Cancer Diagnostics and Target Product Profiles

Annexes

CANCER RESEARCH UK

Together we are beating cancer

About these annexes

These annexes and the research conducted to inform them has been undertaken by researchers at RAND Europe and the Office of Health Economics.

RAND Europe is a not-for-profit policy researchorganisation that helps to improve policy and decision making through research and analysis.

The Office of Health Economics supports better health care policies by providing insightful economic and statistical analyses of critical issues.

About Cancer Research UK

We're the world's leading cancer charity, dedicated to saving and improving lives through research. We fund research into the prevention, detection and treatment of more than 200 types of cancer through over 4,000 scientists', doctors' and nurses' work. In the last 50 years, we've helped double cancer survival in the UK, and our research has contributed to around half of the world's essential cancer drugs. Our vision is a world where everybody lives longer and better lives, free from the fear of cancer.

Our values

Our values help guide our behaviour and culture in an ever-changing world, building on the best of what we do today and what we aspire to be in the future. They unite and inspire us to achieve our ambitious plans and mission of beating cancer together.

Our values are:

Bold
Act with ambition, courage and determination

Human
Act to have a positive impact on people

Credible
Act with rigour and professionalism

Together
Act inclusively and collaboratively

Reference

These annexes should be referred to as follows: 'Cabling, M.L., Dawney, J., Napier, M., Marciniak-Nuqui, Z., Olumogba, F., Kessler, L., Cole, A., Steuten, L., Marjanovic, S. (2024) Cancer Diagnostics and Target Product Profiles - Annexes.

Authors

Mark L Cabling, PhD - Research Analyst, Health and Wellbeing, RAND Europe

Jessica Dawney - Research Analyst, Health and Wellbeing, RAND Europe

Matthew Napier - Economist, Office of Health Economics

Zuzanna Marciniak-Nuqui, PhD - Research Analyst, Health and Wellbeing, RAND Europe

Fifi Olumogba - Junior Analyst, Health and Wellbeing, RAND Europe

Larry Kessler, Sc.D. - Professor, Department of Health Systems and Population Health (HSPOP), School of Public Health and Director of Regulatory Oversight Centre for Dialysis Innovation, University of Washington

