NICE National Institute for Health and Care Excellence

Economic evaluation of multi-indication health technologies: proposed approaches

HTA Innovation Laboratory Report

July 2024

Disclaimer

Readers should be aware that issues and considerations outlined in HTA Lab reports cannot be taken as indicative or suggestive of any future position and will not be regarded as relevant to any future decision that may be taken by NICE.

The contents of HTA Lab reports are based on scientific knowledge that is publicly available and engagement with stakeholders at the time of writing the reports and cannot account for future changes and developments in scientific knowledge or any referenced material from external sources.

Contents

1 Executive summary

Multi-indication health technologies (MIHTs) that can be used across multiple disease areas or multiple indications are becoming more common. MIHTs include diagnostic technologies such as circulating tumour DNA (ctDNA) tests and polygenic risk scores (PRS), and digital health technologies such as virtual ward platform technologies. For example, ctDNA tests can be used to detect various types of cancer and a PRS may be used to identify people with an increased risk of cardiovascular disease or different cancers.

Using currently available approaches, economic evaluation of MIHTs requires multiple separate decision models in each disease area or indication to estimate their value for money. This increases the complexity of the analysis and the time required. This report proposes pragmatic approaches that NICE could adopt for future economic evaluations of MIHTs and provides recommendations to implement the approaches on pilot topics. The report focuses on diagnostic technologies, because these are the most common MIHTs. However, the recommendations apply to any health technology that would require multiple economic analyses to obtain a complete picture of its value for money.

To develop our proposed approaches, we did a scoping literature review of published economic evaluations and health technology assessment (HTA) reports of multi-indication diagnostic technologies, interviewed experts and held a multistakeholder engagement workshop. The literature reviews found that the complexities of modelling multi-indication diagnostics are not yet reflected in published economic evaluations or HTA reports of the technologies included in our reviews. The main findings from the expert engagement activities were that evaluation of MIHTs should:

- early in the evaluation, prioritise key use cases or indications that would be most influential in determining the value of the technology and that should be allocated resources for full economic modelling
- use structured expert elicitation approaches for prioritisation
- consider the heterogeneity and variation in the quality of evidence underpinning the use of these technologies in the different indications

- use pragmatic modelling approaches, when appropriate, to maximise the efficiency of NICE resources, including repurposing existing models
- contribute to developing a repository of economic models to allow for increased collaboration between NICE and its system partners.

Based on the findings of our work, this report recommends using structured expert elicitation to rank the value of the use cases for an MIHT and doing economic modelling for prioritised use cases. These recommendations should be implemented within pilot evaluations first to test the proposed approach so that the lessons learnt from these pilots can be used to refine the final evaluation approach.

2 Background

The NICE HealthTech programme evaluates a broad range of technologies, including diagnostics, devices and digital health technologies, as well as interventional procedures. Some technologies can be challenging to evaluate because they have multiple uses within a disease area or treatment pathway or can be used across multiple disease areas (in this report we refer to these as multi-indication health technologies [MIHTs]). The scope for assessing MIHTs can therefore be far broader than that for a typical technology that focuses on a single indication or single population.

An MIHT could have multiple indications across populations and disease areas in one of two ways:

- Firstly, it could produce diagnostic, predictive or prognostic information for each indication in a single use. Polygenic risk scores (PRS), which can derive susceptibility risks across many disease areas from a single analysis of genetic data, are an example of this.
- Secondly, it could produce diagnostic, predictive or prognostic information for a single indication in a single use that is applicable to different indications within or across disease areas. While the use of these types of technologies relate to a single indication, they might involve large capital expenditure that healthcare commissioners should justify based on the broader range of use cases the technology offers. Computed Tomography (CT) scanning and other medical imaging technologies, which are used in the diagnosis of many different types of disease or injury, or planning of treatment, fall into this category.

Assessments of the overall value of MIHTs to the healthcare system should incorporate their costs and benefits across the different indications in which they can be used. NICE has evaluated a limited number of MIHTs because of the complexities that arise when trying to estimate their value for money. A significant limiting factor is that a full assessment of value of MIHTs may require health economic models for each of the indications. When these are spread across different disease areas, or even at different points in the treatment

pathway within a single disease area, separate decision analytic models may be required.

Developing economic models for MIHTs within standard timelines and under the current resource constraints is a substantial challenge. This led to NICE evaluations that could have considered multiple indications of diagnostic technologies limiting their economic modelling to individual disease areas and applications. Examples of these include NICE's diagnostic guidance on clopidogrel genotype testing after ischaemic stroke or transient ischaemic attack, artificial intelligence (AI) software to help clinical decision making in stroke and therapeutic monitoring of TNF-alpha inhibitors in Crohn's disease. Also, other potential topics involving MIHTs are likely to have not been selected for evaluation because of the lack of resources to do economic analyses that could fully capture their value to the NHS. However, an expected increase in the number of AI-based and genomics-based technologies means that more MIHTs will be assessed by NICE in future. This means that NICE should consider and test alternative and more pragmatic approaches to the economic evaluation of these technologies.

The purpose of this report is to identify pragmatic approaches that NICE could adopt to evaluate MIHTs, with a focus on diagnostic technologies, and provide recommendations for potential implementation on a pilot topic.

The project involved 2 activities:

- Two literature reviews assessing how:
  - published economic evaluation studies have accounted for multiple indications within decision analytic modelling and
  - health technology assessment (HTA) agencies have approached the evaluation of 2 types of MIHTs with a focus on multi-indication diagnostic technologies (PRS and circulating tumour DNA [ctDNA] tests)
- Engagement with academic experts on economic evaluation and evaluation of diagnostics, both:
  - individually and

- in a workshop that reviewed the practical and methodological issues associated with the evaluation of MIHTs to consider potential pragmatic solutions.

The considerations and recommendations highlighted by these activities can help to develop practical proposals for future evaluations. This will support NICE to develop guidance on important topics that are highly relevant to the healthcare system and help make our guidance on these technologies useful to NHS commissioners.

The report focuses on diagnostic technologies because these are the most common MIHTs. However, the challenges and recommendations highlighted in this report also apply to other health technologies that require multiple economic analyses to obtain a complete picture of their value for money.

3 Literature reviews

3.1. Review of economic evaluations

A scoping literature review of published economic evaluations of multi-indication diagnostics was done to identify the methods used to model different disease areas and use cases within a single evaluation. Detailed description of the review and the included studies is provided in Appendix A.

The findings of the review showed that an overwhelming majority of studies of multi-indication technologies focused on a single indication and could not provide insight on how to pragmatically model multiple indications. All identified studies that evaluated polygenic risk scores (PRS; n=22) and circulating tumour DNA tests (ctDNA; n=12) were in a single use case or disease area, so were excluded from this review.

Of the 5 studies included in this review, 3 evaluated next-generation sequencing (NGS) technologies in multiple indications and 2 were economic evaluations of CT scanning.

The approaches taken by 2 of the NGS studies (Azimi et al. 2016; Bennette et al. 2015) provided limited insight into how NICE could approach modelling multiple indications. Firstly, the NGS tests in each of the studies identified a much higher number of genetic conditions than investigated by the authors. The process by which these were narrowed down to the final set of conditions was based on an analysis of the prevalence of the genomic variants. Secondly, both studies used a decision tree to identify the proportion of the cohort with each condition, with those then entering separate disease models to predict their long-term outcomes. These disease models were either simple payoffs identified from existing studies, estimated using models obtained from the original authors, recreated from a published study and, in one instance, a new model developed by the study authors. This greatly reduced the analytical burden but did require resources to identify published models or confirm their absence, do quality assessment and elicit expert opinion when required. The third NGS study (Schofield et al. 2019) used a different approach to examine a range of monogenic disorders by using a trial population of babies from a single hospital and extrapolating outcomes beyond the trial time horizon based on the specific monogenic disorder identified.

