Topic Exploration Report

Topic explorations are designed to provide a high-level briefing on new topics submitted for consideration by Health Technology Wales. The main objectives of this report are to:

- Determine the quantity of evidence available for a technology of interest.
- Identify any gaps in the evidence.
- Inform decisions on topics that warrant fuller assessment by Health Technology Wales (HTW).

Topic exploration report number	TER480
Topic	Artificial intelligence assisted tools for diagnosis of prostate cancer from whole slide digital biopsy images.
Summary of findings	Approximately 90,000 men undergo prostate biopsy to diagnose prostate cancer each year in the UK. Artificial intelligence (AI) assisted tools have been developed for use by pathologists as an aid to diagnosis, potentially improving accuracy and speed. The tools use AI algorithms which have been trained on large pre-existing image datasets to distinguish between cancerous and benign tissue on the slides and to grade cancerous tissue. We identified one health technology assessment (HTA), two systematic reviews, seven individual studies, and one ongoing study. Most studies were retrospective validation studies determining sensitivity and specificity of AI-assisted tools in diagnosing prostate cancer from whole slide images. A meta-analysis found the tools showed high diagnostic accuracy, reported as being comparable to diagnosis by pathologists alone. Four retrospective studies described user-testing where pathologists examined archived slides with and without aid from an AI-assisted tool. When an AI-assisted tool was used, resource use fell and time to diagnosis was shorter. One of these studies also evaluated real-world use of an AI-assisted tool as an aid to diagnosis in a hospital setting in Wales. A reduction in resource use and increased confidence in diagnosis was reported. We did not identify any studies that evaluated how the use of such tools affect clinical outcomes or any full economic analyses. It is unclear how use of AI-assisted tools for diagnosis of prostate cancer from whole slide biopsy images impacts on prostate cancer outcomes. There is a need for more prospective studies conducted in clinical settings and for cost effectiveness to be evaluated.

Introduction and aims

Prostate cancer is the most common cancer in men in the UK. In Wales, it is estimated that there are approximately 2,500 new cases diagnosed each year and incidence is rising. Suspected prostate cancer is investigated initially within primary care through prostate specific antigen tests and digital rectal examination. If indicated, people may be referred for a multiparametric magnetic resonance imaging (MRI) scan followed by a prostate biopsy. A pathologist examines the biopsy samples, using a conventional microscope or by examining a digital image of the whole slide on a computer or using the methods in parallel, to diagnose tissue as cancerous or benign. Cancerous tissue is graded using the Gleason score. Where diagnosis is uncertain, a second opinion is sought. Approximately 90,000 men undergo prostate biopsy for prostate cancer each year in the UK.

Artificial intelligence (AI) tools have been proposed that aid the diagnosis of prostate cancer from a biopsy. Whole slide images are analysed by an AI algorithm that can differentiate between slides containing potentially cancerous tissue and those that are likely benign. The AI tools can flag slides for review and assist with grading and, potentially, in the pre-ordering of additional diagnostic tests. This allows a pathologist to prioritise which slides to review first, order any additional tests that may be necessary more rapidly, and improve the speed and efficiency of diagnosis. Galen Prostate Solution (Ibex Medical Analytics) was identified as a specific example of this technology by the topic proposer. HTW researchers also identified Paige Prostate (Paige AI, Inc) and DeepDx (Deep Bio) as other specific examples.

Health Technology Wales researchers searched for evidence on the clinical and cost effectiveness of AI-assisted tools in the diagnosis of prostate cancer from whole slide digital biopsy images.

Evidence overview

We identified one health technology assessment (HTA), two systematic reviews and seven primary studies not included in the systematic reviews.

Health Technology Assessment

In November 2021, the National Institute of Health and Care Excellence (NICE) published a relevant Medtech innovation briefing (MIB280: Paige Prostate for prostate cancer). The authors presented evidence from five observational studies that included 3,444 biopsy slides reviewed in a pathology lab, although it appears only 2,844 were used in analysis. They found that the technology may increase sensitivity in detecting prostate cancer and had the potential to increase service efficiency when used as an adjunct to standard care. The key uncertainties were that the studies were mainly retrospective, only two compared the technology with standard care alone and none of the studies were conducted in the UK.