Amanda Cole, PhD - Associate Director, Office of Health Economics

Lotte Steuten, PhD - Deputy Chief Executive, Office of Health Economics

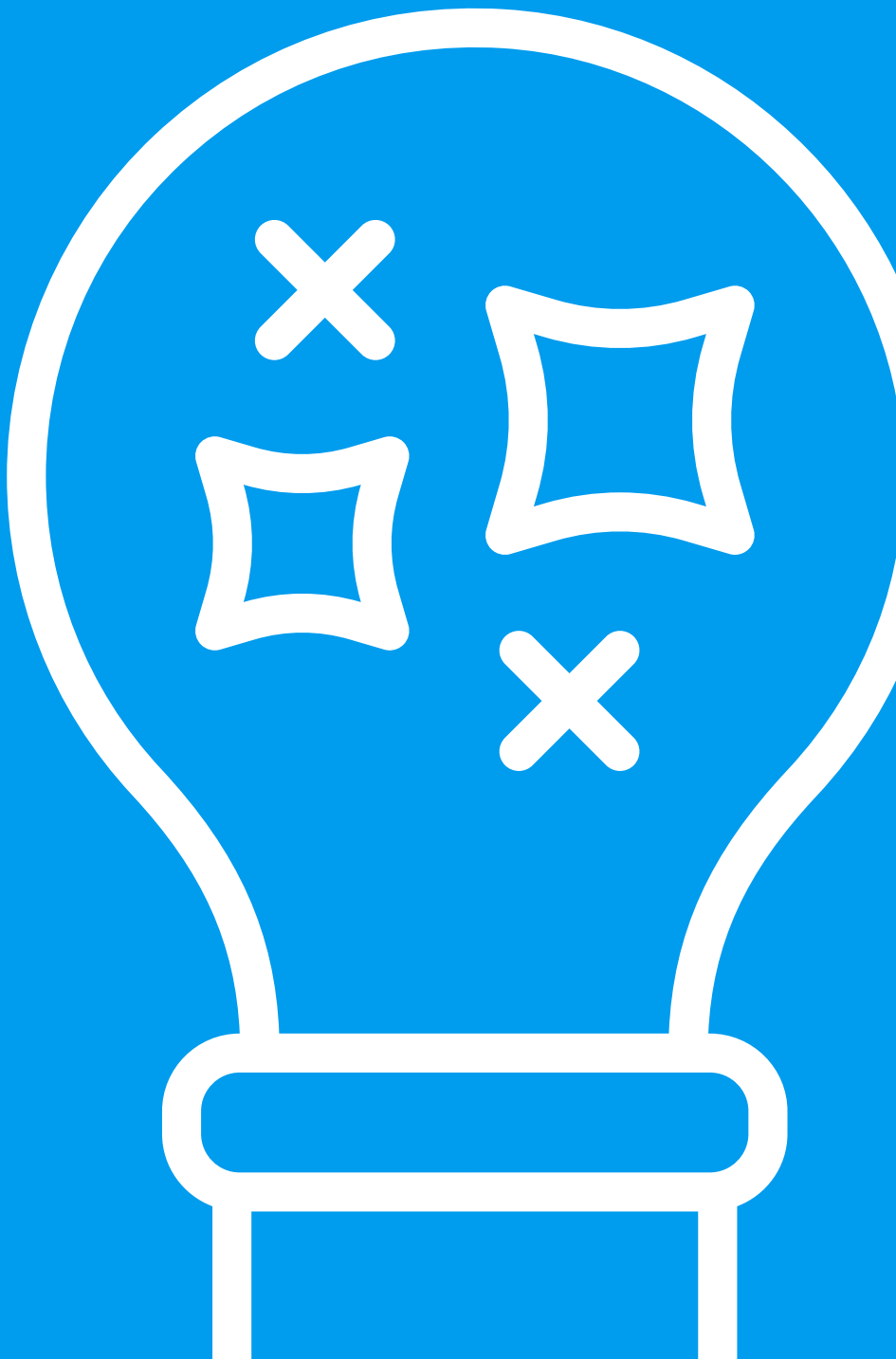
Sonja Marjanovic, PhD* - Director (Healthcare Innovation, Industry and Policy), RAND Europe

*(*corresponding author)*

Contents

Annex A: Scoping Insights Document

Authors: Mark L Cabling, Jessica Dawney, Matthew Napier, Zuzanna Marciniak-Nuqui, Fifi Olumogba, Larry Kessler, Amanda Cole, Lotte Steuten, Sonja Marjanovic



1. Introduction

Annex A is the first of eight annexes complementing the main Cancer Research UK-funded project's final report: 'Advancing the development and use of diagnostic target product profiles for cancer.' The not-for-profit research institute RAND Europe led the project in collaboration with the Office of Health Economics. The project has benefited from ongoing support and advice from Professor Larry Kessler (University of Washington), a key consultant on the work. This document provides the detailed findings and analysis of the scoping desk research and preliminary stakeholder consultation to which the final report refers; thus, Annex A, like all the other annexes, is meant to accompany the final report and is not meant to be read as a standalone document.

1.1. Scoping insights: An overview of our approach

An overview of the scoping phase approach

The scoping work package aimed to lay the foundation for broader stakeholder consultation on a diagnostic test TPP approach for cancer undertaken in later work packages. The scoping focused on examining key insights on TPP development processes (i.e. approaches and methods) for diagnostic tests, desired categories of features in a diagnostic test TPP, and key insights on enablers and barriers to TPP development and use.

We implemented our scoping phase through a combination of desk research and initial stakeholder consultation (to avoid 'reinventing the wheel' given the work already undertaken in this space) and in line with this initial scene-setting work package's limited, brief and exploratory scope.

We used the overall insights gained from the scoping of TPP features and TPP development processes to design and organise discussions at stakeholder workshops undertaken in later phases of the project.

1.1.1. Diagnostic TPP features: Our approach to the scoping task

To understand the types of features already considered in diagnostic TPP efforts, we first explored the systematic review produced by Cocco et al. (2020)[1] in depth. We went beyond an analysis of this key systematic review paper's content to consider the diversity of features informing the authors' core feature typologies. We did this by drawing on the supplementary material published in the article and conducting an analysis of it. This process allowed us to critically reflect on the nature of the TPP features used internationally (as informed by the source TPPs reflected in the systematic review and drawing mainly from the infectious diseases field where diagnostic TPPs are prominent). In the following sections, we highlight key observations from this analysis (see Section 2), which provided useful learning for subsequent phases of this project and informed our approach to designing enquires for the stakeholder consultation workshops.