Although we included 2 studies from 131 economic evaluations of CT scanning uses identified in the searches, one was limited in its range of indications and focused on a single disease area but used 6 different decision models. The other was a trial-based analysis that did not do decision modelling across the different cancer types that were included in the analysis. The reasons for the lack of published economic evaluations of CT scanning including multiple indications are likely different to those for ctDNA tests and PRS. Because a medical imaging technology is primarily used for a specific application rather than for exploratory purposes, evaluations focus on this clinical application and not the investment decision on the imaging technology itself. This simplifies the task for investigators because it can allow them to ignore the acquisition and maintenance costs of the technology.

This is not the case for ctDNA tests and PRS, for which multiple indications are an inherent feature of their outputs and should be an integral part of their

evaluation. Because these are very new technologies, our searches may have identified evaluations of their early iterations in single use cases. Economic evaluations incorporating multiple indications may become more prevalent as the products get closer to launching and being presented to NICE and other HTA agencies to assess.

3.2. Review of HTA reports

The second literature review was done to understand how HTA agencies have evaluated MIHTs, focusing on ctDNA tests and PRS. Detailed description of the review and the included reports is provided in Appendix B.

A total of 7 published HTA reports were included. These included 6 reports on ctDNA tests (also termed liquid biopsy) and 1 report on PRS. The results of this review confirmed that some HTA agencies have started considering the evaluation of ctDNA tests and PRS.

Of the 6 HTA reports on ctDNA tests, 2 were published by NICE, 3 by the Canadian Agency for Drugs and Technologies in Health (CADTH) and 1 by Ontario Health. However, in the identified reports, ctDNA tests have been evaluated in single use cases rather than across a broad range of indications for multiple diseases.

The only available publication for PRS was by Australian Genomics. The report described a consideration of the clinical utility and evidence frameworks required to estimate the value of PRS. The report highlighted there are challenges regarding infrastructure needs, gathering the relevant evidence and translating the benefits of PRS to demonstrate cost effectiveness for reimbursement.

4 Expert engagement

To understand the current landscape of available evidence for the economic evaluation of multi-indication diagnostic technologies, we conducted semi-structured interviews with health economic modelling experts. We also organised a workshop to get a range of expert views on the issues presented by

MIHTs and identify methods that NICE could use to evaluate them pragmatically.

4.1. Interviews

We conducted 3 interviews with 4 experts in health economic modelling who have experience in either NICE committees or within external assessment groups (EAGs). Two interviews were completed before the expert workshop (see section 4.2) to identify further issues and challenges with evaluating multi-indication diagnostics and to calibrate the workshop agenda to focus on the most important concerns. After the workshop, a further interview was done with a representative from a non-academic group to obtain an alternative perspective on the challenges.

4.1.1. Main themes from interviews

Experts were keen to stress that the evidence underpinning technologies that produce information on multiple indications simultaneously (such as ctDNA tests and PRS) will be fundamentally different to the evidence usually included in NICE evaluations. This is because the prevalence of each of the indications will differ, with some being represented by small samples with low precision in test accuracy. This will result in greater decision uncertainty when evaluating the cost effectiveness of these technologies. Methods that allow estimates for low-prevalence indications to 'borrow strength' from higher prevalence ones are still in their infancy.

The approaches of other HTA agencies to doing economic evaluations was discussed with experts to consider what may be achievable within the current evaluation timelines if EAGs were required to model multiple indications. Experts noted that PHARMAC in New Zealand routinely takes less than 2 months to produce economic modelling results, substantially less time than EAGs have in NICE diagnostics evaluations, although this varied depending on factors such as budget impact and level of uncertainty. The different levels of analysis are outlined in section 2.4.1 of the PHARMAC guide on prescription for pharmacoeconomic analysis: methods for cost-utility analysis. However, experts also noted that expedited economic evaluations were able to use HTA reports

from countries where the technologies were already available, such as those produced by NICE.

Experts were mixed in their judgement on whether evaluations of MIHTs could be done within standard resource and time constraints. One noted that even simplified models would still need a substantial amount of clinical input to produce valid results. Others were more optimistic that useful information could be provided to the NICE diagnostics advisory committee. This is providing that pragmatic decisions are made, with clinical input, on the set of indications to be modelled and that all assumptions are transparently communicated.

4.2. Expert workshop

To capture a broad range of perspectives, views and opinions and to identify options for the future evaluation of MIHTs, internal and external stakeholders, system partners and academic groups were invited to participate in an expert workshop. The workshop was held virtually on 11 September 2023 under Chatham House rules and individual names, affiliations, and opinions are not shared. Box 1 lists the workshop expertise of attendees.

The following questions were sent to attendees ahead of the expert workshop to guide discussions.

- What approaches could NICE take to modelling multiple indications using the typical resources of a Diagnostics Assessment Programme (DAP) evaluation?
  - When and how should pre-existing economic models be used in EAG analyses?
  - When and how should results for primary indications be extrapolated to other use cases?
- What role should expert elicitation play in modelling multiple indications?
- What are the key areas for methods development for the evaluation of multi-indication diagnostics?
- What additional support or processes would be needed to support EAGs?

Box 1 Expert workshop attendees

Representatives from:

- NICE guidance-developing teams:
  - DAP
  - Technology Appraisals Programme
  - Medical Technologies Evaluation Programme
  - Centre for Guidelines
- External stakeholders including:
  - NICE Decision Support Unit
  - the UK National Screening Committee
  - NHS England, Scotland and Wales
  - international HTA agencies (the Canadian Agency for Drugs and Technologies in Health [CADTH], the Australian Government Department of Health and the Dutch HTA agency [ZIN])
  - health economists and experts in economic modelling from academic organisations
  - DAP committees
  - EAGs.

4.2.1. Main themes from workshop

Evaluations should be proportional to the system impact of the technology

Diagnostics are more likely to have a larger healthcare system impact than therapeutics because changes to treatment pathways will often require more re-organisation. Multi-indication diagnostics can have impact across different disease areas but the size of impact they have on the system will vary. Experts argued that NICE resources should be allocated proportionately so that technologies that will have substantial system impact are allocated additional resources for their evaluation (and vice versa).

Early evaluation should establish the potential value of the multi-indication technology before allocating resources for full economic modelling

Experts argued that an early evaluation of the expected value of a multi-indication technology needs establishing for each disease indication before allocating resources for the economic modelling (typically done by EAGs). In particular, it is crucial to understand which indications or use cases have the potential for the most and least net value and which indications or use cases have a greater degree of uncertainty because of a weaker evidence base or lower diagnostic test accuracy. These judgements could consider a variety of factors for which published evidence or expert opinion could inform:

- diagnostic accuracy and performance of current tests used in clinical practice
- population size for each use case
- prevalence of disease indication or risk factor
- impact of implementing the diagnostic on NHS delivery services
- suitability of current services to aid implementation
- capital investment costs for implementing the test.

Clinical experts' input will be crucial to rank indications and contextualise each use case for a multi-indication technology as part of a prioritisation exercise using a structured approach (for example, expert elicitation methods). After the lead use case has been identified, the appropriate methods for economic modelling can be agreed along with an assessment of the feasibility of modelling other use cases or extrapolating value to them from the lead use case.

Heterogeneity of evidence

The development of evidence underpinning multi-indication diagnostics is evolving rapidly. Changes to sensitivity and specificity (or other measures of test performance) during the evolution of tests has implications for the quality of evidence. These challenges are similar to those presented by digital health and AI-informed diagnostic technologies.

Cost-effectiveness modelling of MIHTs also needs to account for different types of evidence that are expected to underpin them. Companies could generate

evidence for each potential use case separately. But they are more likely to collect data across use cases in a single study, with implications for the methods used to reflect heterogeneity (such as Bayesian hierarchical models, see Murphy et al. 2021).

The level of evidence for a particular use case may be captured during the initial scoping phase of the technology, and is likely to be greater for indications with a larger population size. The availability of robust evidence to sustain every use case is unlikely, further reinforcing the need to prioritise use cases based on value.

Expert elicitation and consensus agreement for use cases with less relevant evidence sources may help to bridge the knowledge gaps. But it should also be taken into account that these also represent resource-intensive processes.

Pragmatic modelling approaches

Experts acknowledged that NICE recommendations should be based on the most robust evidence and economic modelling. Given that modelling multiple indications will likely result in a simplification of economic models with more assumptions, experts were sceptical of the committee accepting lower standards of evidence.