Systematic reviews

Morozov (2023) conducted a systematic review and meta-analysis of diagnostic accuracy. The primary outcome was accuracy in classifying tissue as benign or cancerous. Secondary outcomes were accuracy in grading, using the Gleason score, and agreement between AI and pathologists. The search period was from 2007 to 2021. Twenty-four studies examining around 8,000 prostate biopsies and 458 radical prostatectomies were included. The studies were retrospective or prospective, and none evaluated use of the tools in routine clinical practice. Meta-analysis found diagnostic accuracy to be high. Pooled sensitivity was 96% (95%CI 95, 97) and pooled specificity was 95% (95%CI 95, 96). Meta-analysis was not possible for Gleason grading, but sensitivity ranged from 77% to 87%, and specificity from 82% to 90%. The authors concluded that accuracy of AI-assisted tools was comparable to that of pathologists and recommended that the tools could be used to aid pathologists in diagnosis of prostate cancer

Evidence overview
to increase speed and quality. The authors do note that in real-world clinical practice AI-assisted tools would probably perform less well due to lower-quality histology slides and differences in preparation. In addition, they comment that they did not identify any articles that reported on cost-efficacy. Authors declared no conflict of interest. Van Booven (2021) conducted a broad systematic review of AI-assisted tools that used artificial neural networks in prostate cancer diagnosis. The dates of the search period were not included but the review identified 19 studies, of which four examined sensitivity and specificity of artificial neural networks in diagnosing prostate biopsies. None of the studies evaluated use in routine clinical practice. The authors reported that AI-assisted tools were comparable to pathologists in terms of accuracy of diagnosis for three studies, but one study found that the artificial neural network was unable to predict biopsy results. This latter study concluded that proper training of artificial neural networks was vital. The authors declared no conflict of interest. **Primary evidence** Aslem (2023) reported on a prospective cohort study conducted in Betsi Cadwaladr University Health Board. The AI-assisted tool was used as an aid to diagnosis for 3,975 slides from 860 patients. Slides were classified by the tool as likely-benign, uncertain, or likely-malignant, with areas of interest highlighted for review and then examined by a pathologist alongside the AI-assisted diagnosis. In comparison with diagnosis without the aid of an AI-assisted tool, request rates for additional tests fell from 8.7% to 4.5%. In addition, the study reported negative predictive values of 99.5% and positive predictive values of 99.4% for diagnosis by an AI-assisted tool alone, The study was not published in a peer reviewed journal and the authors declared no conflict of interest. This report was highlighted by the Topic Proposer who indicated that data collection is ongoing and that data from other studies have been presented at conferences and will be written up for full publication. Eloy (2023) conducted user-testing in a retrospective study using 105 archived biopsy whole slide images, from Portugal, of which 63% were known to be benign and 37% known to show prostate cancer. Diagnosis of prostate cancer by four pathologists using conventional methods was compared with diagnosis when aided by an AI-assisted tool. Diagnostic accuracy without aid was 95.0% vs 93.8% with aid. The pathologists requested approximately 20% fewer additional tests and 40% fewer second opinions when aided by the AI-assisted tool and the median time taken for diagnosis was about 20% lower. The authors concluded that use of the AI-assisted tool reduced resource use whilst maintaining diagnostic accuracy. One of the authors declared a conflict of interest, having received speaking fees and consultancy fees from the manufacturer of the device and a partner organisation. Raciti (2022) conducted a similar retrospective study using 610 whole slide images, taken from 218 different hospitals worldwide. Eighteen pathologists were recruited, and sensitivity increased from 88.7% without aid to 96.6% with the aid (95% MRMC CI, 4.5%, 11.5%) and specificity from 97.3% to 98.0% (95% MRMC CI, 0.1%, 1.2%). The authors reported that the study demonstrated the efficacy, safety and generalisability of the AI-assisted tool and stated that the study was the basis for FDA authorisation to allow the tool to be used in routine clinical practice. The authors declared a conflict of interest being employees of the manufacturer. Jung (2022) validated and conducted user-testing of an AI-assisted tool for diagnosis and grading of prostate cancer. 593 whole slide images from a centre in South Korea were used. Ground truth was determined by regrading of the slides by three expert pathologists. The AI-assisted tool was validated against the ground truth diagnoses, with the original pathology report used as a comparator. An additional pathologist conducted user testing, examining the