The core research team also consulted with Professor Larry Kessler on the nature of features that informed the Cocco et al. (2020) review[1] and specifically on terminology (i.e. similar features being termed differently in different TPPs) and explanations of features. We also took a sample of TPPs (n=8) not covered in the Cocco et al. (2020) review to examine how they organised features into categories and understand how this maps onto the conceptualised developed by Cocco et al. (2020).[2-9] We selected the sample of eight because they included explanatory text and information about their TPP development process and methods (not just a table of features). We sampled and analysed a subset of these TPPs (n=3) from each of the main purveyors of TPPs in more depth to consider the types of features included in the diverse categories.[3,4,6] We did this to consider how well the features and categories map onto the conceptualisation presented by Cocco et al. (2020),[1] and for a more detailed understanding of how TPPs are presented, including optimal and minimal requirements. We detail our selection process for this TPP sample in Section 1.1.2 (below) in line with the task's scoping and exploratory nature.

We used the findings to begin understanding key issues with how TPPs are currently presented and identify opportunities for improvement and implications for future practice potentially relevant to the current project (as we discuss in Section 2.1 and Annex B). This process helped us establish a list of features to consider in workshop discussions conducted later in this project. As part of this, we sought to resolve some duplications that characterise the current TPP landscape and relate to the diverse terminology often used to describe the same or similar features across different TPPs. Based on desk research and consultation with Professor Larry Kessler, we also established some 'working explanations' for features, which we felt was important to enable future meaningful discussion and overcome the lack of readily accessible glossaries for terms used for describing features in current TPPs and literature.

Finally, concerning the scoping of features used in TPPs, we reached out to some members of the project advisory group with particular experience in oncology diagnosis and innovation and Health Technology Assessment (HTA) and regulation for their thoughts on the feature types they consider more likely to apply universally in an oncology diagnostic context and those likely to be more context-dependent (i.e. dependent on test type, clinical use case, clinical use setting or tumour site). This process involved asking the advisors which features in our streamlined list/categorisation are likely to be (a) always essential, (b) sometimes essential (e.g. depending on test type, clinical use case, clinical use setting or tumour site), (c) always desired but never essential, (d) sometimes desired but never essential, and (e) not needed, with an option to comment if they did not understand what a feature means.

This early exercise aimed to explore the diversity of views (or whether there was a strong consensus) at a high level rather than to formally prioritise features. Our working assumption was that there would be considerable diversity, and hence that any effort to overview the feature types used in diagnostic TPPs should err on the inclusive side so that those developing TPPs in future bespoke efforts can use the insights (i.e. a relatively comprehensive list of features) to select features relevant to their unique TPP development effort).

This exercise also helped inform the design and organisation of our future workshop discussions. Our early consultation with a small sample of expert advisors (leading experts in the field) laid the foundations for further work packages. We obtained six responses in total.

Colleagues at the Office of Health Economics (OHE) also conducted a scoping review of the literature on early economic modelling (EEM), and this document summarises the critical learning points.

1.1.2. TPP development processes: our approach to the scoping task

We started by synthesising insights from the Cocco et al. (2020) review[1] on the approaches and methods used to develop diagnostic TPPs. This synthesis covered the key stages, methodological options in each stage, stakeholders involved, governance and oversight (and commissioning), and the TPP development process. We examined the systematic review paper's content and some of the underlying supplementary material as part of this process.

We also identified and prioritised a sample of TPPs (n=8) because they included explanatory text and information about their TPP development process and methods (not just a table of features) for more in-depth analysis as part of our scoping work.[2–9] We did this to explore how the processes detailed in the sample relate to the process typology presented in the Cocco et al. (2020) review,[1] and dig deeper into insights on TPP development approaches and processes by analysing a sample of individual TPPs. We sought to explore the nuance and detail of diagnostic TPP development in a way that could help inform the design of further project work packages and further stakeholder consultation.

We achieved the above through an exploratory approach involving the following key steps:

1. Search the major TPP developers' database for TPPs, e.g. the World Health Organisation (WHO), PATH (formerly known as the Program for Appropriate Technology in Health), FIND, and the United Nations Children's Fund (UNICEF), and collate a list of TPPs available. While PATH, FIND and UNICEF tend to provide information only on TPPs they support or draft, resources like the WHO database provide a TPP directory

that enables an Excel export providing a long list of various TPPs across diseases and developers.