However, in some instances this might be preferable to using many existing models with different assumptions because the overall results will be more comparable across different indications. Considering the example of ctDNA tests, an approach may be to demonstrate cost effectiveness for the highest and lowest frequency of tumour types to provide upper and lower quality-adjusted life-year (QALY) limits and implicitly interpolate the value accruing in other tumour types that fall between these limits.

For technologies within a disease area, a similar model structure could be used to expedite this stage of the evaluation. If applicable, use cases could also be divided into:

- those that represent relatively small changes to current clinical practice and

- those that will need additional interventions and substantial pathway changes that are likely to be resource intensive (focusing modelling resources into these will capture a greater proportion of value).

Repurposing existing models

Experts considered the potential to use existing economic models to help modelling multiple use cases. Using existing models, however, should be done under strict caveats. First, a model should be developed for the UK setting and preferably 'approved' by NICE, such as models developed by EAGs or within NICE's Centre for Guidelines.

It is highly likely that models would need to be adapted, a process that is not always successful and may result in a new model still being developed. If extended to multiple existing models, experts felt that this strategy would entail extending the modelling timeline beyond the current standard 24 weeks used by NICE.

Consideration may be given to a simplified approach with a common model structure and common data sources when appropriate (for example, for a range of different cancers), with a clear justification for the modelling strategy and rationale for the underpinning assumptions.

Establish a repository of economic models

Multiple stakeholders and system partners are working to develop economic models to evaluate the cost effectiveness of multi-indication technologies. These include academic partners, the UK National Screening Committee, NHS England and Genomics England. A valuable resource would be an open-source repository of economic models for information sharing and effective collaboration. This would allow better identification of existing models that could be used by NICE to evaluate multi-indication technologies.

Summary

Experts noted that additional NICE resources should be allocated to evaluating multi-indication diagnostics given their complexity, especially when the technologies generate large system or resource impacts, or both. However,

given the current resources available, there was broad consensus that a pragmatic solution is needed and should focus efforts early in the evaluation on:

- identifying the use cases with the most potential value or highest levels of uncertainty, or both
- encouraging companies to generate robust evidence across potential use cases
- considering how the evaluation committee can extrapolate from indications with economic modelling to those without.

5 Recommendations

MIHTs are expected to be challenging for NICE and other HTA agencies to evaluate. The breadth of potential applications, across diverse populations, will require new approaches to do evaluations within existing resource and time constraints that are useful for NHS commissioners and that ensure timely access to these health technologies.

Based on our findings in the literature reviews and expert engagement, we make 2 sets of recommendations, one related to methods and the other to processes of the evaluation of MIHTs.

5.1 Methods-related recommendations

5.1.1. Prioritise use cases for modelling using expert elicitation

For most MIHTs, it would be infeasible to model all potential use cases. The complexity of modelling just a single use case can be challenging for committees and EAGs for particular disease areas using current resources and timescales. A process is therefore required to prioritise use cases for modelling with input from experts using structured expert elicitation methods.

It is recommended that the main criterion for prioritisation should be the expected value to the NHS. To elicit the expected value ranking, the following factors are likely to be the key ones for experts to consider, but factors can vary depending on the nature of the technology and its intended purpose:

- prevalence of the disease or genomic variant
- disease burden
- expected resource impact
- availability of effective subsequent management options

Ranking can be done using a qualitative (deliberative) or quantitative (multi-criteria decision analysis) approach. This ranking exercise should identify the highest and lowest value use cases, to set an indicative range of cost effectiveness, and establish the relative cost effectiveness of the intermediate

value use cases compared with the highest value use case to help extrapolation. An example of how this could work is shown in box 2.

Box 2 Example of expert elicitation approach

A hypothetical MIHT has 3 use cases (A, B and C) within its scope. Experts are asked to rank the indications in terms of 2 criteria:

1. The expected value of diagnosing and treating the indication to the health system, in terms of net health benefit and other relevant aspects of value. In practice this could be directly or indirectly informed by evidence on the size of the patient population, the severity of the indication or disease, and treatment effect size.
2. The complexity of producing economic modelling outputs. In practice, this would incorporate what is known about the typical modelling approaches within a disease area, the availability of data and the availability of existing models that could be repurposed.

The experts are also asked to score the indications on a relative numerical scale to account for the size of the differences in each criterion.

Use case	Expected value rank	Relative value score	Model feasibility rank	Relative model complexity score
A	1	100%	2	80%
B	2	70%	3	90%
C	3	50%	1	100%

An overall score is aggregated across the criteria using the relative criteria scores. Equal weights are applied to the criteria for simplicity.

Use case	Score
A	90%
B	80%
C	75%

Ranking according to the feasibility of developing detailed economic modelling should also be done. This is particularly relevant because developing detailed models for some use cases might be hindered by issues such as availability and quality of data, given the heterogeneity in the evidence available for all potential use cases. EAGs could also consider more routine use of expert elicitation to decide on the feasibility of modelling additional indications. For example, sources such as the University of York Centre for Health Economics' structured expert elicitation resources (STEER) could be validated and formalised for use by EAGs.

The final priority rank order should combine both attributes (expected value and feasibility of modelling) to guide which use case to develop a detailed model for.

5.1.2. Identify the most informative and pragmatic modelling strategy

Once the use cases have been prioritised, a modelling strategy can then be agreed with the EAG. The optimal approach will be context-specific and depend on the outcome of the priority ranking exercise. The priority ranking exercise will determine how complex each of the use cases will be to model, whether any pre-existing models can be repurposed and elicit expert opinion on how valid it is to extrapolate between use cases. The feasibility of repurposing or extending existing models should be assessed carefully, with both internal and external stakeholders noting the difficulties with adapting complex models compared with developing new models from scratch. Two types of modelling strategy could be pursued given these factors.

Strategy 1: modelling highest and lowest value use cases

Detailed models could be developed for the highest and lowest value use cases. This will help with empirically defining the full range of cost-effectiveness estimates within which the remaining use cases would fall and allow for more accurate estimation of the value in the non-prioritised use cases. Estimating the value of all other use cases relative to the modelled use cases would be done by extrapolation based on relative cost-effectiveness ranking established during the prioritisation exercise. If the experts deem extrapolation not feasible, another

approach to estimating the expected value of the non-prioritised use cases could be to use the published literature on available or relevant disease models to source aggregate or average payoff, as has been done in the published literature. Less detailed modelling could also be done for these intermediate value use cases, which could be in the form of calculating outcome payoffs.

Strategy 2: stepwise modelling approach

A stepwise approach could be followed whereby only the highest value use case (or cases) is (are) prioritised for detailed modelling and results used to guide the decision of whether any more use cases should be modelled. This means that if the highest value use case does not demonstrate cost effectiveness, then no more modelling would be required. If it demonstrates cost effectiveness, a lower value use case would then be modelled, to define the range of cost effectiveness. An earlier assessment of what constitutes 'lower value' could be made based on population size, disease severity or indicators of effectiveness (for example, diagnostic accuracy). Less complex modelling of intermediate value use cases could then also be done, if required, or extrapolation based on relative cost-effectiveness ranking would then be done, as described in strategy 1.

Further considerations

In both modelling strategies described in this section, the uncertainty around value in the non-modelled use cases will not be possible to fully quantify or characterise given that no probabilistic modelling is done for these use cases. Committees would instead have to rely on deterministic sensitivity analyses or plausible scenario analyses identified by the expert working group.

Committees would then be able to assess a broader picture of expected impact of the MIHT on outcomes and costs. This could be achieved by using the cost-effectiveness evidence on the different indications (as shown in figure 1) and allowing the committee to use the information deliberatively when making their recommendations. Alternatively, the evidence can be combined into a weighted incremental cost-effectiveness ratio (ICER), as shown in table 1, which uses the information on the expected size of the population of the indications.

A further complication is that the allocation of fixed technology costs will be spread over the range of indications the technology will be recommended for, which in turn depends on the cost-effectiveness of each use case. For example. In Figure 1, the committee would be likely not to recommend the technology for indication C due to an ICER of £50,000 per QALY. However, doing so would increase the proportion of fixed costs counted within the modelling of indications A and B and would increase the respective ICERs.

