Evidence overview: MRI fusion biopsy for people with suspected prostate cancer

This overview summarises the main issues the diagnostics advisory committee needs to consider. It should be read together with the final scope and the diagnostics assessment report.

1 Aims and scope

The NICE guideline on prostate cancer recommends that a multiparametric magnetic resonance imaging (mpMRI) test should be offered to people with suspected clinically localised prostate cancer. People with a significant lesion should be offered a mpMRI-influenced prostate biopsy.

Targeted biopsies, which take only a small number of tissue samples or cores, are done for people who have a suspicious lesion identified by MRI. A systematic biopsy approach, in which multiple samples are taken from different regions of the left and right side of the prostate, can be done alongside targeted biopsy. This approach can be taken if lesions are equivocal (when radiologists are unsure if it is cancer), and clinical suspicion is high. Clinical experts explained that the biopsy approach is dependent on the information from the mpMRI and individual clinician preference. They commented that practice in the NHS varies.

Targeted biopsies are usually done using visual estimation (cognitive fusion) to compare the previously captured MR image with the live transrectal ultrasound (TRUS) image used during the biopsy procedure to guide the biopsy needle. Because of the differences in positioning when a person has an MRI versus when they have an ultrasound, the prostate shape differs on MR and ultrasound images, which can make targeting the lesion difficult. Technologies are available to fuse the MR image onto the live ultrasound (MRI fusion biopsy) to aid biopsy targeting. MRI fusion systems are indicated for targeted biopsies of suspicious lesions where a small number of tissue

samples or cores are taken. Clinical experts commented that, as for cognitive fusion, systematic biopsies may be done alongside targeted biopsies done using MRI fusion.

Before the MRI fusion procedure, the prostate is contoured (outlined to define the shape of the prostate) on the MR images, and lesions or other suspicious areas are also marked on the image. MRI fusion software combines the MRI image to match the shape of the prostate in the live ultrasound. A 3D image of the prostate generated by the fusion software helps the clinician visualize the position of the biopsy cores.

The more samples taken during a prostate biopsy, the higher the risk of adverse events. Refined targeting of the prostate for biopsy could avoid taking unnecessary samples. This could reduce the risk of adverse events such as urinary retention, infection and sepsis following the biopsy. More accurate targeting of suspicious prostate lesions could increase prostate cancer detection rates (missing fewer cases) particularly for people with small lesions, and could reduce the number of repeat biopsies needed by reducing the risk of missing the cancer in the first biopsy.

The NICE guideline on prostate cancer recommends that a Cambridge Prognostic Group (CPG) risk category: low (CPG 1), favourable intermediate (CPG 2), unfavourable intermediate (CPG 3), high risk (CPG 4) or very high risk (CPG 5), should be assigned to all people with newly diagnosed localised or locally advanced prostate cancer. They are based on PSA result, Gleason score determined by histological analysis of the biopsy and clinical stage based on the mpMRI scan. The guideline also outlines the different treatment options for people in each category (such as active surveillance, radical prostatectomy and radiotherapy). The recommended management strategies for each CPG category are summarised in table 1 on page 29 of the diagnostics assessment report.

The aim of the assessment was to review existing evidence on the potential clinical and cost effectiveness of MRI fusion biopsy for people with suspected prostate cancer.

Decision question

Do MRI fusion biopsy systems offer a clinically and cost effective use of NHS resources in people with suspected prostate cancer?

Populations

People with suspected prostate cancer who have had an MRI scan that indicates a significant lesion (Likert or PI-RADS score of 3 or more).

Where data permits, the following subgroups may be considered:

- People with anterior lesions
- People with posterior lesions
- People who have had a previous negative prostate biopsy and are referred for a repeat biopsy within 12 months.

Interventions

The technologies included in the assessment are systems that include MRI fusion software to assist targeting of prostate biopsies. The assessment includes targeted transperineal or transrectal prostate biopsy using MRI Fusion software with or without systematic biopsy, under local or general anaesthesia:

- Artemis (InnoMedicus Artemis)
- Biojet (Healthcare Supply Solutions)
- BiopSee (Medcom)
- bkFusion (BK Medical UK Ltd and MIM Software Inc)
- Fusion Bx 2.0 (Focal Healthcare)
- FusionVu (Exact Imaging)
- iSR’obot Mona Lisa (Biobot Surgical)

- KOELIS Trinity (KOELIS and Kebomed)
- Philips UroNav (Phillips)

The software systems differ in terms of how many ultrasound devices they can be used with or the requirement for proprietary hardware such as specified ultrasound probes or workstations. Some only work with a specific brand device, others are interoperable with multiple third-party technologies. Software packages may also differ in their compatibility with fixed (using a probe holder for steadying) or freehand (without additional supports) ultrasound probes. Some systems also offer additional features such as a stabilising arm or robotic arm to help with the biopsy procedure. Further details on the technologies are in section 2.2 of the final scope.

Comparator

The comparator for the assessment is targeted transperineal or transrectal prostate biopsy using cognitive fusion biopsy (use of an MRI image to visually estimate the location of interest) with or without systematic biopsy, under local or general anaesthesia.

Healthcare setting

Secondary healthcare.

Further details, including descriptions of the interventions, comparator, care pathway and outcomes, are in the final scope for MRI fusion biopsy for people with suspected prostate cancer.

2 Clinical effectiveness evidence

The external assessment group (EAG) did a systematic review to identify evidence on MRI fusion biopsy for people with suspected prostate cancer. Find the full systematic review results in table 4 on page 59 in the diagnostics assessment report.

Overview of included studies

There were 23 studies that met the inclusion criteria for the systematic review. The EAG did network meta-analyses using 14 out of the 23 studies that compared 2 or more of software fusion, cognitive fusion, systematic biopsy or a combination of software and cognitive fusion with systematic biopsy. The EAG stated that to reduce bias in the meta-analysis, it only included prospective studies reporting within-patient comparisons (that is, 2 or more approaches done on the same person), or RCTs reporting comparative results for 2 or more of the interventions used on different people. Table 1 provides an overview of these 14 studies. The other 9 studies were included in the narrative synthesis only, most of these studies were naïve, unadjusted comparisons between cohorts.

No studies included in the meta-analyses were done in the UK. No identified studies assessed Fusion Bx 2.0 or FusionVu. No studies meeting the EAG's criteria for inclusion in the network meta-analysis (described above) were identified for bkFusion or iSR'obot Mona Lisa.

There are more details in tables 4 to 6 on pages 59 to 63 of the diagnostics assessment report.

Study Quality

Quality assessment of studies included in the meta-analysis

The EAG used the QUADAS-C and QUADAS-2 risk of bias tools to critically appraise the 14 studies used in the meta-analyses. All studies were judged to be at high risk of bias for at least 1 domain.

The EAG stated that the common issue with the between patient comparisons in the meta-analysis was that whether a test was deemed positive or not was informed at least partly by the index test, which also differed between the two arms. Also:

- Data for KOELIS Trinity was informed entirely by evidence on a previous version (KOELIS Urostation)
- Biopsy naïve patients were underrepresented in the study populations; 4 studies included only patients with at least 1 prior negative biopsy, and 3 studies had a population where around 50% had had a prior biopsy. Clinical experts highlighted that in current practice biopsy naïve would form the large majority (over 90%) of the population of interest.