2. Narrow down the longlist of TPPs using the following criteria, ensuring the TPPs are:
    a. Focused on diagnostics (in line with this work's scope, i.e. we were not focused on TPPs for products).
    b. Published from 2017 onward for a recent sample of relevant TPPs (published in the last five years; this was also in line with a desire to identify TPPs that were unlikely to be included in the Cocco et al. review[1] given the timeline that review covered).
    c. Include substantial detail and information about the TPP development process and methods (note that we excluded some TPPs with a background paragraph detailing a general or generic process as they did not provide useful specificity or detail).
    d. Suitable for analysis within this task's scope and resources (we aimed for 6–10 TPPs for in-depth analysis): i.e. pursue selective sampling in line with the scoping work package's relatively bounded and focused nature and aiming to use the sample to check for compatibility and relationships with insights from the Cocco et al. (2020) review,[1] in a way that could help inform the design of stakeholder workshops undertaken later in the project.
3. We then conducted a critical analysis of the documents. This analysis involved drawing out key learning points and observations relevant to further work phases, including commonly shared aspects of TPP development processes, diversity in the methods and approaches used, and types of oversight/governance and stakeholder involvement. To aid this TPP analysis, we developed a coding sheet that identified the following in each TPP: (a) the scoping process (identifying the reason for developing a TPP, the steps taken and methods used in this stage and the stakeholders involved), (b) the TPP drafting process (identifying the steps taken to draft TPPs, the methods used and the stakeholders involved at this stage), and (c) the consensus process (identifying the approach to exploring and reaching consensus and the methods used, as well as stakeholders involved in this stage).

We used these insights to shape discussions on methods and approaches for TPP development for oncology in further workshops, including discussions on trade-offs between different methodological options and approaches and broader considerations when designing a TPP development process, such as when to include specific stakeholders and how. We also used insights from this analysis to identify individuals we thought would be relevant to interview in the scoping interviews.

1.1.3. Scoping interviews

As part of the scoping task, we also conducted a small number of semi-structured exploratory interviews with a mix of individuals spanning experts involved in prior diagnostic TPP development efforts (in the international infectious diseases space, industry experts and consultants and a CRUK representative). Depending on who we were interviewing, the interviews served to explore and learn from prior experiences with developing diagnostic TPPs or to understand the nature of a diagnostic TPP as a demand signalling document that stakeholders would find helpful. After gaining informed consent, we conducted all interviews online between May and June 2023 using MS Teams.

Individuals consulted (and named with their consent) included:

- Dr Joy Allen (in a personal capacity with experience in HTA and pharma and research).
- Dr Phillip Beer, CSO of Step Pharma, Chair of BIVDA's genomics working group.
- Dr David Boyle, Co-lead diagnostics programme, and Roger Peck, Associate Director of technology advancement, diagnostics at PATH.
- Professor Jon Deeks, Institute of Applied Health Research, University of Birmingham.
- Dr Christopher Hanna, Principal at Kattner-Thalmann Partners (who has been involved with WHO efforts).

- Ms Samantha Harrison, Head of Strategic Evidence & ICBP lead at Cancer Research UK.
- Ms Lucy Hattingh, Principal at LH Consulting (who has been involved with FIND efforts).

1.1.4. Early economic modelling (EEM) scoping task

We began by searching the literature on the economic evaluation of diagnostic tests to establish how the value generation for diagnostic technologies differs from other types of healthcare technologies and essential considerations when evaluating these technologies. We conducted searches in Google Scholar and identified relevant articles.

Following this, we synthesised key insights from Cocco et al. (2020, 2021)[1,10] to identify the EEM analytical approaches relevant to diagnostic technologies and how these fit within different phases of TPP development. Finally, we examined the NICE methods guide[11] to establish the routes a diagnostic technology may take through NICE and any key methodological or evidential considerations for diagnostics.

These insights informed how early economic analyses could fit into the development of a TPP. In turn, these helped define what characteristics the diagnostic must have to be evaluated favourably by the payer in terms of cost-effectiveness.

