriate point in the clinical pathway for PDD to be used is in conjunction with rigid WLC during the initial TURBT, and possibly also in conjunction with rigid WLC during surveillance monitoring of high risk patients. The advantages of higher sensitivity (fewer false negative results, better detection of higher risk tumours) of PDD compared with WLC have to be weighed against the disadvantages of lower specificity (more false positive results, leading to additional unnecessary biopsies, potentially additional unnecessary investigations and the resulting anxiety caused to patients and their families). In terms of the photosensitising agents used, HAL would result in fewer false positives than 5-ALA (based on data for both patient and biopsy-level analyses), although it is possible that

other factors apart from the agent used may also have contributed to the specificity values reported.

The literature continues to develop with regard to PDD in conference abstracts. The study by Stenzl and colleagues (13) is noteworthy because it reports for the first time a HAL-based phase III multicentre RCT (PC B305) with clinical effectiveness outcomes. Of 766 patients randomised in 28 European and USA centres, the recurrence rate at nine months was 36% following HAL-based TURBT and 46% following WLC-assisted TURBT (p=0.029). Although full publication is awaited, the FDA in December 2009 approved HAL as an adjunct to WLC in the detection of non-muscle invasive bladder cancer.

We are aware of one other systematic review of PDD in non-muscle invasive bladder cancer, by Kausch and colleagues (5). Although Kausch and colleagues considered studies published in English, French or German, of 21 reports of 17 trials included, only two were non-English language (both German). Their review presented a patient-based meta-analysis of additional detection rate of PDD compared with WLC and considered effectiveness outcomes such as residual tumour and recurrence-free survival but did not report diagnostic accuracy measures such as sensitivity and specificity. However, similar to our review, Kausch and colleagues (5) concluded that PDD detects more patients with bladder tumours, especially more with CIS, than WLC, and that more patients have a complete resection and a longer recurrence-free survival when diagnosed with PDD.

Further research is needed in the form of RCTs comparing PDD alone, with PDD or rigid WLC plus single dose adjuvant chemotherapy at TURBT in patients presumed to have non-muscle invasive bladder cancer. Study design should take into account participant risk factors, for example smoking and age and allow outcomes to be reported based on risk categories at randomisation. Clinical effectiveness outcomes should include residual tumour rates at first check cystoscopy, recurrence-free survival, tumour recurrence rates, time to first

recurrence, and progression. Provision should be made for longer term (up to 10 years) follow-up.

1. Babjuk M, Oosterlinck W, Sylvester R, et al. *Guidelines on TaT1 (non-muscle-invasive) bladder cancer*. [document on the Internet]. Arnhem: European Association of Urology; 2009. [accesssed January 2010] URL: http://www.uroweb.org/fileadmin/tx_eauguidelines/2009/Full/TaT1_BC.pdf

2. Babjuk M, Soukup V, Petrik R, Jirsa M, Dvoracek J. 5-aminolaevulinic acid-induced fluorescence cystoscopy during transurethral resection reduces the risk of recurrence in stage Ta/T1 bladder cancer. BJU Int 2005; 96:798-802.

3. Hendricksen K, Moonen PM, der Heijden AG, Witjes JA. False-positive lesions detected by fluorescence cystoscopy: any association with p53 and p16 expression? World J Urol 2006; 24:597-601.

4. Jichlinski P, Wagnieres G, Forrer M, et al. Clinical assessment of fluorescence cytoscopy during transurethral bladder resection in superficial bladder cancer. Urol Res 1997; 25:S3-S6.

5. Kausch I, Sommerauer M, Montorsi F, et al. Photodynamic Diagnosis in Non-Muscle-Invasive Bladder Cancer: A Systematic Review and Cumulative Analysis of Prospective Studies. Eur Urol 2010; 57:595-606.

6. Koenig F, McGovern FJ, Larne R, et al. Diagnosis of bladder carcinoma using protoporphyrin IX fluorescence induced by 5-aminolaevulinic acid. BJU Int 1999; 83:129-135.

7. Kriegmair M, Zaak D, Rothenberger KH, et al. Transurethral resection for bladder cancer using 5-aminolevulinic acid induced fluorescence endoscopy versus white light endoscopy. J Urol 2002; 168:475-478.

8. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008; 149:889-897.

9. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. J Clin Epidemiol 2004; 57:925-932.

10. Masters JRW, Popert RJM, Thompson PM, et al. Intravesical chemotherapy with epirubicin: A dose response study. J Urol 1999; 161:1490-1493.

11. Mowatt G, Zhu S, Kilonzo M, et al. Systematic review of the clinical effectiveness and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer. Health Technol Assess 2010;14(4).

12. Sangar VK, Ragavan N, Matanhelia SS, Watson MW, Blades RA. The economic consequences of prostate and bladder cancer in the UK. BJU Int 2005; 95:59-63.

13. Stenzl AS, Roessler WR, Fradet YF, et al. HEXVIX fluorescence cystoscopy improves detection and resection of papillary bladder cancer and reduces early recurrence: a multicentre prospective randomized trial. Abstract 1010. Eur Urol Suppl 2009; 8:373.

14. Sylvester RJ, Oosterlinck W. An Immediate Instillation after Transurethral Resection of Bladder Tumor in Non-Muscle-Invasive Bladder Cancer: Has the Evidence Changed? Eur Urol 2009; 56:43-45.

15. Sylvester RJ, Oosterlinck W, Van Der Meijden APM. A single immediate postoperative instillation of chemotherapy decreases the risk of recurrence in

patients with stage Ta T1 bladder cancer: A meta-analysis of published results of randomized clinical trials. J Urol 2004; 171:2186-2190.

16. *UK bladder cancer statistics* [webpages on the Internet]. Cancer Research UK; 2010. [accessed January 2010] URL: http://info.cancerresearchuk.org/cancerstats/types/bladder/?a=5441

17. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3(25):Nov 10.

18. Zaak D, Karl A, Knuchel R, et al. Diagnosis of urothelial carcinoma of the bladder using fluorescence endoscopy. BJU Int 2005; 96:217-222.

Table 1 Sensitivity of PDD and WLC in detecting stage/grade of tumour

	PDD sensitivity % Median (range)	WLC sensitivity % Median (range)	Number of patients (biopsies)	Number of studies
Less aggressive/lower risk				
Patient based detection	92 (20 to 95)	95 (8 to 100)	266	3
Biopsy based detection	96 (88 to 100)	88 (74 to 100)	1206 (5777)	7
More aggressive/higher risk including CIS				
Patient based detection	89 (6 to 100)	56 (0 to 100)	563	6
Biopsy based detection	99 (54 to 100)	67 (0 to 100)	1756 (7506)	13
CIS				
Patient based detection	83 (41 to 100)	32 (0 to 83)	563	6
Biopsy based detection	86 (54 to 100)	50 (0 to 68)	1756 (7506)	13

Notes:

1. The number of biopsies is the overall total reported by the studies.
2. Number of biopsies. In some studies more biopsies were taken for PDD than WLC and in these cases the higher number used for PDD has been used in the table. In the less aggressive/lower risk category, Hendricksen and colleagues (3) reported 217 biopsies for PDD and 123 for WLC while Koenig and colleagues (6) reported 130 biopsies for PDD and 67 for WLC. Hendricksen and colleagues and Koenig and colleagues were also included in the more aggressive/higher risk category, as was Jichlinski and colleagues (4), who reported 421 biopsies for PDD and 414 for WLC. The studies by Hendricksen and colleagues, Jichlinski and colleagues and Koenig and colleagues were also amongst those reporting detection of CIS.