There is further detail on the quality assessment of the included studies in tables 7 and 62 on pages 68 and 272 respectively, of the diagnostics assessment report. A brief overview of the study numbers, designs, sizes and quality assessment for each of the technologies included in the meta-analysis is presented below. A network diagram showing what the software fusion technologies was compared to in each study (for example, systematic biopsy, cognitive fusion) is shown in figure 4 in the diagnostics assessment report. Three included studies did not assess fusion software but were included by the EAG in the network meta-analysis as they contributed indirect evidence to the comparisons of interest.

Table 1 Overview and risk of bias of studies included in the network meta-analysis

Software fusion	Author	Study design	Study size	P	I	R	FT
Artemis	Elkhoury 2019 (PAIREDCAP)	Prospective, within-patient	248	Low	High	Low	Low
Artemis, Artemis + systematic biopsy	Filson 2016	Prospective, within-patient	538	Low	Low	High	Low
Artemis + systematic biopsy	Izadpanahi 2021	RCT, between patient	199	Low	Low	High	Low
Artemis	PROFUS	Prospective, within patient	101	Low	High	Unknown	Low

Artemis, Biojet	Rabah et al, 2021	RCT, between patient	307	Unknown	High	High	Low
BiopSee	Wegelin et al, 2019 (FUTURE)	RCT, between patient	157	Unknown	High	High	Low
Urostation* (KOELIS), Urostation + systematic biopsy	Fourcade et al, 2018	Prospective, within-patient	191	Low	Low	High	Low
Urostation Touch* (KOELIS)	Cornud et al, 2018	Prospective, within-patient	88	Low	High	Low	High
Urostation (KOELIS)*, Urostation + systematic biopsy	Alberts et al, 2018	Prospective, within patient	48	Low	Low	High	Low
Urostation* (KOELIS), Urostation + systematic biopsy	Albisinni et al, 2018	Prospective, within-patient	74	Low	Low	High	Low
UroNav, UroNav + systematic biopsy	Wajswol et al, 2020	Prospective, within-patient	169	Low	Low	High	Low
None*	Thangarasu et al, 2021	Prospective, within-patient	75	Low	Low	High	Low
None*	Kulis et al, 2020	Prospective, within-patient	63	Low	Low	High	Low
None*	Gomez-Ortiz, 022	Prospective, within-patient	111	Low	Low	High	Low

* these studies were only used to help form the network meta-analysis, P: patient selection, I: index test, R : reference standard/test(s) used to derive overall test positive rates, FT : flow and timing

Detection of cancer

The EAG extracted data on the number of people with and without cancer detected by biopsy for studies included in the meta-analysis. For people with

cancer, this was split by which grade was detected (using the ISUP Grade; see table 3 in the diagnostics assessment report). Few studies reported data on people with ISUP grade 3 or higher. Details of the included studies are in table 8 on page 71 of the diagnostics assessment report.

The EAG did a network meta-analyses based on these data. Find full detail of the network meta-analysis model in section 4.2.1 of the diagnostics assessment report. In its base case, the EAG pooled data from different software fusion technologies; that is, assuming they had identical performance. It explained that this was based on limited direct evidence comparing different fusion devices and clinical advice.

Base-case meta-analysis model 1a (base case)

The EAG included 13 studies in its base-case network meta-analysis (model 1a). Rabah et al. (2021) was excluded as it compared 2 software fusion devices (which were assumed in this analysis to have identical effects), so could not contribute to the analysis. The network meta-analysis structure of model 1a is illustrated in figure 3 on page 72 of the diagnostics assessment report.

Estimates were produced in odds ratios. For model 1a and 1b, this was a multinomial model which estimated the differential odd ratios for classification in each of 4 cancer grade categories (CPG 1, 2, 3, or 4 to 5) compared to the reference category ‘no cancer’ for software fusion compared to cognitive fusion. The EAG noted that these odds ratios are hard to interpret and so also provided the results as probabilities of people from the same cohort being classified as having no-cancer or different ISUP grades by different biopsy methods (see page 72 in the diagnostics assessment report for details). The probability results are shown below (table 2). See table 9 on page 76 of the diagnostics assessment report for odds ratio results.

2.1 Model 1a results

Table 2 Probabilities (median and 95% Credible Interval) of being classified at different ISUP grades for biopsy-naïve patients

ISUP	Cognitive†	Software†	Cognitive + SB*	Software + SB*
No cancer	0.55 (0.48, 0.62)	0.47	0.41 (0.21, 0.56)	0.36
1	0.17 (0.13, 0.22)	0.16	0.21 (0.10, 0.33)	0.22
2	0.12 (0.08, 0.16)	0.20	0.10 (0.03, 0.23)	0.22
3	0.09 (0.06, 0.14)	0.11	0.21 (0.06, 0.59)	0.12
4-5	0.06 (0.03, 0.10)	0.06	0.02 (0.00, 0.18)	0.08

†Artemis probabilities from Filson. (2016) biopsy-naïve data, *Artemis + SB probabilities from Filson. (2016) biopsy-naïve data. SB: systematic biopsy; SF: software fusion biopsy; ISUP grade: International Society of Urological Pathology

Results suggest that compared to software fusion biopsy, people having cognitive fusion biopsy may have:

- a higher probability of being classified as not having cancer,
- a similar probability of being classified as having non-clinically significant cancer (CPG 1), and
- a lower probability of being classified at higher CPGs, particularly CPG 2.

The EAG highlighted the uncertainty in their meta-analyses results, particularly for the highest category (ISUP 4 to 5) because of the small number of studies reporting results broken down by all ISUP grades, and the small number of people in this category.

There is further detail on these results on pages 76 to 77 of the diagnostics assessment report.

2.2 Meta-analysis model 1b

The EAG did a further network meta-analysis assessing specific software fusion technologies, rather than pooling across technologies (as in model 1a).

Results could not be estimated for Uronav or Uronav plus systematic biopsy because the 1 study for this device did not split diagnosis by grade of cancer. The EAG cautioned that model fit may be poor because of limited data. The network meta-analysis structure of model 1b is illustrated in figure 4 of the diagnostics assessment report on page 74.

Because of limited data results could only be produced by the model for all grades of cancer for Artemis.