2. Key insights from the scoping exercise

2.1. Insights on TPP features

This part of the scoping task examined overarching categories of potential features in a TPP's scope and the specific features within these categories. The observations draw on insights from the Cocco et al. (2020) review[1] and from a sample of TPPs selected for analysis in the scoping phase,[2-9] alongside a recent article on a novel framework for evaluating diagnostic strategies for early cancer detection (CanTest framework).[12]

2.1.1. General analysis observations and their implications

1) Feature categories: How TPPs are organised

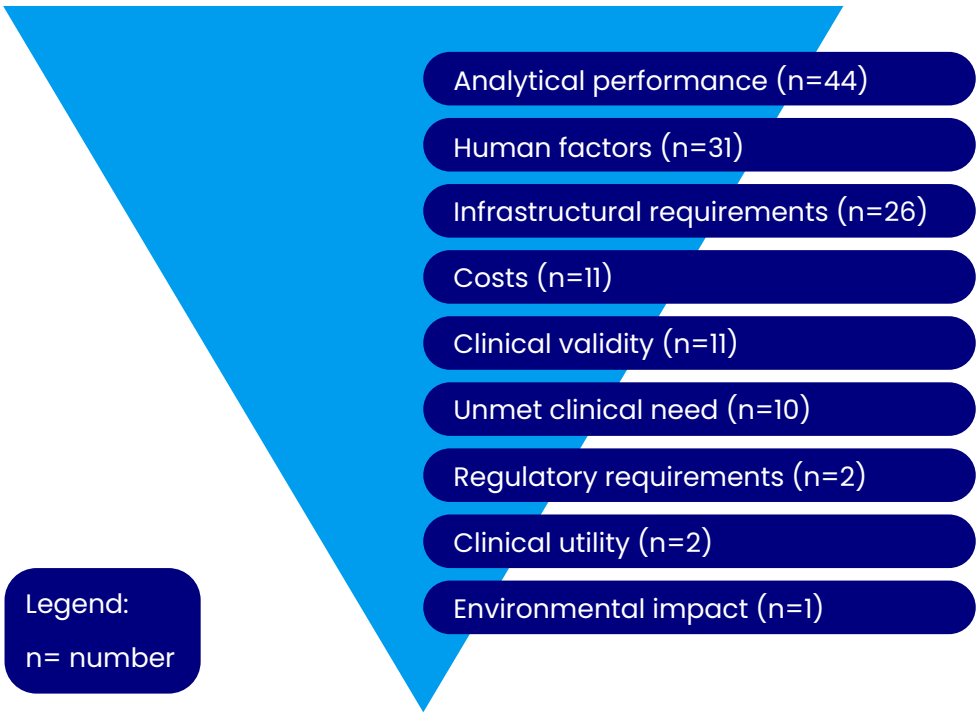
Observation: **While individual features presented in TPPs are grouped into conceptual categories, their organisation into categories differs across TPPs (i.e. the category structures TPPs used are different). This variety may partly relate to the nature of the diagnostic test for which a TPP exists or is being developed. However, it may also relate to differences in terminology and organisational practices among those commissioning and developing TPPs. We observed similarity in most higher-level concepts covered through the features included in different TPPs** (such as concepts related to a test's technical performance, cost considerations or human factors) even if different TPPs are not structured (i.e. organised) in the same way. To elaborate, we reviewed a sample of TPPs, finding that no overarching feature categories used the same terminology as that describing conceptual categories of TPP features in the Cocco et al. (2020) framework[1] for structuring test characteristics. This finding is not entirely surprising given that Cocco et al. (2020) arrived at their conceptualisation fresh, based on a substantial number (n=44) of diverse TPPs (see Figure 1 for the categories they used, and the number of features identified in each category across the TPPs they analysed).

Reassuringly, however, the categories in our sample were conceptually similar to and/or compatible with those in Cocco et al.'s framework (i.e. we did not find that the Cocco et al. framework omitted concepts and features that were prominent in the sample of TPPs we analysed). To illustrate, the terms 'scope,'[2] 'scope of the app,'[8] 'general scope,' and 'scope of the test'[7] all seem to relate to the wider category of features covered under 'unmet clinical need' in the Cocco et al. (2020) conceptualisation. Similarly, categories such as 'test performance,'[4] 'scope of toolkit components,'[5] and 'test performance characteristics'[7] are similar to the category of 'analytical performance' in Cocco et al.'s review. A few categories seemed specific to the diagnostic type in question, e.g. the 'clinical decision support algorithm' category in a TPP for electronic clinical-decision support algorithms incorporating point-of-care diagnostic tests in low-resource settings.[5]

Based on their systematic review, Cocco et al. (2020) list nine major categories of TPP features from their analysis of 44 TPPs,[1] as shown in Figure 1 below.