Figure 1 Using different sources for cost-effectiveness outcomes for a hypothetical MIHT

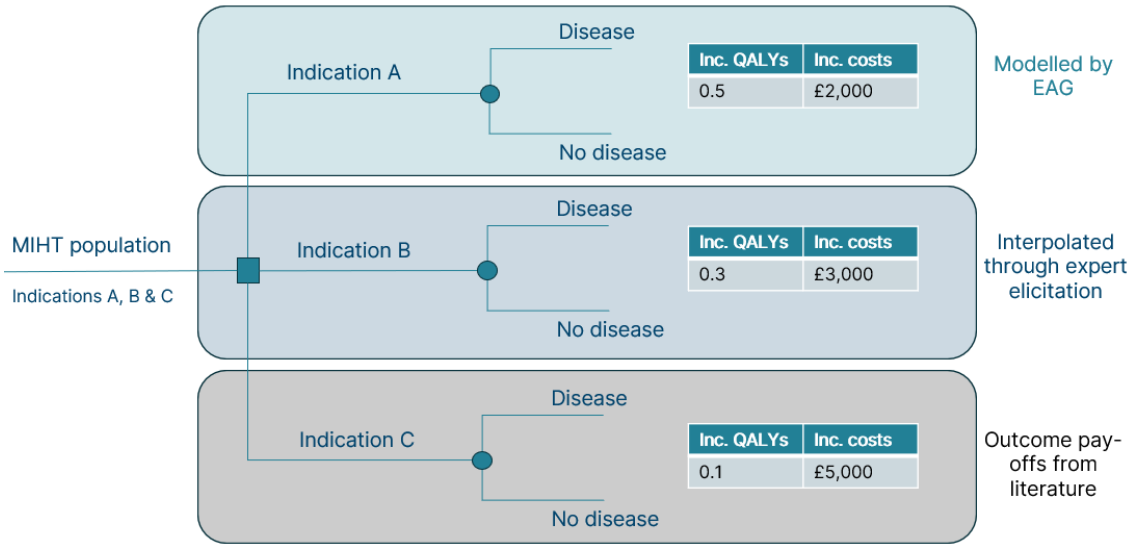


Table 1 Example of weighting cost-effectiveness estimates of a hypothetical MIHT using the indications' target population size

Indication	Population size	Incremental cost	Incremental QALY	ICER
A	1,000	£1,000	0.5	£2,000
B	10,000	£3,000	0.3	£10,000
C	100,000	£5,000	0.1	£50,000
Total	111,000	£531,000,000	13,500	£39,333

5.2 Process-related recommendations

5.2.1. Establish an expert working group to support the evaluation

An expert working group can be used throughout the evaluation process to support the NICE technical team, EAGs and committees. This group should include clinical specialists from across the different disease areas covered by the use cases, commissioners, and health economists. For example, a group for ctDNA tests would need to include oncologists specialising in different tumour sites whereas a group for PRS might need to reflect a broad range of specialties such as oncology, endocrinology and psychiatry.

In addition to the task of prioritising use cases noted in section 5.1, the working group could also provide advice to committees on how it could extrapolate value from the modelled to unmodelled use cases.

5.2.2. Allow flexible time and resource for economic modelling

Because the principal challenge in evaluating MIHTs is the extent of modelling required to robustly estimate value, processes could be adapted to allow more time for full economic modelling to be completed, if essential. The additional financial resources required could be justified based on system impact.

However, allowing more time will not address other technical barriers to modelling that may be identified during feasibility assessment such as the lack of good quality data to model some use cases. This is why it would be necessary to assess feasibility of modelling in all use cases very early on during the prioritisation exercise. The time extension could be agreed on a case-by-case basis and based on an assessment of how much additional resource would be required to model additional use cases.

5.2.2. Improve efficiency of modelling other use cases

Activities that would reduce the time and resource costs of the alternative methodological approaches proposed should also be considered.

The literature review identified that muti-indication economic evaluations used existing decision models to estimate the value of additional use cases. The

experts consulted in the workshop and interviews considered this to be a resource-intensive task that was unlikely to be feasible within current timelines. However, available models that have been accepted or developed by NICE, for example, disease pathway models, can make this more feasible. Additionally, developing libraries of accepted or validated models would make it easier to identify models that could be used in MIHT evaluations and determine how well they fit the population and resources needed for adoption.

Simplified modelling approaches that reduce the complexity of the model structure or data requirements should also be considered. This would allow the EAG to commit more resources to evaluating more use cases. Feasibility would need to be assessed on a case-by-case basis and would depend on the strength and validity of the assumptions that would result from any simplifications.

5.2.3. Allow company-submitted models

For MIHTs with large potential impacts on the NHS, a possible process change is to allow companies to submit economic models for multiple use cases. These would then be reviewed by the EAG as is currently done in medicines evaluations. This would reduce the burden on the EAG and enable the committee to consider evidence on multiple indications.

Adopting this process change would mean that the committee would not be considering independent economic modelling. It will also require consultation with companies to assess the feasibility of implementing this process change in practice.

As well as being more complex to develop, MIHT models would also be more resource intensive to critique. The time required for EAGs to review company-produced analyses of MIHTs may also need to be increased to be commensurate with this complexity.

Alternatively, responsibility could be placed on the company to do early modelling and provide high-quality evidence across the array of use cases to enable the EAG to do the modelling in more use cases.

6 Conclusion

Multi-indication health technologies (MIHTs) can be used for a wide array of purposes within or across disease areas, making their full value difficult to quantify within the constraints of a health technology assessment (HTA) process.

The complexities of evaluating multi-indication diagnostics are yet to be reflected in modelling conducted within NICE assessments. The inclusion of multiple indications in published economic evaluations is limited. Evaluations of next-generation sequencing (NGS) approaches are an exception; these used expert elicitation and pre-existing decision models to model genomic conditions pragmatically.

While circulating tumour DNA (ctDNA) tests and polygenic risk scores (PRS) are being considered by international HTA agencies, ctDNA tests have so far been considered as a single use case only and PRS implementation is in its infancy.

Experts in the economic evaluation of diagnostics had differing opinions on how NICE could do economic modelling of MIHTs within the resources of a typical NICE evaluation. However, there was consensus that a more pragmatic approach would be necessary and that more clinical expertise should guide the choice of which use cases to model in detail based on expected value.

We recommend that NICE considers adopting bespoke methods and processes for multi-indication diagnostics to ensure that its guidance is as useful as possible to NHS commissioners. The recommended changes should be discussed with key system partners and implemented within a pilot evaluation, so that the learnings can be used to finalise the approach for evaluating MIHTs.

Although our report focused on multi-indication diagnostics, the proposed approach can be equally valid to adopt and pilot for the evaluation of other MIHTs such as digital health technologies and medicines.

7 Project team

James Koh, Scientific Adviser

Roshni Joshi, Scientific Adviser

Magda Bugajska, Project Manager

Dalia Dawoud, Associate Director

8 Acknowledgements

The NICE DAP team

NICE Scientific Advice

Workshop participants

9 References

Australian Genomics (2022) Polygenic Score Incubator Project [online, accessed 21 September 2023]

Azimi M et al. (2016) Carrier screening by next-generation sequencing: health benefits and cost effectiveness. Molecular Genetics & Genomic Medicine 4(3): 292–302

Basharat and Horton (2022) The CADTH Horizon Scan Emerging Multi-Cancer Early Detection Technologies [online, accessed 9 September 2023]

Bennette CS et al. (2015) The cost-effectiveness of returning incidental findings from next-generation genomic sequencing. Genetics in Medicine 17(7): 587–595

Burgers LT et al. (2017) Cost-effectiveness analysis of new generation coronary CT scanners for difficult-to-image patients. The European Journal of Health Economics 18(6): 731–742

Cowling and Loshak (2019) The CADTH Overview of Liquid Biopsy for Screening and Early Detection of Cancer [online, accessed 9 September 2023]

Hill et al. (2020) Circulating Tumour DNA Testing for the Identification of Genetic Mutations: Diagnostic Test Accuracy and Clinical Utility (CADTH rapid response report: reference list) [online, accessed 9 September 2023]