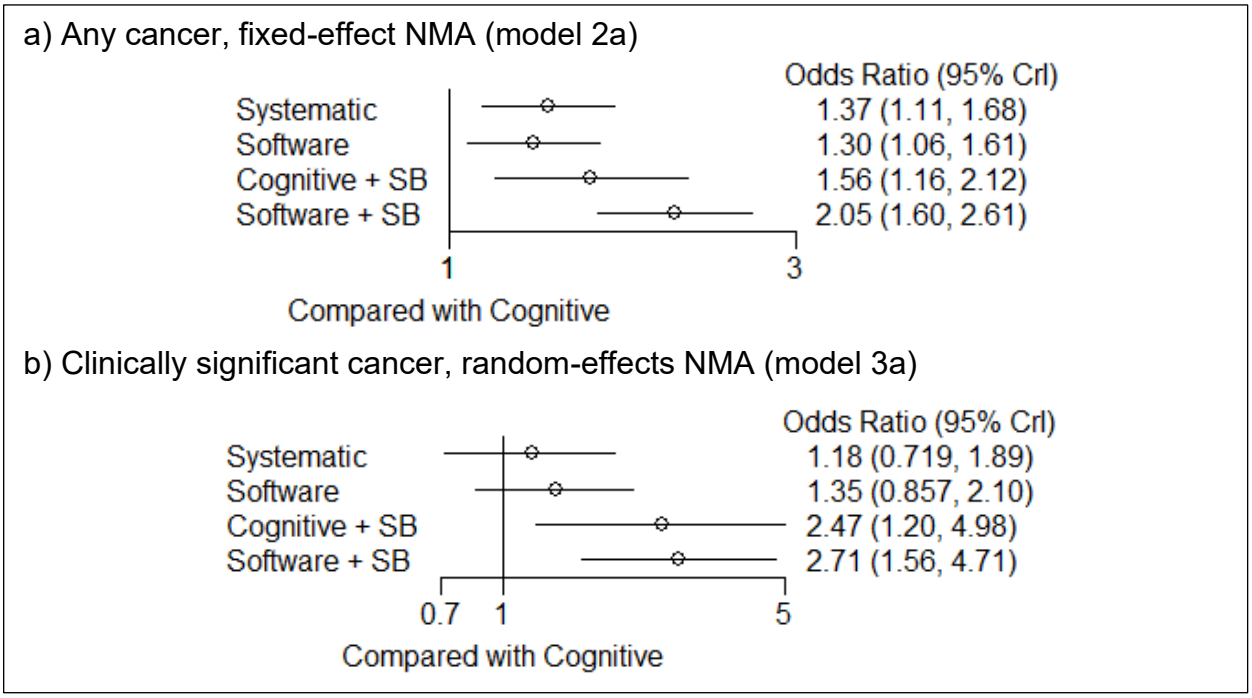
Probabilities of being classified at different ISUP grades could only be produced for Artemis (see table 67, appendix 6 of diagnostics assessment report). These were very similar to probabilities for the pooled software fusion devices (see table 2 above).

There is further detail on the results of model 1b on pages 78 to 79 of the diagnostics assessment report. Results can be found in table 66 in appendix 6.

2.3 Other meta-analyses

The EAG did further network meta-analyses looking at detection of all prostate cancer (ISUP 1 or higher versus no cancer) and clinically significant prostate cancer (ISUP 2 or higher versus no cancer or ISUP 1). Models were run pooling data from all different software fusion technologies (models 2a and 3a) and considered the different technologies separately (models 2b and 3b). The EAG highlighted that, while using essentially the same data, the models differed from models 1a and 1b (discussed above), which had 5 categories, rather than 2.

Figure 1 Odds ratio of detection (median and 95% CrI) of cancer

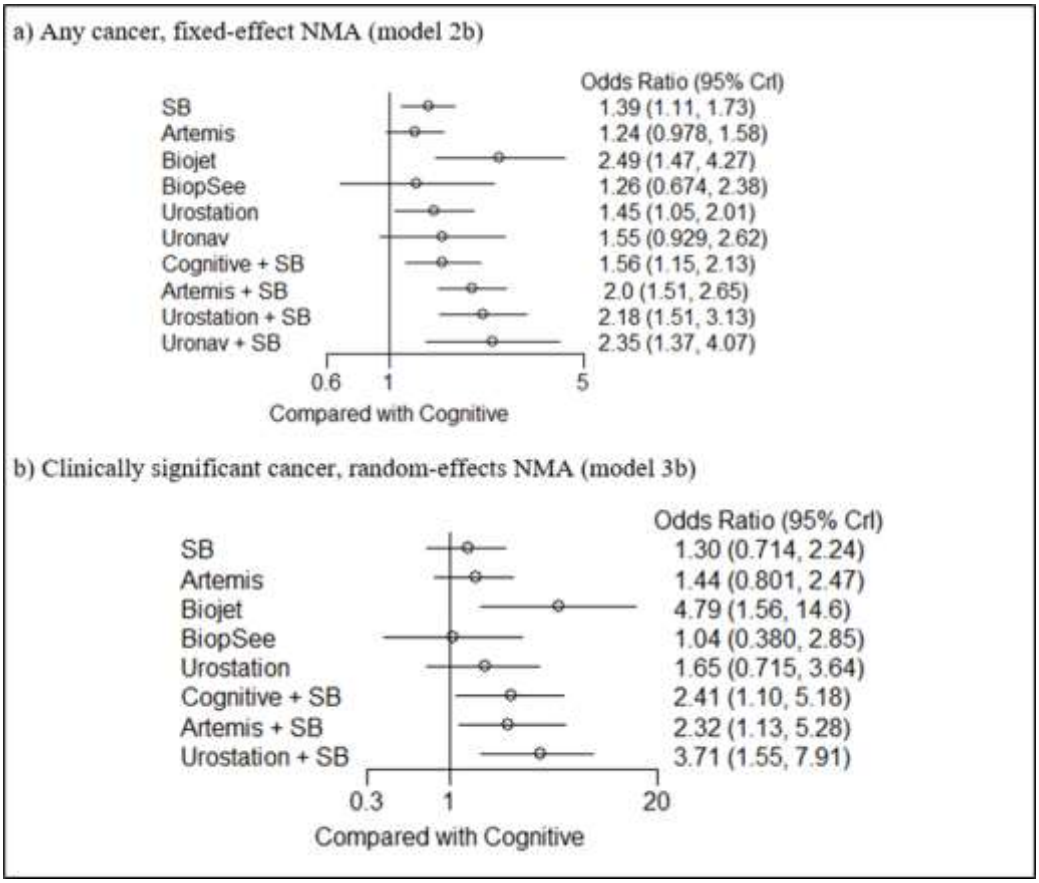


95% CrI: 95% credible interval, SB: systematic biopsy

The EAG stated results show there is evidence that software fusion biopsy may detect more cancers than cognitive biopsy alone (odds ratio 1.30, 95% Credible Interval [CrI] 1.06 to 1.61; figure 1a). Software fusion detected more clinically significant cancers also (ISUP 2, 3, 4-5) as opposed to no cancer or ISUP 1, compared to cognitive fusion biopsy but the credible interval included no benefit (odds ratio 1.35 95% CrI 0.86 to 2.10; see figure 1b; model 3a).

Figure 2 illustrates the results from model 2b and 3b for detection of any cancer and clinically significant cancer, respectively. The EAG stated that there was evidence that Biojet software fusion increases clinically significant cancer detection, as opposed to no cancer or ISUP 1, compared to cognitive fusion. The EAG highlighted that this was based on 1 study.

Figure 2. Odds ratio of detection (median and 95% CrI) of cancer with individual device effects



95% CrI: 95% credible interval, SB: systematic biopsy

There is further detail on the EAG's additional meta-analyses methods and results in section 4.4.2 of the diagnostics assessment report.

Narrative synthesis of studies not included in the meta-analysis

Nine studies reported data on prostate cancer detection but were not included in the EAG's meta-analysis because they did not use a within-patient comparison, or a randomised comparison between software fusion and cognitive fusion or between two or more eligible software fusion technologies. Therefore, the EAG considered the studies to be at higher risk of confounding compared with studies included in the meta-analysis.

All 9 studies reported a comparison between separate cohorts. Most were naïve, unadjusted comparisons, but Ferriero et al. (2022; discussed below) used propensity score matching to adjust for differences in participant characteristics.

A full description of the studies and results can be found in section 4.4.3 in the diagnostics assessment report.