Figure 1 Flow of studies through review process

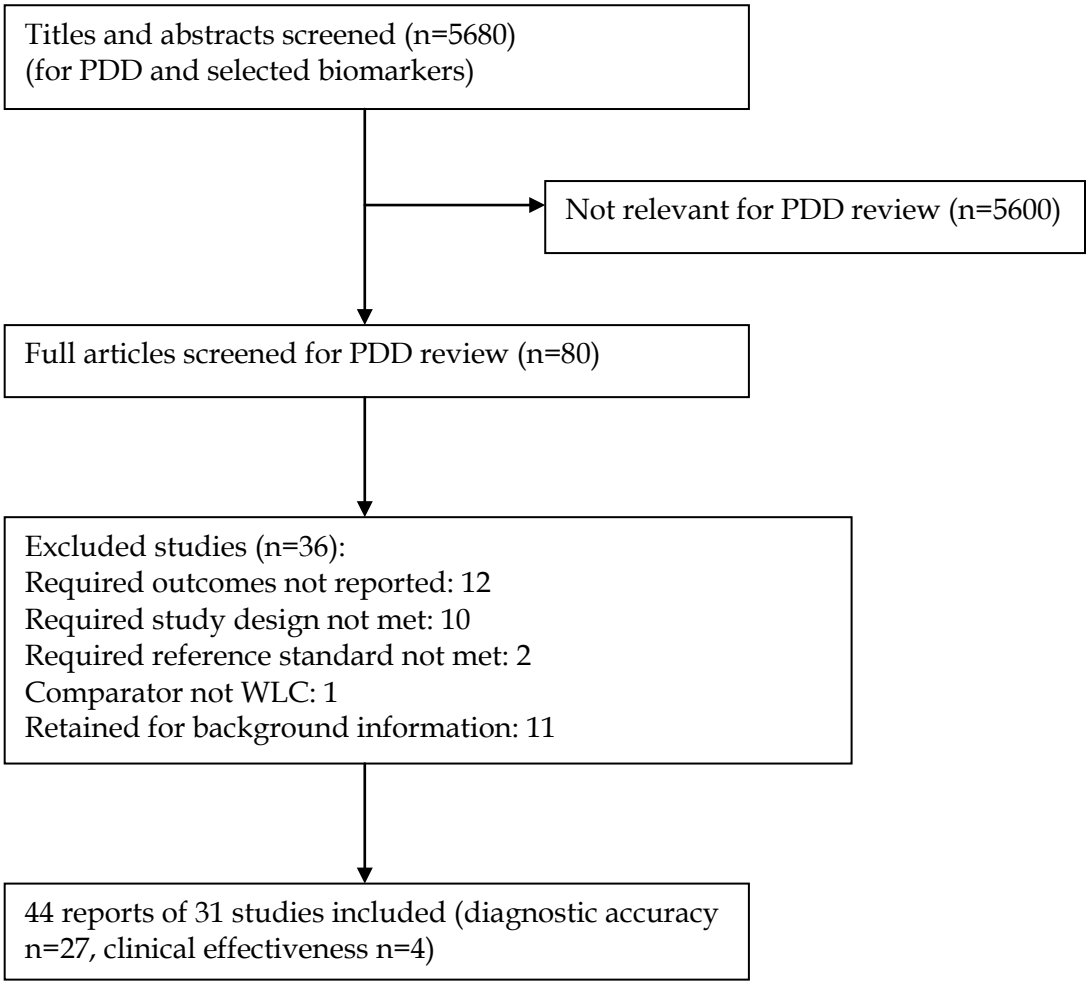

Figure 2 Summary of quality assessment of the diagnostic studies (n=27)