Hubbard W et al. (2022) Development and validation of paired MEDLINE and Embase search filters for cost-utility studies. BMC Medical Research Methodology 22(1): 310

Mayerhoefer ME et al. (2020) PET/MRI versus PET/CT in oncology: a prospective single-center study of 330 examinations focusing on implications for patient management and cost considerations. European Journal of Nuclear Medicine and Molecular Imaging 47(1): 51–60

Ontario Health (Quality) (2020) Cell-free circulating tumour DNA blood testing to detect EGFR T790M mutation in people with advanced non–small cell lung cancer: a health technology assessment. Ontario Health Technology Assessment Series 20(5): 1–176

Medtech innovation briefing on plasma EGFR mutation tests for adults with locally advanced or metastatic non-small-cell lung cancer (2018)

Schofield D et al. (2019) Long-term economic impacts of exome sequencing for suspected monogenic disorders: diagnosis, management, and reproductive outcomes. Genetics in Medicine 21(11): 2586–2593

Medtech innovation briefing on signatera for detecting molecular residual disease from solid tumour cancers, Medtech innovation briefing (2022)

Appendix A: Review of economic evaluations

A.1. Objective

To understand what methods had been used to model different disease areas and use cases within a single evaluation.

A.2. Methods

The review focused on identifying the key studies that are likely to be informative. We therefore only included studies when abstract or full-text screening found that outcomes and costs across multiple indications had been estimated by the authors to evaluate the cost effectiveness of a technology.

We included economic evaluations from 2013 or later and focused on 3 types of multi-indication diagnostic technologies:

- CT scanning
- polygenic risk scores (PRS)
- circulating tumour DNA (ctDNA) tests.

This focus was necessary because there is no universally agreed terminology for describing multi-indication technologies that could be used to identify records. Also, screening all economic evaluations for all types of multi-indication diagnostics was infeasible given time and resource constraints.

To maximise the relevance of the search results, the types of economic evaluations were restricted to cost-effectiveness analysis, cost-utility analysis and cost–benefit analysis.

The searches were completed on 28 April 2023. The following databases were searched:

- MEDLINE ALL (Ovid)
- Embase (Ovid)
- EconLit (Ovid).

The search was limited from 2013 to 2023 in accordance with the requirements for the review. Search filters (precise version) from Hubbard et al. (2022) were applied to the search strategies in MEDLINE and Embase to identify cost-utility studies.

We imported the search results into systematic review screening software (EPPI Reviewer) and screened titles for their relevance for inclusion in the review. This was based on whether an economic evaluation had been done and whether multiple indications had been evaluated within the study.

We also included a further set of 7 records for title and abstract screening. These were identified from a systematic review of economic evaluations of next-generation sequencing (NGS) techniques, allowing for an analysis of a fourth type of multi-indication technology.

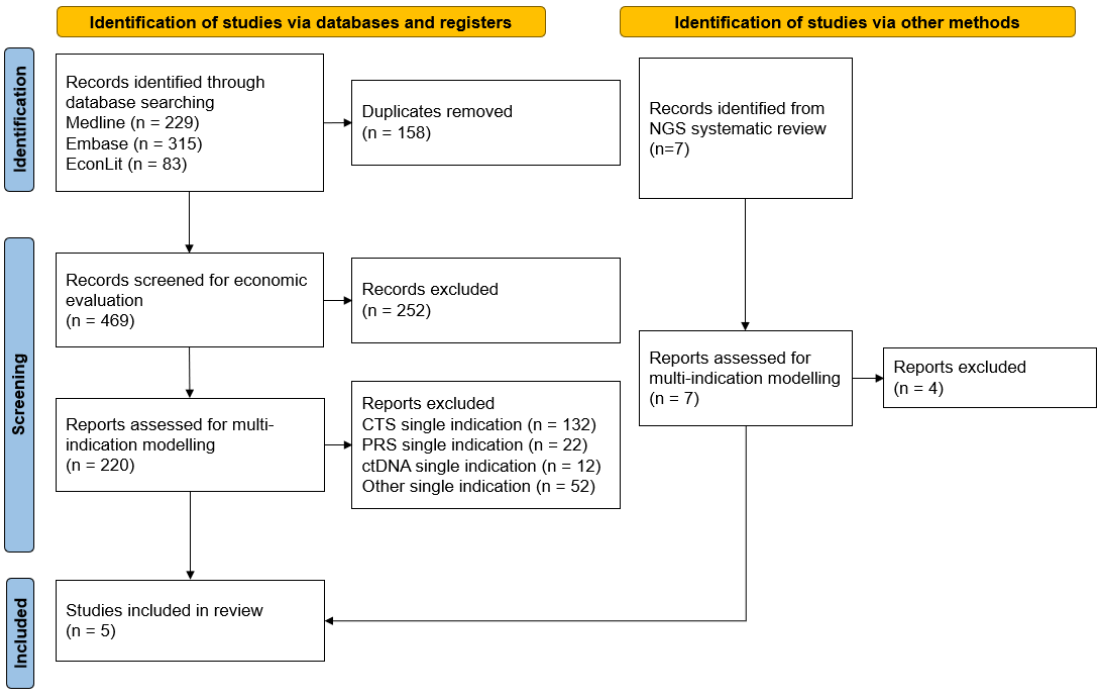
We developed a data extraction form for the following main sets of data from each included economic evaluation:

- study details (title, lead author)
- indications
- intervention and comparators
- study outcomes
- modelling approach (model type and structure, analysis of uncertainty)
- parameters and data sources
- main results.

A.3. Results

We identified a total of 627 studies through our electronic search strategies and a further 7 records from the NGS systematic review. After removing 158 duplicate records, we screened the titles and abstracts of the remaining 476 records. From these, we identified 227 records in which an economic evaluation had been done. We excluded 222 records which did not model multiple indications, resulting in 5 included studies. We have shown the details of the selection process in a PRISMA diagram (Figure A1).

Figure A1 PRISMA diagram of the economic evaluations review search and screening results



A3.1. Included studies

Of the 5 studies included in this review, 3 studies evaluated NGS techniques for screening genetic disorders and 2 studies evaluated medical imaging technologies. Two of the NGS studies focused on newborn screening, and the other looked at incidental findings from the screening of 3 adult populations (healthy individuals, people with cardiomyopathy and people with colorectal cancer). One imaging study included use of new generation CT scanners in 'difficult-to-image' patients with suspected coronary artery disease (CAD) and known CAD. The other imaging study included routine staging and follow up of people with cancer using positron emission tomography (PET)-CT scanning. No economic evaluation studies of ctDNA tests or PRS were included in this review because none accounted for multiple indications.

A3.2. Approach to modelling multiple indications

Each included study took a different approach to accounting for multiple indications.

A.3.2.1. NGS studies

In a study of carrier screening by Azimi et al. (2016), the authors compared NGS screening with genotype screening and no screening for identifying 14 genetic conditions in adult couples before conception or before birth of their child. The authors did not describe how the 14 genetic conditions were selected but noted that they were among the most prevalent and had been recommended for carrier screening. A decision tree was used to model how many affected births were averted because of the accuracy of the screening technology and the prevalence of the genetic mutation.

Couples with more accurate and sensitive screening results were more likely to either not conceive or pursue other options such as adoption or donor egg or sperm. The study calculates the life-years gained and costs saved from the number of affected births averted using estimates from the published literature for each of the 14 genetic conditions. The authors note that '[w]here data were lacking… values were estimated using conservative assumptions, by consensus among the study authors and external authorities with expertise on the disorders of interest'.

Relative to no screening, screening by NGS and genotyping averted 233 and 202 affected births per million couples, respectively. NGS dominated genotype screening and had a cost per life-year of $29,498 compared with no screening.

Schofield et al. (2019) did an economic evaluation comparing exome sequencing to identify monogenic disorders with usual diagnostic care. The study extrapolated the long-term outcomes of 80 infants aged 0 to 2 years with suspected monogenic disorders, recruited from a single tertiary paediatric centre in Australia. Both diagnostic strategies were used for all participants and cascade testing was offered to all first-degree relatives.