Comparison of different software fusion devices

Three studies compared different fusion software against each other. One study comparing Biojet and Artemis was included in the network meta-analysis (Rabah et al., 2021). It reported that the Biojet biopsy positivity rate was significantly higher than that of Artemis (43.5% vs 21.1% respectively, p= 0.0002). The EAG cautioned that the comparison was confounded due to different biopsy routes and anaesthesia methods between the study arms. The other 2 studies (that were not included in the meta-analysis; Ferriero et al, [2022] and Sokolakis et al. [2021]) found no statistically significant difference in test positive rates of prostate cancer and clinically significant prostate cancer between software fusion devices. See section 4.4.3.3 in the diagnostics assessment report for further detail.

The EAG stated that there is insufficient evidence to conclude on the relative accuracy and clinical effectiveness of different software devices.

Subgroups

The EAG stated that separate network meta-analysis for biopsy-naïve or prior negative biopsy subgroups were not conducted due to the limited number of studies identified.

One study included test positivity rates by lesion location (FUTURE, 2019). It reported no significant differences in the rates of prostate cancer or clinically significant prostate cancer between software fusion (BiopSee) and cognitive fusion for posterior and anterior located lesions (or between peripheral and

transition zones). This study was also at high risk of confounding due to different routes and anaesthesia methods between the study arms.

Test positive rates for patients receiving a repeat biopsy following a prior negative biopsy and for biopsy naïve patients are presented in Appendix 7, Table 74 and Table 75 respectively.

Intermediate outcomes

Evidence identified on further outcomes such as biopsy positivity rates, procedure length and operator preferences can be found in section 4.5 of the diagnostics assessment report.

Usability

One study evaluated the usability of software fusion biopsy. It reported that rigid systems (Biojet and Uronav) were easier to use compared to the elastic registration system (KOELIS) for transrectal biopsies under local anaesthesia. The EAG highlighted that the study was small and at high risk of bias.

Adverse events

The EAG identified 3 studies that compared complication rates and adverse events of software fusion and cognitive fusion, and 2 compared different software fusion devices. It stated that there is no evidence of a significant difference in safety outcomes between biopsies conducted with software fusion and cognitive fusion, but highlighted that the evidence is limited by poor reporting and at high risk of confounding due to differences in biopsy routes and anaesthesia methods. Details on the 3 studies that compared adverse events with software fusion against cognitive fusion are summarised in table 20 in the diagnostics assessment report.

Ongoing studies

The EAG highlighted the ongoing IP7-PACIFIC trial (NCT05574647). This is a UK-based randomised trial (estimated n=3,600). It aims to determine whether software fusion biopsy is superior to cognitive fusion at detecting clinically

significant prostate cancers in patients with suspicious MRI in patients randomised to either mpMRI or bpMRI (biparametric MRI). It is anticipated to complete in January 2026.

3 Cost-effectiveness evidence

The EAG did a systematic review to identify any published economic evaluations of MRI fusion biopsy for people with suspected prostate cancer. Find the full systematic review results on page 296 of the diagnostics assessment report.

Systematic review of cost-effectiveness evidence

The EAG found 1 study that met its full inclusion criteria (Pawha et al. 2017). See pages 110 to 116 in the diagnostic assessment report for further discussion of this study. The study reported that strategies with cognitive fusion components had higher net health benefit than the corresponding strategies with software fusion biopsy. The EAG considered that this study has several features that limited its generalisability and relevance to the decision problem in the current assessment. This included that the study took a US societal perspective rather than that of NHS and personal and social services, the diagnostic and care pathway modelled differs from current UK practice, and that the study predated use of MRI so the population is not limited to people with a significant lesion identified by MRI (as in this assessment).

The EAG commented that the economic evidence submitted by the companies largely consisted of resource use and cost data (mostly acquisition, maintenance, and training costs) on their software fusion. This evidence was considered for the parameterisation of the model. Three pieces of evidence submitted by KOELIS and Kebomed did not meet the inclusion criteria, so were not considered further.

The EAG also did further literature searches to identify studies to support model conceptualisation and identify potential parameter values. Further

details on the value components are discussed on pages 125 to 128 and summarised in table 34 on page 126 of the diagnostics assessment report.

Economic analysis

The EAG developed a de novo health economic model. It modelled cost effectiveness for 2 populations of people with suspected prostate biopsy:

- people who are biopsy naive, and
- people who are having a repeat biopsy

Model results were presented separately for i) targeted software fusion biopsy compared with targeted cognitive biopsy and ii) combined (targeted and systematic) software fusion biopsy compared with combined cognitive biopsy.

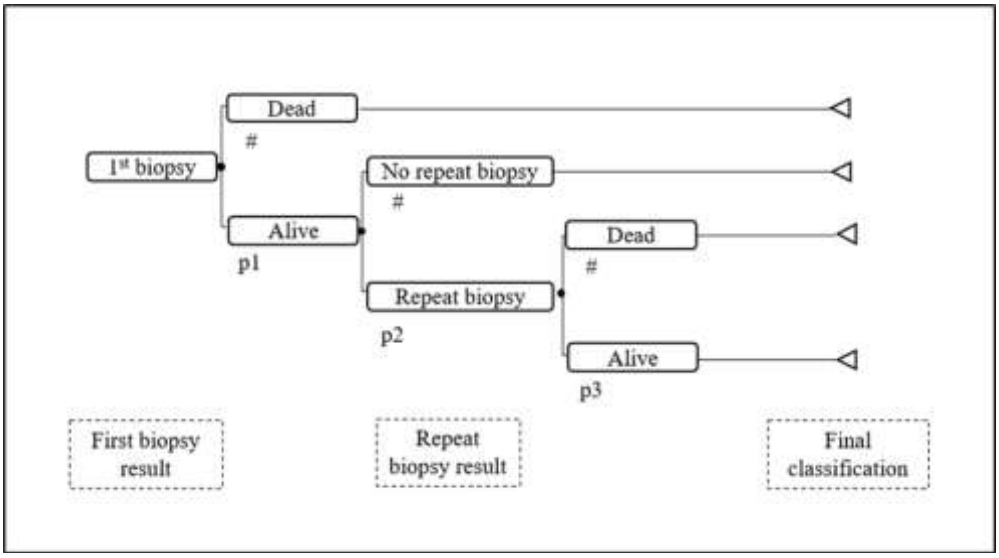
Model structure

The model consisted of a decision tree that estimates short term diagnostic outcomes and a cohort state transition (Markov) model that predicts longer term outcomes.

Decision tree

The diagnostic pathway is structured as a decision tree that captures adverse events from biopsy and incidence of repeat biopsies (which only happened if initial biopsy was negative or ISUP 1). For individuals who receive repeat biopsy, the final classification was assumed to correspond to the highest ISUP grade result of the two biopsies. Ultimately people are classified according to the result of their biopsy (or biopsies) as having no cancer or prostate cancer (ISUP grades 1, 2, 3 or 4 to 5). This may match their true condition or be a lower grade (or no cancer if they do have it). People were assumed not to get a higher-grade diagnosis of cancer than they really have (or cancer at all if they do not have it). Figure 3 shows a simplified schematic of the decision tree.

Figure 3 Decision tree schematic



Complement probability (1-probability): p1: probability of surviving the 1st biopsy, p2: probability of repeat biopsy, p3: probability of surviving 2nd biopsy.

Figure 4 summarises the 15 possible classification outcomes of the decision tree.