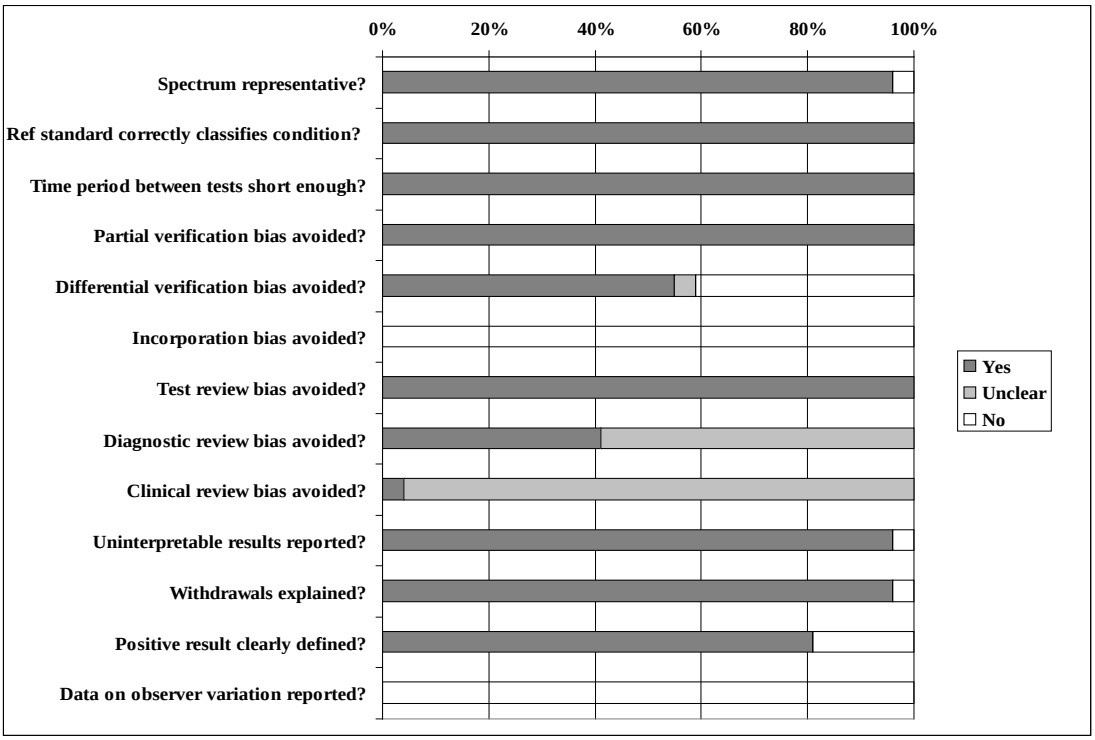

Figure 3 SROC plot for biopsy level analysis (n=14 studies)

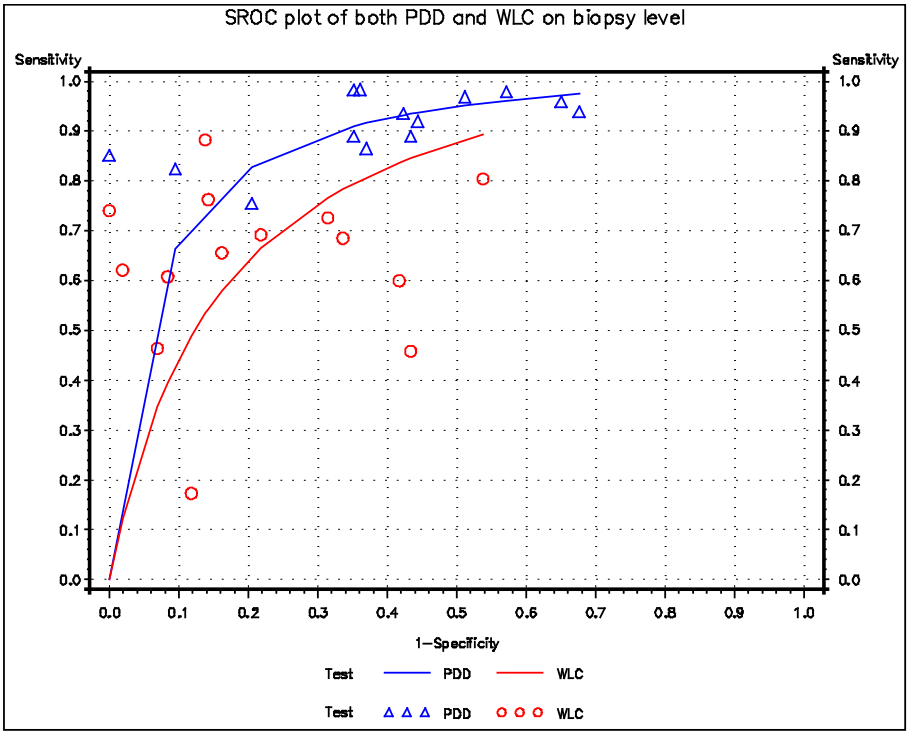

Figure captions

Figure 1 Flow of studies through review process

Figure 2 Summary of quality assessment of the diagnostic studies (n=27)

Figure 3 SROC plot for biopsy level analysis (n=14 studies)