The study modelled the outcomes and costs for infants and first-degree relatives over a 20-year time horizon. Bespoke extrapolations were applied to each individual based on the specific monogenic disorder identified. For example, exome sequencing diagnosis resulted in people having a reduced number of blood transfusions, which was used to predict quality of life improvements in

each subsequent year. The authors did not model any improvements to mortality from diagnosis.

Exome sequencing resulted in an additional 25 diagnoses of monogenic disorders among the 80 infants and an additional 7 diagnoses in first-degree relatives, compared with usual diagnostic care. This resulted in 11.62 quality-adjusted life years (QALYs) gained at an additional cost of AU$242,154, discounted at 5% per annum, yielding an incremental cost-effectiveness ratio (ICER) of AU$20,840.

Bennette et al. (2015) evaluated the benefits of incidental findings (IFs) from using NGS technologies for other purposes. Three different cohorts of individuals were included: healthy individuals, people with cardiomyopathy and people with colorectal cancer. The 7 different types of findings that account for 95% of expected IFs were included in the modelling, narrowed from a list of 24 based on the judgement of a working group. This included presence of the BRCA1/2 genes associated with breast cancer risk and genetic mutations associated with hypertrophic cardiomyopathy.

The decision model produced in the study placed cohorts of individuals aged 45 into a decision tree. When an IF was identified, individuals entered disease-specific models to predict their lifetime QALYs and healthcare costs. The authors identified disease-specific models by searching PubMed for cost-effectiveness analyses in each of the 7 IFs. Outcome payoffs from these models were used for 1 IF in which the model results were transferable to the patient population being considered in the IF model. For some IFs, no published cost-effectiveness analyses were identified, requiring the study authors to develop new decision models. Others were recreated using the model structures and parameters from published studies because the incremental payoffs of diagnosis for the patient population of interest were not provided. Sequencing costs were included in a sensitivity analysis. Some conditions were not considered incidental to certain cohorts because they were related to the patient population and would have otherwise been discovered. For example, NGS diagnosis (and associated health and cost effects) of familial

hypercholesterolaemia, hypertrophic and dilated cardiomyopathy, and arrhythmic right ventricular cardiomyopathy were not considered when modelling the cardiomyopathy cohort.

For healthy individuals, 116 IFs would be expected per 10,000 people, resulting in 67 incremental QALYs gained and additional costs of $3.9 million. For people with cardiomyopathy, equivalent numbers would be 56 IFs per 10,000 people, 20 incremental QALYs and additional costs of $896,000. For people with colorectal cancer, equivalent numbers would be 68 IFs per 10,000 people, 25 QALYs gained and additional costs of $2.9 million. This yielded ICERs of $58,600 for healthy individuals, $44,800 for people with cardiomyopathy and $115,000 for people with colorectal cancer.

A3.2.2. Imaging technology studies

Burgers et al. (2017) evaluated a next-generation coronary CT (NGCCT) scan in 'difficult-to-image' patients with known or suspected CAD. Two strategies, with and without invasive coronary angiography (ICA), were compared with ICA alone.

Patients' outcomes and costs were estimated through a series of 5 interlinked decision models for the diagnostic pathway, CAD progression, radiation exposure from tests and treatments, non-CAD mortality, and strokes triggered by CAD interventions. Four of the 5 models had been developed by the study authors and their collaborators in previous studies, with only the stroke model developed from scratch in the current study.

The authors concluded that NGCCT alone was the cost-effective option for the suspected CAD population, both for the overall population and in all subgroups. For the known CAD population, NGCCT plus ICA was the cost-effective option for the overall population and in all subgroups.

Mayerhoefer et al. (2020) did a trial-based economic evaluation of PET-MRI scanning against PET-CT scanning in people with cancer referred for routine staging or follow up. This included 24 different types of cancer across the 330 individuals in the trial population. Primary outcomes in the study were

accurate diagnoses and changes in clinical management of patients, and did not extend to direct health measures. Costs were based on total cost of investment, maintenance and use of the imaging technology, which included cost for the system, and did not extend to downstream health resource use.

The cost per percentage point increase in diagnostic accuracy was estimated to be €14.26, and the cost per percentage point increase in correctly managed patients was €23.88. The trial was not sufficiently powered to estimate results by subgroup, with 17 out of the 24 cancers represented by fewer than 10 patients.

Appendix B: Review of HTA reports

B.1. Objective:

The objective of this review was to understand how health technology assessment (HTA) agencies have evaluated specific multi-indication technologies that are expected to become more widely used in the future.

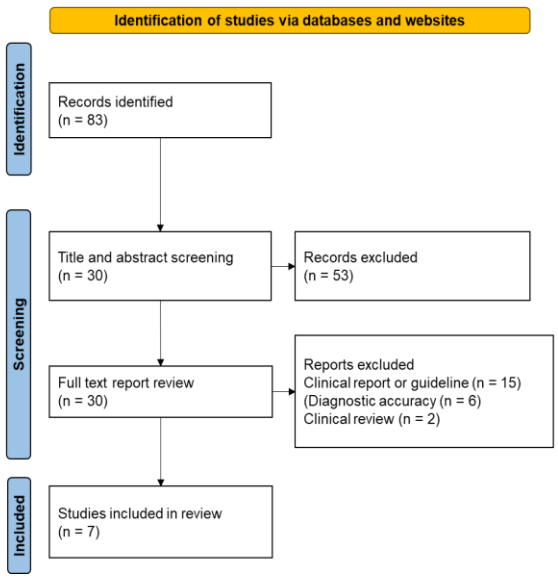
B.2. Methods

A search was run in June 2023 of the TRIP (Turning Research into Practice) Database, INHATA (International Network of Agencies for Health Technology Assessment) database and international HTA agencies' websites using the site command in Google. MEDLINE and Embase databases were searched with a precise filter for clinical practice guidelines. Key search terms combined terms for the target technology and published HTA reports. Records from each database were stored in a Word document. Results were screened for relevance on their titles. Clinical studies investigating clinical efficacy and accuracy of the multi-indication technology and clinical guidelines were excluded. To limit the results to a manageable size and maintain consistency with the first review, we focused the searches for this review on circulating tumour DNA (ctDNA) tests and polygenic risk scores (PRS). We included all reports for these technologies regardless of whether they evaluated multiple indications to develop a more complete picture of how HTA agencies approached issues around multi-indication technologies. Full-text review was carried out by 1 reviewer. The literature review summary was completed by 1 reviewer.

B.3. Results

The database searches returned 83 hits. After title and abstract screening, 53 records were excluded and 30 studies were selected for full-text review. Following full-text review, a total of 7 published HTA reports were included. These included 6 HTA publications for ctDNA (also termed liquid biopsy) and 1 report on PRS. Figure B1 presents the search flow diagram.

Figure B1 PRISMA diagram of the HTA reports review search and screening results



B.3.1. ctDNA tests

The review included 6 publications by HTA agencies on ctDNA tests. Two were published by NICE, 3 by the Canadian Agency for Drugs and Technologies in Health (CADTH) and 1 by Ontario Health. These are summarised in this section.

B.3.1.1 NICE publications

NICE has published 2 medtech innovation briefings (MIBs) on ctDNA tests. MIBs are documents to support NHS commissioners and staff when considering using new medical devices or diagnostic technologies. The information provided in MIBs includes a description of the technology, a review of the relevant evidence, how the technology is used and the position in the treatment pathway. MIBs do not include an evaluation of cost effectiveness of the technology.

Plasma EGFR mutation tests (2018)

MIB on plasma epidermal growth factor receptor (EGFR) mutation tests for adults with locally advanced or metastatic non-small-cell lung cancer (NSCLC)

Technology: Plasma EGFR mutation ctDNA tests do not need a biopsy to be taken and are a less-invasive alternative to tissue EGFR mutation tests. There are 7 technologies for plasma EGFR mutation tests described in the MIB.

Indication: The ctDNA test would be used as an alternative to tumour tissue EGFR testing, or before tumour testing, to inform decisions about prescribing EGFR tyrosine kinase inhibitors (TKIs).