The EAG made a simplifying assumption that ISUP grade can be used as a proxy for an individuals' CPG score (for example, CPG1 corresponds to ISUP grade 1) to allow data from the network meta-analysis to be linked to downstream clinical management (which is in terms of CPG) in the model.

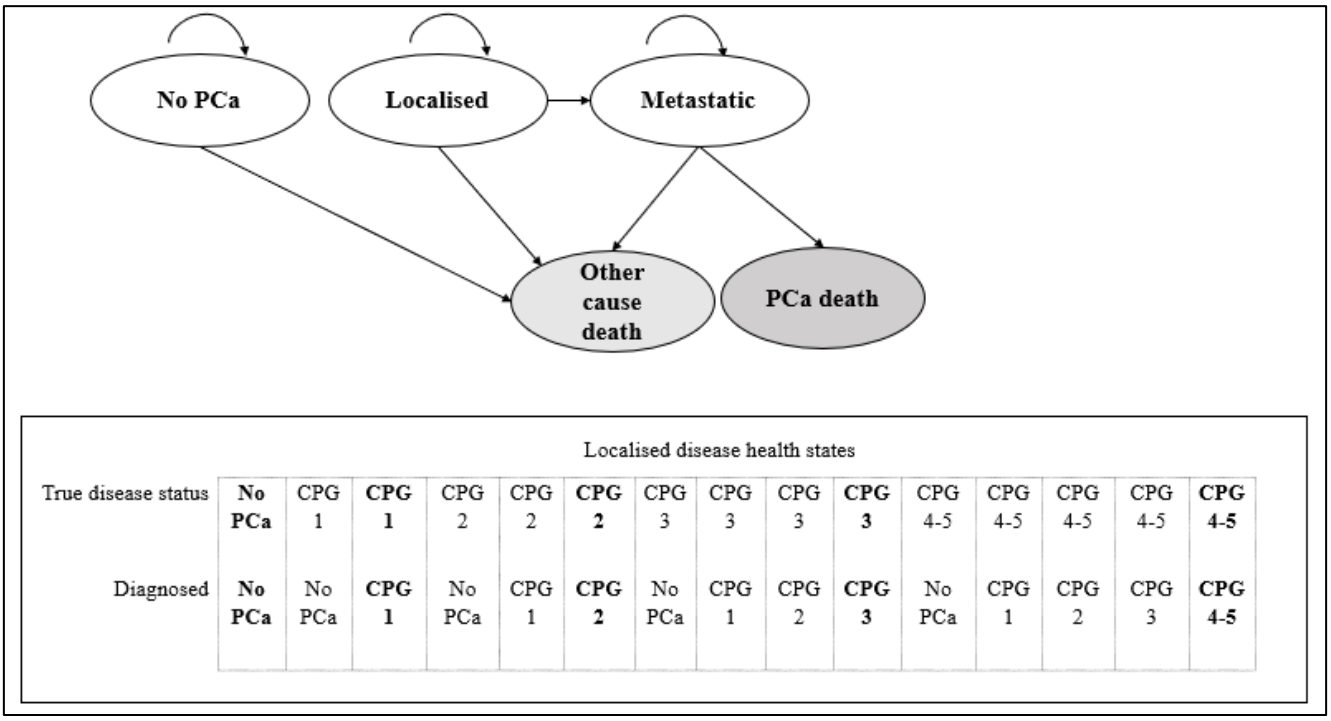
The costs and QALY pay-offs in the decision tree capture the short-term impacts of first and repeat biopsy. Cost includes the biopsy procedure (which varies depending on whether software or cognitive fusion is used) and of associated adverse events. QALY loss is also associated with biopsy procedural complications.

Markov model

The EAG's Markov model had 3 transition states based on true disease status: No prostate cancer (PCa), localised PCa and metastatic PCa, and 2 absorbing death states: death from cancer (PCa death) and death from other

causes (see figure 4). The model has yearly cycles (with a half-cycle correction applied) and a lifetime time horizon (40 years).

Figure 4 Long-term outcomes Markov model structure



CPG, Cambridge Prognostic Group; PCa, prostate cancer

People who survived the biopsy procedures in the diagnostic pathway enter the model through the no prostate cancer state if they are disease free or the localised disease state if they have prostate cancer. The people who die from biopsy enter the 'other cause' death state. People enter the model state that reflects their true disease status, but within this they are further in a 'localised disease health state' based on their diagnosis, which may be incorrect. The 15 possible localised disease health states are illustrated in the box in figure 4. Progression between CPG states was not possible in the model. It was also assumed that after 2 years in the localised disease state, people would receive treatment according to their true disease status, if misdiagnosed.

Patients with prostate cancer at model entry can remain in the localised disease health state or transition to the metastatic disease state at each yearly model cycle. The treatment or care that people get in the model is

based on their diagnosed status (rather than true disease status). What treatment or care they get determine the likelihood that they will progress to metastatic disease. Moving to the metastatic disease state increases costs for treatment, worsens people's health related quality of life and exposes individuals to risk of death from prostate cancer.

A further sub-model was used for people who develop metastatic disease. Further detail on the metastatic health state in on pages 184 to 185 of the diagnostics assessment report. Prostate cancer related mortality only applies to patients in the metastatic disease states.

Population

The population in the model is people with suspected prostate cancer who have had an MRI scan that indicates a significant lesion (Likert or PI-RADS score of 3 or more). The mean age of the initial cohort is 66 years.

Comparator

The comparator in the model was targeted transperineal or transrectal prostate biopsy using cognitive fusion with or without systematic biopsy, under local or general anaesthesia.

Model inputs

The full list of model parameters is in table 109 on pages 361 to 363, and further details on the parameterisation of the model can be found on pages 151 to 160 of the diagnostics assessment report.

The EAG used data on test performance from network meta-analysis 1a in its base case; that is, using data pooled from studies using 6 different software fusion technologies. Costs were varied across technologies (but not data on performance) in further analyses.

Diagnostic accuracy

The EAG's evidence synthesis, described in section 2.1, gave the distribution of results by biopsy technique across different cancer grades but did not

consider the accuracy in relation to a reference standard. So, the EAG developed an extension to the network meta-analysis in model 1a to include diagnostic accuracy data (this is described in full in section 4.7.1.1 in the diagnostics assessment report). An assumption made was that there would be no false positive results from biopsy. That is, if biopsy detected cancer (no matter which technique was used to obtain it) then the person did not have a lower grade of cancer (or no cancer if detected). But they could have a higher grade of cancer (or cancer, if none was detected) than identified by the biopsy.

The EAG did a further review (including studies identified in the systematic review) to identify evidence on the accuracy of MRI targeted biopsy against a 'gold-standard' test (template mapping or saturation biopsy; see section 4.7.1.3 in DAR for full details). The EAG preferred Mortezavi et al. (2018) over Zhou et al. (2018) to represent the diagnostic accuracy of software fusion in the base case analysis because it considered it more closely reflects the lower accuracy observed in UK-specific evidence sources (Zhou was used in a scenario analysis).

By combining data from Mortezavi with the results of the network meta-analysis, the economic model included people with each cancer grade in the modelled cohort, how many would be correctly identified by software and cognitive fusion in turn, and how many would have only a lower grade detected (and, if more than 1 lower grade possible, how people would be classified over these). Full results can be found on page 142 in the diagnostics assessment report.