Evidence overview
slides with and without the aid of AI-assistance. Prostate cancer detection for the AI-assisted tool alone was comparable with the original pathology report and superior for Gleason grading. When the pathologist was aided by AI, accuracy improved and time for each slide decreased from 55.7 to 36.8 seconds/slide. The authors concluded that the study supported the use of the AI-assisted tool. The authors declared no conflict of interest. Three other studies reported on the development and validation of AI-assisted tools. Duenweg (2023) conducted a prospective study to compare two machine learning methods to grade tumours on slides. Data came from 47 patients with confirmed prostate cancer from the USA. Accuracy at determining cancerous from benign tissue was 89% and 88%, but only one machine learning method was able to grade cancers. The authors declared no conflict of interest. In a retrospective study, Tsuneki (2022) developed and validated a deep learning model to classify prostate cancer into indolent or aggressive, using 2285 slides from Japan. Higher accuracy was shown for diagnosing aggressive cancers than indolent, but the authors note that the model was unreliable for borderline cases. Two authors declared a conflict of interest being employees of the manufacturer. Oner (2022) developed a deep learning tool to aid with diagnosis of low-grade cases of prostate cancer using training and testing data from 99 patients from Singapore. An external validation study conducted using 10,512 European slides showed a high level of accuracy, which the authors report as showing that the tool was generalisable. However, before validation, the tool was trained on a subset of the European slides, which meant validation was not conducted in data from an unseen dataset. The authors declared no conflict of interest. **Economic evidence** We identified no full economic analyses for AI-assisted diagnosis of prostate cancer from whole slide digital biopsy images. The NICE Medtech innovation briefing (MIB280, 2021) included a brief description of the costs associated with the Paige Prostate system. The final pricing model was still being developed at the time of the NICE MIB but was stated to be available on a subscription basis with prices starting at £1 per slide. There is also a one-off fee associated with integrating Paige Prostate into the wider information management system. This fee varies but typically starts at £15,000. The initial cost of the system may be offset, at least partially, by savings through improved productivity or eliminating the costs of unnecessary treatment or progression of a disease and its related costs. The NICE MIB concluded that further evidence was needed to assess real-world cost effectiveness and impact on systems in the UK. **Ongoing studies** One ongoing study was identified (NCT05228197). The study is a cross-sectional cohort study to assess the diagnostic accuracy and cost-effectiveness of the Galen Prostate solution in diagnosing prostate cancer on prostate biopsy from 780 patients. The study is being conducted in the NHS with an estimated study completion date reported as the end of March 2023. **Technology classification** AI-assisted diagnosis of prostate cancer from whole slide digital biopsy images is a digital health technology and was determined to be a Tier C technology according to the Evidence Standards Framework for Digital Health Technologies. Technologies within this classification provide information that will be used to aid treatment or diagnosis, to triage or identify early signs of a disease or condition or will be used to guide next diagnostics or next treatment interventions. For technologies of this classification, it is recommended that satisfactory evidence for effectiveness is produced to demonstrate effectiveness of the technology. This includes studies conducted in a setting like the UK health and care system, peer-reviewed studies, and prospective studies. Evidence to support the claimed benefits of AI-assisted diagnosis of whole slide digital biopsy images should include real-world evaluations of its

Evidence overview
clinical utility and include one or more high-quality studies that support the claimed benefits in a relevant setting, showing improvements in relevant outcomes. Similarly, appropriate assessment of the economics of AI-assisted diagnosis of prostate biopsy should be undertaken.

Areas of uncertainty
• Direct evidence on the impact of AI-assisted diagnosis on patient outcomes when compared with standard care. • The impact of AI-assisted prostate cancer diagnosis on pathologists' decisions. • Evidence for validation of AI algorithms conducted by independent researchers. • Evidence for cost effectiveness.

Literature search results

Health technology assessments and guidance

NICE. (2021). Paige Prostate for prostate cancer. Medtech innovation briefing (MIB280). Medtech innovation briefing. Available at: https://www.nice.org.uk/advice/mib280/chapter/summary

Evidence reviews and economic evaluations

Morozov A, Taratkin M, Bazarkin A, et al. (2023). A systematic review and meta-analysis of artificial intelligence diagnostic accuracy in prostate cancer histology identification and grading. Prostate Cancer Prostatic Dis. Epub ahead of print. doi: https://dx.doi.org/10.1038/s41391-023-00673-3