Implication: **There is merit in exploring the potential for a more standardised framework for organising and grouping TPP features in future efforts to develop diagnostic TPPs for use as demand signalling tools in oncology.** This is especially so given that there are no established practices for developing TPP features in an oncological context – unlike infectious diseases, where different organisations commissioning innovation and developing TPPs may have preferred and established approaches to TPP structure. Developing diagnostic TPPs for oncology would be a novel step (at least in the context of a demand signalling tool) and could benefit from building the foundations for a more standardised, streamlined approach to structuring and organising TPPs (with an awareness that some test types may still require additional, test-specific categories of features). While there may be established

Figure 1. Overarching categories of features listed in the Cocco et al. (2020) review and the number of features in each category[1]



diagnostic TPPs on industry-developed cancer test specifications, information about this is not publicly available. Besides, the importance of TPPs as demand-signalling tools that can respond to industry needs must be considered.

2) Feature diversity within and across categories

Observations: **An extensive range of features appear in different categories used in TPPs, and relatively few standard features apply across different TPPs (i.e. are likely to be important irrespective of test type, clinical use case and use setting).** There are some common features across different TPPs and many unique ones. For example, a relatively small number of features included in the Cocco et al. (2020) analysis[1] appear in half or more of the TPPs the authors analysed (see Table 1 below), suggesting there may only be a limited number of features considered vital to all TPPs.

However, the findings on relative frequency presented in Cocco et al. (2020)[1] may be an artefact of the nature of TPPs considered in the review, thus an underestimate. Our analysis found a range of features in individual TPPs that are conceptually very similar, if not identical, across individual TPPs. However, they are termed differently and hence counted as individual features regarding their frequency in the Cocco et al. (2020) review's supporting data. For example, features like 'price/cost of individual test' and 'cost per diagnosis' come under the 'costs' category or features like 'instrument-infrastructural requirement' and 'infrastructural requirement' under the 'infrastructural requirements 'category' in Cocco et al. (2020) review. These are conceptually similar features described differently in different TPPs. Hence, combining multiple similar or conceptually identical but different features into one feature type may make them appear more frequent. Cocco et al.'s findings on frequency may be an artefact of the diverse terminology used in the TPP landscape to some degree. Nonetheless, they provide a useful reference and starting point for understanding common TPP features. We provide further insights about this observation in Annexes B and G (the latter giving working explanations of individual features).

Some features in TPPs are unique to specific test types. For example, a 'Target Product Profile' for a mobile app to read rapid diagnostic tests to strengthen infectious disease surveillance'[8] listed some features specifically relevant for electronic devices,

such as features used to describe 'compatible mobile devices (smartphones and tablets)' or 'handling of intermittent connections'. Such features are unlikely to be relevant for tests that do not use mobile-device interfaces. It may be that in-vitro diagnostics are more represented in work to date. Although this project does not seek to establish new features for specific test types, we have enquired whether any key features matter for imaging-type tests during stakeholder workshops and general practitioner (GP) or pathology lab interviews and brought these into our analysis.

The number of features linked to any single TPP category can vary widely, as evident in the Cocco et al. (2020) review[1] and our sample TPPs. Of the nine categories Cocco et al. (2020) reported, 'Analytical Performance' had the highest number of features (n=44) reported across the TPPs analysed. In contrast, the 'Environmental Impact' category listed only one feature.[1] This result suggests that the 'Environmental Impact' category has either been less developed to date (i.e. is considered to a lesser extent in existing diagnostic TPPs) or that the features considered are more common, i.e. there is less diversity across the TPPs in the specific feature(s) used to understand and specify requirements related to 'Environmental Impact' (or a combination of both). This interpretation is supported by our review of sample TPPs, since no TPP mentioned 'Environmental Impact' either as a category or a feature.