The results suggested that targeted software fusion increases the probability of detection at the correct grade of cancer across all ISUP grades, particularly for ISUP grades 1 and 2 (approximately 5% increase in each with software fusion compared to cognitive fusion). For combined biopsy, the results suggest software fusion increases the probability of detection at the correct grade of cancer for ISUP 1, 2 and 4 to 5. The EAG noted that results for both

sets of comparisons (targeted and combined) are not statistically significant, as the 95% credible intervals for each probability largely overlap when comparing the accuracy for software and cognitive fusion. See page 142 of the diagnostics assessment report for more detail.

Modelling of long-term outcomes

Calibration was used to estimate transition probabilities for progression from localised to metastatic disease, based on a person's CPG (true disease status) and treatment received (conservative management or radical treatment; informed by diagnosis received; see section below). The EAG assumed no one was treated with watchful waiting. People either had radical treatment (radical prostatectomy or radiotherapy) or active surveillance. See section 6.3.4.1 of the diagnostics assessment report.

Treatment effects on localised and locally advanced prostate cancer

Treatment allocation based on CPG score was sourced from Parry et al. (2020), a study on the differences in localised and locally advanced treatment according to CPG in clinical practice in England. More people were assumed to have radical treatment at higher CPG.

Further detail on the distribution of treatments to different CPG groups are summarised in table 42 on page 161 of the diagnostics assessment report.

For people with a grade of cancer, modelled transition probabilities from localised to metastatic cancer were lower for people with a correct diagnosis (rather than cancer incorrectly diagnosed at a lower grade, or missed altogether).

Treatment effects on metastatic prostate cancer

Details on these parameters are in table 94 on page 345 of the diagnostics assessment report.

Adverse events

The model captured adverse events related to the biopsy procedure**.** No difference in biopsy related adverse events was modelled based on whether cognitive or fusion biopsy was used. Longer term adverse events did vary by which treatment was used, which was affected by which CPG people were diagnosed as having (and therefore cognitive or software fusion technique used). The EAG modelled incidence of erectile dysfunction, urinary incontinence and bowel dysfunction.

See tables 64 and 95 in the diagnostics assessment report for a full list of adverse events model inputs and data sources.

Costs

The full details of costs used in the model are detailed in section 6.3.7 to 6.3.14 of the diagnostics assessment report.

Cost of fusion software

The base-case model used an average cost of fusion software across technologies. Individual technology costs were used in a further analysis. Table 3 details the costs of individual fusion software and cognitive fusion for first-time and repeat biopsies. Cost per biopsy for software fusion was higher because of the cost of the technology, cost of staff training time, and increased costs relating to doing the biopsy which was assumed to take 10 minutes longer.

Details of the biopsy costs are on pages 180 to 182 of the diagnostics assessment report.

Table 3 Cost of software fusion and cognitive fusion biopsy use per biopsy

Technology	1st biopsy cost	Repeat biopsy cost
bkFusion	£430.89	£481.00
FusionVu	£452.89	£503.00
KOELIS Trinity	£434.62	£484.80
BiopSee	£376.47	£426.39
Fusion Bx 2.0	£441.79	£491.92
Average cost software fusion	£427.33	£477.42
Cognitive fusion	£332.13	£380.05

Health-related quality of life

Table 4 summaries the disutilities that were applied for biopsy related adverse events (for between 1 and 3 months), treatment associated adverse events for localised disease (for example, bowel dysfunction; applied for the duration of stay in the health state) and for people who progressed to metastatic disease (applied for the duration of stay in the health state). These were based on a diagnostics assessment report produced for a previous piece of diagnostics guidance (see page 367 in the diagnostics assessment report).

Table 4 Disutility values for biopsy and treatment related adverse events.

Biopsy adverse events	Disutility value
Mild adverse events	-0.289
Leading to non-elective admissions	-0.490
Death	-0.490
Baseline health state	Age and sex dependent
Localised treatment	
Sexual dysfunction	-0.0230
Urinary dysfunction	-0.0950
Bowel dysfunction	-0.2090
Metastatic disutility	- 0.137

Base-case results

Base-case cost effectiveness for MRI fusion software

The base-case probabilistic analysis results are shown in table 5. The deterministic results are presented in table 55 on page 192 of the diagnostics assessment report. Probabilistic and deterministic results were similar.

Table 5 Probabilistic base-case cost-effectiveness results for i) targeted and ii) combined biopsy

i) Strategy	**Total QALYs**	**Total Costs**	**ICER**	**NHB at £20,000**	**NHB at £30,000**	**Probability CE at £20,000**	**Probability CE at £30,000**
Targeted CF	8.30	£28,179	-	6.89	7.36	0.36	0.32
Targeted SF	8.31	£28,245	-	6.90	7.37	0.64	0.68
Targeted	**Inc QALYs***	**Inc Costs***	**-**	**INHB at £20,000****	**INHB at £30,000****	**-**	**-**
SF versus CF	0.01	£65	£6,197	0.01	0.01	-	-
ii) Strategy	**Total QALYs***	**Total Costs***	**ICER****	**NHB at £20,000****	**NHB at £30,000****	**Probability CE at £20,000****	**Probability CE at £30,000****
Combined CF	8.30	£28,164	**-**	6.89	7.36	0.27	0.25
Combined SF	8.32	£28,213	**-**	6.91	7.38	0.73	0.75
Combined	**Inc QALYs***	**Inc Costs***	**-**	**INHB at £20,000****	**INHB at £30,000****	**-**	**-**
SF versus CF	0.02	£49	£2,199	0.02	0.02	**-**	**-**

SF: software fusion, CF: cognitive fusion, CE, cost-effectiveness: INHB: incremental net health benefit, Inc: incremental, NHB: net health benefit

In both the targeted biopsy and the combined biopsy analyses, software fusion is more costly, but yields more QALYs than cognitive fusion.

Targeted biopsies using software fusion cost about £95 more per biopsy than with cognitive fusion. But software fusion also leads to fewer repeat biopsies compared to cognitive fusion; this has a small impact on incremental costs

and QALY loss. In the longer-term model, more QALYs were generated with use of software fusion. The EAG stated this is likely because of reduced numbers of people progressing to metastatic cancer.

The EAG provided model results showing what proportion of people with different grades of cancer were detected by software and cognitive fusion. Software fusion increased the correct classification across all CPGs compared to cognitive fusion, particularly for CPG 2 (21% more) and CPG 1 (16% more). See section 3.1 and table 2 in the EAG's addendum for further details.

The EAG highlighted that increased correct detection of CPG1 does not improve cost effectiveness. There is little benefit in terms of reducing risk progression to metastatic disease for CPG1 detected versus undetected.

The improved cost effectiveness of software fusion is driven by improved detection of CPG 2, 3 and 4 to 5. The largest effect is from CPG 2 (which, of these grades of cancer, was present at the highest prevalence).

The EAG also highlighted uncertainty about the prevalence at each cancer grade used in the model (particularly at higher cancer grades, where there was little data).

Further break down of results and how classification and misclassification impacted on cost effectiveness results can be found in the EAG's addendum, sections 3.2 and 3.3.

When targeted fusion was combined with systematic biopsy, there were more cost savings and health benefits in the long-term model for software fusion compared to cognitive fusion. The EAG suggested that this may be partially driven by the increased detection of CPG 4 to 5 for software fusion compared to cognitive fusion, which is proportionally higher in the combined biopsy comparison.