Van Booven DJ, Kuchakulla M, Pai R, et al. (2021). A systematic review of artificial intelligence in prostate cancer. Res Rep Urol. 13: 31-9. doi: http://dx.doi.org/10.2147/RRU.S268596

Individual studies

Duenweg SR, Brehler M, Bobholz SA, et al. (2023). Comparison of a machine and deep learning model for automated tumor annotation on digitized whole slide prostate cancer histology. PLoS One. 18(3): e0278084. doi: https://doi.org/10.1371/journal.pone.0278084

Eloy C, Marques A, Pinto J, et al. (2023). Artificial intelligence-assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. Virchows Archiv. 482(3): 595-604. doi: https://dx.doi.org/10.1007/s00428-023-03518-5

Jung M, Jin MS, Kim C, et al. (2022). Artificial intelligence system shows performance at the level of uropathologists for the detection and grading of prostate cancer in core needle biopsy: an independent external validation study. Modern Pathology. 35(10): 1449-57. doi: https://doi.org/10.1038/s41379-022-01077-9

Oner MU, Ng MY, Giron DM, et al. (2022). An AI-assisted tool for efficient prostate cancer diagnosis in low-grade and low-volume cases. Patterns. 3(12): 100642. doi: https://doi.org/10.1016/j.patter.2022.100642

Raciti P, Sue J, Retamero JA, et al. (2022). Clinical Validation of Artificial Intelligence-Augmented Pathology Diagnosis Demonstrates Significant Gains in Diagnostic Accuracy in Prostate Cancer Detection. Archives of Pathology & Laboratory Medicine. doi: https://doi.org/10.5858/arpa.2022-0066-OA

Tsuneki M, Abe M, Ichihara S, et al. (2023). Inference of core needle biopsy whole slide images requiring definitive therapy for prostate cancer. BMC Cancer. 23(1): 11. doi: https://doi.org/10.1186/s12885-022-10488-5

Ongoing research

Ahmed HU. (2022). A Study to Assess the Clinical and Cost-effectiveness of the Galen Prostate Artificial Intelligence Histology System in Diagnosing Clinically Important Prostate Cancer on Prostate Biopsy Tissue [NCT05228197]. Clinical Trials Registration. Available at: https://clinicaltrials.gov/study/NCT05228197 [Accessed 13th July 2023].

Evidence supplied by the Topic Proposer
Aslam M, Health A. (2023). Successful deployment of an Artificial Intelligence solution for primary diagnosis of prostate biopsies in clinical practice. Trillium Pathology. 1: on-line. doi: https://doi.org/10.47184/tp.2023.01.03 Two additional conference abstracts: Comperat E, Rioux-Leclercq N, Levrel O, et al. (2021). CP-03-001: Clinical level AI-based solution for primary diagnosis and reporting of prostate biopsies in routine use: a prospective reader study 33rd European Congress of Pathology - Abstracts. Virchows Archiv. 479(Suppl 1): 1-320. doi: https://doi.org/10.1007/s00428-021-03157-8 Raoux D, Yazbin I, Arbov S, et al. (2021). 497: Novel AI-Based Solution for Supporting Prostate Cancer Diagnosis Increases the Efficiency and Accuracy of Reporting in Clinical Routine. Abstracts from USCAP 2021: Genitourinary Pathology (Including Renal Tumors) (422-531). Modern Pathology. 34(2): 647-766. doi: https://doi.org/10.1038/s41379-021-00760-7

Date of search	July 2023
Concepts used	Artificial Intelligence, AI, Machine learning; Prostate cancer; biopsy

Proposed research question and evidence selection criteria (if selected)

Proposed Research question	What is the clinical and cost effectiveness of AI-assisted diagnosis of prostate cancer from prostate biopsy, compared to standard care?

	Inclusion criteria	Exclusion criteria
Population	People with suspected prostate cancer undergoing prostate biopsy	
Intervention	AI-assisted diagnosis of prostate biopsy	AI-assisted diagnosis of MRI scans
Comparison/ Comparators	Prostate biopsy diagnosed without AI assistance.	
Outcome measures	Sensitivity Specificity Accuracy Procedure time Use of immunohistochemistry Time to diagnosis Number of repeat biopsies	
Study design	Any: Prioritise systematic reviews/RCTs Include unpublished case studies of high relevance to Wales Economic evaluations Exclude opinion pieces, letters, conference abstracts	

Proposed specialities	**Cancer; Genitourinary system**