Key uncertainties in the evidence: The evidence summarised in the MIB is from 7 non-UK-based prospective studies with 2,106 adults. The evidence demonstrates that the diagnostic accuracy of plasma EGFR mutation testing using ctDNA tests has a similar specificity, but lower sensitivity, compared with tissue EGFR mutation testing in adults with locally advanced or metastatic NSCLC. The evidence and the technology for identifying EGFR tyrosine kinase mutations are rapidly evolving and there is no established gold-standard test against which to evaluate them.

Cost and budget impact: The MIB provides a range for the cost of testing as between £138.05 to £230.22 per unit. The resource impact is comparable to standard care, but plasma EGFR mutation testing could be cost saving if it led to fewer tissue biopsies. There is no published evidence on the impact of adopting plasma EGFR mutation tests for people with NSCLC on the use of healthcare resources.

Signatera MRD assay (2022)

MIB on Signatera for detecting molecular residual disease (MRD) from solid tumour cancers

Technology: The Signatera test is an MRD assay that measures ctDNA in people with solid tumours, including colorectal, breast, bladder, renal and lung cancer.

Indication: Signatera is indicated for use in people with solid tumour cancers in addition to standard care. Breast, prostate, lung and bowel cancer make up

more than half of new cases. There are 2 testing settings depending on which treatment the person has had:

- Adjuvant setting: Signatera used in secondary care within 6 months after surgery to evaluate the need for adjuvant therapy. Follow-up testing may be done to increase sensitivity of detecting MRD.
- Surveillance setting: Signatera can also be used after the initial 6-month postoperative period to detect recurrence and to monitor treatment response for up to 5 years.

Key uncertainties in the evidence: The MIB reports that there is a large evidence base for Signatera including many full-text papers and abstracts. All ctDNA analysis in the evidence base was retrospective and healthcare professionals and patients were blinded to test results. There is therefore no evidence on using Signatera in clinical decision making or treatment choice. Therefore, there are gaps in the evidence base for prospective studies comparing the concurrent use of Signatera with standard care tests and imaging in larger sample sizes. The MIB recommends that future studies should also include randomised trials evaluating the use of Signatera in NHS clinical practice, including its effect on treatment decisions, outcomes, and resource use.

Cost and budget impact: It was concluded that there are budget impact considerations related to selecting the right population for adjuvant therapy which could have cost savings. The MIB suggests potential implementation of the technology could result in increased healthcare system burden if the frequency of imaging was increased to confirm ctDNA-positive results. The MIB concluded that detailed economic analysis in a NICE evaluation is needed across different tumour types and clinical settings to demonstrate any cost savings as described.

B.3.1.2. CADTH publications

CADTH Issues in Emerging Health Technologies (2019)

The CADTH Overview of Liquid Biopsy for Screening and Early Detection of Cancer (Cowling and Loshak 2019)

This CADTH Horizon Scanning bulletin provides an overview of liquid biopsy and available evidence. The bulletins are not systematic reviews and do not provide a critical appraisal of the findings. The reports are not intended to provide recommendations for ctDNA testing.

Methods: A search was done to identify relevant published literature between 1 January 2017 and 9 July 2019. Studies were considered for inclusion if the review included a liquid biopsy that could be used for screening or for the early detection of cancer.

Key findings: The search identified 12 liquid biopsies in development for oncology. While some of the tests focus on specific cancers, other tests are in development to be used as screening tools for multiple solid tumour cancers. The CELLSEARCH CTC test kit is the only test approved by Health Canada. It has also received US Food and Drug Administration (FDA) clearance to be used for monitoring metastatic breast, colorectal or prostate cancer. CELLSEARCH fulfils the requirements for CE marking in the European Union. The CADTH report described that the full utility of ctDNA technology for screening is not yet realised and many of the identified tests are to be used in addition to diagnostic procedures or as companion diagnostics to aid decision making.

Key uncertainties in the evidence: The report concluded that ctDNA testing for screening purposes is in its infancy. Liquid biopsy is increasingly adopted and explored for clinical use, however, analytic and clinical validation addressing the biopsy analytes are needed. A limitation of the assays is that different ctDNA assays vary in performance and have a different threshold for detection. This reduces the comparability of ctDNA tests. Beyond testing validity, translation of ctDNA tests into clinical practice is a key concern. To establish clinical utility for decision making, evaluation of tests in prospective clinical trials or retrospective analysis of collected samples will be required.

Cost effectiveness: The CADTH Horizon Scanning bulletin concluded that as tests continue to develop, a key consideration will be the cost effectiveness of new liquid biopsy technologies and whether they offer better diagnostic outcomes and cost savings compared with standard of care. There are a limited number of studies that have addressed the value of ctDNA tests for various cancer types. It is difficult to determine if ctDNA testing is cost effective as a screening or diagnostic tool, especially because diagnostic capabilities are still being researched. There are multiple considerations including staging, treatment course, therapeutic options and prognosis that differ across cancer types. It was concluded that all these factors contribute to the limited ability to apply a broad evaluation of cost effectiveness at that time.

CADTH Rapid Response Report 2020

The CADTH Rapid Response Report on Circulating Tumour DNA Testing for the Identification of Genetic Mutations: Diagnostic Test Accuracy and Clinical Utility (Hill et al. 2020)

This rapid response report is a reference list intended to help Canadian healthcare decision makers and healthcare professionals make well-informed decisions to improve the quality of healthcare services. The rapid report review outlined 3 research questions to identify test accuracy and clinical utility of, and cost-effectiveness literature on, ctDNA testing for the identification of genetic mutations. The research questions underpinning the search strategy were:

- What is the diagnostic test accuracy of ctDNA testing for the identification of genetic mutations?
- What is the clinical utility of ctDNA testing for the identification of genetic mutations?
- What is the cost effectiveness of ctDNA testing for the identification of genetic mutations?

Search strategy: A literature search was done on key resources including MEDLINE via Ovid, the Cochrane Library, the University of York Centre for Reviews and Dissemination (CRD) databases, and the websites of Canadian

and major international health technology agencies. Filters were applied to limit retrieval to HTAs, systematic reviews, meta-analyses, network meta-analyses, any types of clinical trials or observational studies, economic studies, and diagnostic test accuracy studies.

Key findings: The rapid response report did not identify any HTA reports. Four systematic reviews, 3 randomised controlled trials and 44 non-randomised studies were identified investigating the diagnostic test accuracy and clinical utility of ctDNA testing for the identification of genetic mutations. The report lists the references that were identified but does not go any further to synthesise the evidence in a narrative review or meta-analysis. No economic evaluation studies were identified in the search.

CADTH Horizon scan (2022)

The CADTH Horizon Scan on Emerging MultiCancer Early Detection Technologies

This CADTH Horizon Scan report provided a summary of information for new and emerging health technologies identified through the Horizon Scanning Service. The report summarised the available information on ctDNA tests for multi-cancer early detection tests for cancer screening. The report focused on the Galleri and CancerSEEK tests, both of which are further along the development pipeline and are being assessed in clinical studies.

Key findings: The Galleri and CancerSEEK tests were not approved for clinical use in Canada or the US, but both received the FDA Breakthrough Device Designation in 2019. This signals an expedited process for authorisation if either manufacturer submits an application for reimbursement. No economic evaluations were included. The report provided a summary of some of the published evidence demonstrating diagnostic accuracy and clinical utility for the Galleri and CancerSEEK tests. It did not provide a review of the breadth of ongoing research on these 2 tests.

Key uncertainties in the evidence: The report concluded that research for test sensitivity and specificity is in early stages and estimates of positive predictive

values may not be reliable. It is important to be cautious because a low positive predictive value may lead to increased follow-up tests. Moreover, there may not be enough genetic material to be detected by the ctDNA test in the early stages of different cancer types. For complex technologies such as the multi-cancer detection tests, that examine the presence of multiple cancer types, which are heterogenous diseases, a detailed evaluation is needed. This is likely to be a complex analysis of the benefits and harms for those multiple conditions.

The report highlights 1 approach to assess ctDNA testing. This is a quantitative framework model that incorporates the numbers of cancers detected with the rates of follow-up testing and impact on clinical outcomes and mortality, given the assumption early detection will lead to improved clinical outcomes.