Further detail on the main drivers for the higher cost and associated QALYs are on pages 194 to 195 of the diagnostics assessment report.

Using different costs for software fusion technologies

The EAG also ran the base-case analysis using different costs for software fusion technologies that were provided to NICE. Results are shown in table 6. Table 55 summarises the probabilistic base-case results on page 192 of the diagnostics assessment report. No costs were provided for Artemis, Biojet, iSR'obot Mona Lisa, or UroNav Fusion Biopsy System, so these tests were not included in this analysis.

Table 6 Deterministic base-case cost-effectiveness results: targeted software fusion technologies pairwise comparisons with targeted cognitive fusion

Strategy	Inc costs	Total QALYs	Total Costs	ICER vs. cognitive fusion*	NHB at £20,000*	NHB at £30,000*
Targeted CF	-	8.29	£28,364	-	6.87	7.34
Targeted software fusion	£98	8.30	£28,428	£5,623	6.88	7.35
Targeted bkFusion	£101	-	£28,431	£5,954	6.88	7.35
Targeted FusionVu	£125	-	£28,454	£8,001	6.88	7.35
Targeted KOELIS Trinity	£105	-	£28,435	£6,302	6.88	7.35
Targeted Fusion Bx 2.0	£113	-	£28,443	£6,968	6.88	7.35
Targeted BiopSee	£44	-	£28,374	£890	6.88	7.35

*Per additional QALY; CE, cost-effectiveness; INHB, incremental net health benefit; Inc, incremental; NHB, net health benefit, CF: cognitive fusion

The EAG stated that results were similar for combined biopsy.

Sensitivity analyses

Threshold analysis

The EAG did a threshold analysis to determine the maximum cost at which software fusion is cost-effective at maximum acceptable ICERs of £20,000 and £30,000 per QALY. Below £586 and £695 cost per biopsy targeted software fusion biopsy was cost effective at £20,000 and £30,000 per QALY, respectively. For combined software fusion biopsy this increased to £874 and £1,116, at £20,000 and £30,000 per QALY, respectively.

Results of the threshold analyses are illustrated in figures 17 and 18 on pages 375 and 376 of the diagnostics assessment report.

Subgroup analysis

The ICERs for the EAG's subgroup analysis of people who have had a previous negative biopsy were similar to the base-case results. For full analysis see section 6.3.2 of the diagnostics assessment report. There was an increased likelihood of correctly classifying individuals with prostate cancer across all CPGs for software compared to cognitive fusion in both the targeted and combined biopsy analysis.

Full results are in table 57 on page 199 of the diagnostics assessment report.

Scenario analysis

Table 7 outlines the 7 scenario analyses done. ICERs for all scenarios were below £12,000 per QALY, except for scenario 4. In this scenario the EAG modelled no difference in detection of cancer between software fusion and cognitive fusion, but reduced the likelihood that people having software fusion would need a repeat biopsy. The EAG commented that the small incremental benefits from fewer repeat biopsies were insufficient to offset the higher costs of software fusion biopsy compared to cognitive fusion.

Table 7 Scenario analysis cost-effectiveness results: i) targeted and b) combined biopsy

Scenario	Inc QALYs*	Inc Costs*	ICER per QALY
Base-case	0.01	£63	£5,623
1. PAIREDCAP (2019) baseline	0.01	£39	£4,428
2. Zhou (2018) diagnostic	0.03	£83	£3,105
3. Degradation of repeat biopsy accuracy	0.01	£63	£5,477
4. Software fusion as quality assurance	0.000099	£87	£875,042
5. Radical treatment for all identified CPG equal or more than 2	0.03	-£117	Dominates
6.1 Throughput (150 per year)	0.01	£129	£11,425
6.2 Throughput (450 per year)	0.01	£42	£3,689
Base-case	0.02	£49	£2,199
1. PAIREDCAP (2019) baseline	-	-	-
2. Zhou (2018) diagnostic	-	-	-
3. Degradation of repeat biopsy accuracy	0.03	£46	£1,801
4. Software fusion as quality assurance	0.000139	£81	£582,123
5. Radical treatment for all identified CPG equal or more than 2 and conservative treatment for CPG 1	0.05	-£300	Dominates
6.1 Throughput (150 per year)	0.03	£110	£4,275
6.2 Throughput (450 per year)	0.03	£26	£1,009

*Software fusion compared to cognitive fusion, CPG: Cambridge Prognostic Group, QALY: Inc: Incremental, Quality-adjusted life year, ICER: incremental cost-effectiveness ratio

A full description of the scenario analyses results is on pages 200 to 201 of the diagnostics assessment report.

4 Summary

Clinical effectiveness

Varying levels of data were identified for the different technologies. The EAG judged all studies to be at high risk of bias for at least one of the quality assessment domains. No evidence was identified for Fusion Bx 2.0 and

FusionVu. The bkFusion and ISR'obot Mona Lisa studies were not included in the meta-analysis because they were non-randomised.

In the network meta-analysis, there were 5 studies for Artemis (n=1,393 total), 1 for Biojet (n= 307), 1 for UroNav (n= 169), 1 for BiopSee (n= 157), and 4 for KOELIS (n=401). All studies included for KOELIS used a previous version of the current device, which does not have an integrated ultrasound.

The EAG pooled data from different software fusion technologies in its base case analysis. Software fusion biopsy, compared to cognitive fusion, seemed to detect more people with higher ISUP grades. The results were similar when comparing the same biopsy methods in combination with systematic biopsy.

There was evidence that software fusion biopsies identify more cancer overall than cognitive fusion biopsies (OR 1.30, 95% CI 1.06 to 1.61; see table 70 in the diagnostics assessment report on page 283). There was slightly higher detection of cancer and clinically significant cancer by software fusion plus systematic biopsy compared to cognitive fusion plus systematic biopsy, but results were similar.

There was not enough evidence to perform subgroup analyses to obtain diagnostic performance estimates on lesion location, between biopsy naïve and prior negative biopsy patients, or according to operator experience.

Cost effectiveness

In the EAG's base case analysis, targeted software fusion biopsy had a probabilistic ICER of £6,197 per QALY, when compared with cognitive fusion. The ICER was lower when comparing a combination of software fusion and systematic biopsy with cognitive fusion and systematic biopsy, with a probabilistic ICER of £2,199 per QALY. The probability of being cost-effective for both targeted and combined software fusion at maximum acceptable ICERs of £20,000 and £30,000 per QALY was 64% or higher.

Scenario analyses for targeted and combined software fusion resulted in only small changes in ICERs (all below £12,000 per QALY), except for 2 scenarios. First (scenario 4), when the only benefit of software fusion is assumed to be to inform the selection of cases for repeat biopsy. That is, no difference in cancer detection between software and cognitive fusion is assumed. This increased the ICERs to over £580,000 per QALY gained for software fusion. In scenario 5, the base-case was changed so that all individuals diagnosed with CPG 2 or higher had radical treatment at initial diagnosis and those diagnosed CPG 1 had conservative treatment (and do not switch for radical treatment). In this scenario, software fusion dominated cognitive fusion.

5 Issues for consideration

Clinical effectiveness

Limitations of clinical effectiveness evidence

The EAG judged all studies used in the meta-analysis to be at high risk of bias. It stated that no high-quality RCTs have been published. For its economic model, the EAG needed data broken down by detection at different grades of cancer, which was not reported by all studies. This meant that the EAG's meta-analyses used imprecise estimates where data was most sparse, particularly at higher grades of cancer where few cases were detected and reported only in a few studies. The EAG commented that there was large uncertainty in all estimates due to the limited evidence. Meta-analyses showed moderate heterogeneity that could not be explained by differences in individual software fusion devices. In addition, the EAG stated that a lack of evidence on prevalence of prostate cancer by CPG across the population of interest, complicated interpretation and lowered its confidence in the results.

Differences between devices

Evidence levels varied across different software fusion technologies. The EAG combined data from different technologies in its base case analysis, based on

advice from clinical experts. But there was limited data directly comparing the different technologies. The EAG stated that this made it difficult to draw conclusions for relative accuracy of individual devices. When individual software fusion technologies were assessed in network meta-analysis, only Artemis had enough data to allow estimates to be produced for all grades of cancer. Three technologies included in the network meta-analysis (Biojet, BiopSee and UroNav only had 1 study each). There were 4 studies for KOELIS Urostation (a previous version of the KOELIS Trinity).

Outcomes with no identified data

The EAG found no evidence on the following outcomes included in the scope: biopsy sample suitability/quality, number of repeat biopsies performed, procedure completion rates, software failure rate, time to diagnosis, length of hospital stay, time taken for MR image preparation, subsequent prostate cancer management, re-biopsy rate, hospitalization, overall survival, progression free survival, patient- and carer reported outcomes (including tolerability and health-related quality of life), barriers and facilitators to implementations. During scoping it was raised that the value of the technology may differ in various subgroups according to the location of the lesions, operator experience or between biopsy naïve and prior negative biopsy patients. The evidence identified was too limited for the EAG to draw conclusions on the performance of software fusion in these subgroups.

Cost effectiveness

Improved detection of prostate cancer drives cost effectiveness for software fusion

Cost effectiveness estimates for software fusion were below £12,000 per QALY or dominating, in all scenario analyses except where no benefit of detecting prostate cancer was modelled (scenario 4). The EAG's network meta-analysis underpinned the economic model, so any limitations of this

analysis (see previous section) therefore apply to the cost effectiveness estimates.

The EAG commented that the cost-effectiveness of software fusion is driven by i) comparative diagnostic accuracy for evidence that is particularly sparse (especially for cancer grades above CPG 2), and by prevalence, which is also uncertain due to limited evidence.

Cost effectiveness estimates for different software fusion technologies

Cost effectiveness results were only provided based on a pooled estimate of performance using data from several different software fusion technologies. It was unclear how different software fusion technologies performed relative to each other (see earlier section). The EAG did model individual costs for the different technologies which did not have much impact on the ICERs. However, not all technologies included in the network meta-analysis (Biojet, Artemis and UroNav) provided costs.

Ongoing studies

The ongoing PACIFIC RCT is a UK trial assessing the use of software and cognitive fusion (see section 2). Its primary outcome is the proportion of clinically significant cancers (defined as ISUP 2 or above) detected amongst people who had a biopsy with a suspicious MRI (MRI score 3, 4, 5 on either Likert or PIRADS schema). [redacted] This study is reported on ClinicalTrials.gov (NCT05574647, accessed 14 November 2022) as not yet recruiting, with an estimated study completion date of January 2026.

6 Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

- All people with cancer are covered under the disability provision of the Equality Act (2010) from the point of diagnosis.
- Radical treatment for prostate cancer can affect fertility.
- Prostate cancer is more common in older people, people of African family background and people with a family history of prostate cancer.
- People with learning disabilities are often disproportionally impacted by cancer.
- Trans women should have access to prostate biopsy if needed.
- Enlarged prostate is most common in older people and prevalence may vary by ethnic background.
- Some people are at a greater risk of complications during general anaesthetic. This might include people with diabetes, older people, people who are overweight, people with heart disease and people with high blood pressure.
- The technology is contraindicated for people who cannot have an MRI, for example, people with implanted non MRI-compatible pacemakers, intracranial aneurysm clips and cochlear implants.
- The technology may not be suitable for people who are not eligible for a transrectal ultrasound, for example people who have had a proctectomy (removal of the rectum). This is because the technology overlays transrectal ultrasound images with the MRI scan. This may be more prevalent in people who have inflammatory bowel diseases, such as ulcerative colitis.

7 Implementation

IT issues

Interoperability issues (the ability of computer systems or software to exchange and make use of information from the devices) and capacity issues in NHS Trusts may be a potential barrier to implementing MRI fusion software. Disruptions to local IT networks could prevents the download of MRI images

into the MRI Fusion systems. Some devices may offer non-networked image transfer options.

Clinical expertise

The level of clinical expertise of the operator may influence their choice to use the technology. A perceived limit in the advantages of using the technology, coupled with the slightly longer time to prepare images, may deter use.

Differences between monitoring devices

The devices vary in terms of their compatibility with third party ultrasounds and freehand and stabilised biopsies. Utilisation of rigid and elastic estimation also differs, and devices may be better suited for specific populations depending on the location of their lesion. Local practices may delay purchasing these systems until their existing ultrasounds need replacing.

8 Authors

Vera Unwin
Topic lead

Thomas Walker
Technical adviser

Glossary

Multiparametric MRI

Multi-parametric magnetic resonance imaging (mpMRI) works by creating detailed images of the prostate that enable clinicians to detect suspected prostate cancer. It uses different MRI settings including T2-weighted imaging, diffusion-weighted imaging (a different MRI setting) and dynamic contrast enhancement.

Cambridge Prognostic Group

Prostate cancer is categorised into 5 risk groups: Cambridge Prognostic Group (CPG) 1 to 5.

Prostate-specific antigen

Prostate-specific antigen (PSA) is a protein produced by normal, as well as malignant, cells of the prostate gland. The blood level of PSA is often elevated in people with prostate cancer. The PSA test is used to assess the risk of prostate cancer in by measuring PSA in the blood.

Biparametric MRI

Biparametirc MRI uses T2-weighted images combined with diffusion-weighted imaging (a different MRI setting).

Cognitive fusion

The use of an mpMRI scan to overlay a mental image of where regions of interest in the prostate might be on a live-ultrasound image during the procedure.

Transrectal ultrasound guided biopsy

This is where core biopsies of the prostate are taken via the rectum under local anaesthetic.

MRI fusion biopsy

The overlay of MRI scan images on to live ultra-sound images, to assist targeted prostate biopsies. MRI fusion systems comprise specialist software with or without proprietary hardware.

Targeted biopsy

A biopsy approach which utilises an MRI scan to identify lesions from which a small number of tissue samples or cores are taken.

Systematic biopsy

A biopsy approach where multiple samples are taken from different regions of the left and right side of the prostate.

Transrectal biopsy

This is a biopsy approach where core samples are taken from the prostate via the rectum.

Transperineal biopsy

This is a biopsy approach where core samples are taken from the prostate via the perineum.

Gleason score

The Gleason grading system is used to help evaluate the prognosis of men with prostate cancer using samples from a prostate biopsy. Together with other parameters, it is incorporated into a strategy of prostate cancer staging which predicts prognosis and helps guide therapy. Gleason scores of 5 or lower are not used. The lowest Gleason score is 6, which is a low-grade cancer. A Gleason score of 7 is a medium-grade cancer, and a score of 8, 9, or 10 is a high-grade cancer.

Proctectomy

The surgical removal of the rectum and all or part of the colon.