The report suggested multi-cancer early detection technologies function differently compared with the traditional model of screening for a single cancer with a single test. If the tests are implemented, they would require a substantially different approach to screening because many healthcare professionals would be required to review the results. The report concluded that there are uncertainties about whether healthcare systems would cope with the need for additional tests such as biopsies and imaging for confirmatory diagnosis.

Ontario Health (2020)

The Ontario Health Technology Assessment Series report on Cell-Free ctDNA Blood Testing to Detect EGFR T790M Mutation in People With Advanced NSCLC

This Ontario Health Technology Assessment Series report is the only identified HTA agency report that evaluated cost effectiveness. The report evaluated the diagnostic accuracy and clinical utility, and the cost effectiveness and budget impact of publicly funding cell-free ctDNA blood testing to detect the EGFR T790M mutation in people with advanced NSCLC in Ontario.

Clinical effectiveness: The literature search done by Ontario Health identified 12 studies that met the inclusion criteria to address the research question. This

was diagnostic accuracy and clinical utility of cell-free ctDNA blood testing as a triage test compared with tissue biopsy to detect the EGFR T790M mutation in people with NSCLC. The pooled sensitivity and specificity of liquid biopsy to detect EGFR T790M in people with NSCLC was 68% (95% credible interval [CrI], 46% to 88%) and 86% (95% CrI, 62% to 99%; GRADE: Moderate). The positive predictive value and negative predictive value were 89% and 61%, respectively.

Uncertainties in the evidence: The report highlighted there are limitations and inconsistencies in the literature about the diagnostic accuracy of ctDNA tests, affecting the sensitivity and specificity of liquid biopsy. Future research should standardise the minimum biological threshold that will guide treatment decisions in clinical practice.

Cost effectiveness: The literature search to assess the cost effectiveness of cell-free ctDNA blood testing alone or in combination with tissue biopsy compared with alternative testing strategies to detect the EGFR T790M mutation in people with advanced NSCLC identified only 1 study. This study was deemed partially applicable to the research question because it did not consider the cost effectiveness of liquid biopsy compared with tissue biopsy alone. To overcome this limitation and gap in the evidence, Ontario Health did a primary economic evaluation.

Primary economic evaluation: What is the cost effectiveness of ctDNA blood testing as a triage test or alone, compared with tissue biopsy for the detection of the EGFR T790M mutation in people with advanced NSCLC from the perspective of the Ontario Ministry of Health?

Table B.2 Summary of primary economic evaluation

Model type and health states	• Decision tree combined with a cohort health state transition (Markov) model. The decision tree was used to model the mutation testing and initial treatment decision. • The Markov model was used to capture disease progression, survival and treatment modifications over time. • Over time, people could do any of the following: ○ continue to have treatment

	○ finish treatment and have maintenance therapy, if applicable ○ progress and have additional active treatment ○ progress and move to best supportive care ○ die.
Population	• Cohort of mostly males (about 55%) aged 64 years or older. This cohort was based on people included in a Canadian multicentre validation study of liquid biopsy for EGFR T790M mutation testing.
Intervention	• Liquid biopsy alone.
Comparator	• Tissue biopsy alone (standard of care). • Liquid biopsy as a triage test (followed by tissue biopsy if the result is negative).
Outcomes	• Incremental costs (including only testing and testing-related adverse event costs). • Incremental number of tissue biopsies avoided. • Incremental number of correct treatment decisions (such as when people positive for EGFR T790M have treatment with the third-generation EGFR-TKI, osimertinib, and people without EGFR T790M people have treatment with chemotherapy). • Incremental cost per tissue biopsy avoided. • Incremental cost per additional correct treatment decision.
Perspective	• The Ontario Ministry of Health.
Cycle length	• 3 weeks corresponding to approximately 1 round of chemotherapy.
Time horizon and discounting	• The time from disease progression after first-line (first- or second-generation) EGFR-TKI therapy to the treatment decision under various testing strategies. • In long-term analyses, a 10-year (life-long) time horizon. • Annual discount rate of 1.5% to both costs and quality-adjusted life years (QALYs). • Sensitivity analyses to look at a range of discount rates (0% to 5%).
Key model assumptions	• After each progression event, 50% of people have an additional line of treatment and 50% have best supportive care. • People have only 1 treatment at a time (for example, no combination chemotherapy and EGFR-TKI). • Pneumothorax is the only adverse event associated with tissue biopsy that substantially affects resource use and quality of life. • One-time costs associated with treatment-related adverse events are applied during the first cycle of each treatment. • Ongoing disutility are applied for treatment-related adverse events.
Mortality	• Survival and progression estimates were taken from the published literature and from the clinical review.
Costs	• EGFR T790M testing. • Drug acquisition (purchase), administration and monitoring. • Adverse events related to tissue biopsy or treatment.

	• General and end-of-life care.
Utility inputs	• Utility inputs were derived from the literature for people having second-line treatment for NSCLC.
Results: Long-term (lifetime) base case	• Liquid biopsy alone: CAD$122,938 per QALY. • Liquid biopsy as triage test: CAD$175,502 per QALY.
Results: Probabilistic sensitivity analysis	• At willingness-to-pay values below CAD$125,000 per QALY, liquid biopsy had the highest probability of being cost effective (0.65). At willingness-to-pay values above CAD$200,000 per QALY, liquid biopsy as a triage test had the highest probability of being cost effective (0.82).

Key uncertainties in the evidence: Only 2 laboratories in Ontario offer liquid biopsy testing for the mutation. For the implementation of the ctDNA test, other testing locations may be required. It was estimated that if funded as a triage test, the budget impact for liquid biopsy would cost the public payer an additional $60,000 to $3 million yearly. The costs are attributed to costs of treatment, adverse events and end-of-life care. The budget impact is dependent on factors such as implementation and the need for capital investment. The report highlights that many long-term costs attributed to treatment and care would be funded through existing drug programmes, the Ontario Health Insurance Plan and hospitals.

B.3.2. PRS

For PRS, 1 published HTA report was identified. This was published by Australian Genomics (Polygenic score incubator project published by Australian Genomics [2022]).

The project report published by Australian Genomics provides recommendations to the Australian Government Medical Research Futures Fund (MRFF). The project authors sought advice from national and international experts and did a qualitative interview survey (n=13) and a workshop with national experts (n=31). This was to identify priorities to advance PRS (from now, PGS) research that will inform future implementation and drive clinical translation of PGS in the Australian healthcare system. Based on national consultation comments and expert input, the report makes recommendations for 3 streams of

multidisciplinary PGS research in Australia: stream 1 to inform PGS assay and test development, stream 2 to inform evaluation and implementation and stream 3 for education, workforce and understanding. Stream 2 is of relevance to the HTA Lab project to identify approaches for the evaluation of multi-indication technologies. The key points from stream 2 are:

- development of PGS clinical tools or integration of PGS into existing risk tools
- implementation of PGS clinical tools in practice: population-level implementation studies, for example, into existing screening programmes
- frameworks tailored to help evaluation of validity and utility of PGS
- evaluation at every point along the PGS pipeline from laboratory to patient health outcomes
- health economic evaluation at every stage, from informing PGS implementation protocols (for example, using discrete choice experiments) to cost–benefit analysis to inform governments or healthcare systems and identify impact on budgets
- intervention and behaviour change studies, including long-term health outcome or behavioural follow-up studies
- health technology assessment.

Uncertainties in the evidence: The report concluded that the use of PGS in clinical practice is still in development and there are substantive research gaps for implementation into healthcare in a responsive, ethical, and cost-effective manner. There is no gold standard for analysing the analytic validity of PGS and key evidence gaps include a test evaluation framework with consistent methodology. A determination of the appropriate regulatory oversight is required to demonstrate the clinical utility of PGS.

Frameworks to understand the clinical utility of genetic tests have been proposed. This includes a framework based on the ACCE model (Analytic validity, Clinical validity, Clinical utility, Ethical, legal, and social implications) developed by the US Centres for Disease Control and Prevention (CDC) Office of Public Health Genomics. The ACCE model has been used extensively and

may be appropriate for evaluating PGS. However, a broader framework that combines the ACCE model with an HTA approach is needed.