The number of distinct features in a category speaks to the diversity of potential requirements of interest across different tests and/or – in the context of the Cocco et al. (2020) dataset – potentially also to the diversity of terminology used to describe any single feature across different TPPs. However, it does not indicate how developed a category is regarding how far requirements related to that category are addressed in different TPPs. For example, the 'Clinical Utility' and 'Regulatory Requirements' categories only have two features based on Cocco et al.'s review. However, the features within the 'Regulatory Requirement category' (i.e. 'regulatory requirements' and 'product registration path') were covered in 35% and 25% of TPPs reviewed by Cocco et al. (2020), respectively.[1] In contrast, Cocco et al. (2020) only identify two feature types within the Clinical Utility category ('intended outcome and linkage to care' and 'what is the risk of an inaccurate test result?'), with each rarely used in TPPs – only 2% of TPPs analysed by Cocco et al. (2020).[1]

Implications: **While there will likely be a core set of common features relevant to multiple TPPs** (e.g. diagnostic sensitivity and specificity), **some features will also be unique to different test types, clinical use cases, use settings and/or tumour sites. Any efforts to develop a standard framework for TPPs must be sensitive that there may be (a) some essential features across TPPs for different tests, (b) some desirable even if not essential features across TPPs for different tests, but also other features which may be (c) essential or (d) desirable only in TPPs for some test types and clinical use cases**. There will likely be a trade-off between what is ideal and what is feasible, and these decisions may need making on a case-by-case basis; each must consider what is feasible and necessary without compromising flexibility and innovation.

As noted earlier, we also conducted high-level sense-checking to see how different people approach features potentially relevant to TPP development. This process was not a quantitative exercise but sought to gauge perceptions and approaches to add value to the project. We outline the key insights across the six inputs we received below:

- The majority of respondents considered most existing features either *always or sometimes essential* or *always or sometimes desirable*, supporting the value of a comprehensive options list to build on in future efforts. This result also points to the scope for bespoke adaptation.
- One notable discrepancy was a respondent who felt that many of the broader systems-related features concerning human factors or infrastructure were not necessarily appropriate for TPPs and would be better addressed in other documents (e.g. manufacturer specifications). This illustrates that some people may have different views on the breadth of what TPPs should address regarding the interplay of technical performance and wider system features.
- Some respondents considered features unclear (though usually only one or two respondents), reinforcing our earlier point

about the need for future work to clarify feature definitions in bespoke efforts. This referred mainly to the following features (it is well outside the scope of this project to embark on definitional matters; however we did some desk research and consulted with members of the project advisory group regarding the definitions we used for these features—see Annex G for details): Target user; Target level of health system; Proof of concept; Strain specificity; Volume of sample specimen; Result; Reagent kit (nature, transport, storage and stability, supplies not included in kit); Control/ comparative reference method; Type of analysis; Precision/concordance; Quality control; Indeterminate test result; Generic sensitivity and specificity; Field performance; Precision/concordance; False recent ratio; Instructions for use; Patient identification capability; Multiuse platform; Costs per diagnosis; Market size, nature and segmentation; Competitive market; and Product registration path. This lack of clarity refers to a minority of features in the landscape we analysed and as covered by Cocco et al. (2020).[1]

Individual TPP development efforts will want to consider the granularity level associated with any single feature. Thus, some features in the lists below may merit splitting into more granular individual features in specific TPP development efforts (e.g. separating infrastructure features associated with physical equipment from those associated with supplies/consumables).

In addition, related to the plurality of terminology used to name features, **there is scope for clarifying what different features mean and potentially for streamlining the diversity of terminology used for the same/ similar feature types in future efforts to develop TPPs.** Any efforts to do so must be mindful that what falls within a feature's scope – in terms of its specifications – will likely have a mix of unique and diverse elements across TPPs for different test types. A broad scope may apply to some features, e.g. 'supplies needed' or 'quality control'.

Some feature categories listed in the Cocco et al. (2020)[1] review were easier to understand and define than others. For example, the indicators used for 'costs' seemed more self-explanatory than the diverse indicators used for 'analytical performance'.[1] To proceed with the project, we felt that it was necessary to formulate a way of explaining and understanding what different features mean (to the degree possible). To clarify our understanding of specific features, we combined desk-based searches with a consultation with Professor Larry Kessler (project consultant) wherever possible (considering the detail in the Cocco et al. [2020] review and supporting materials). We discussed proposed explanations for a feature with Professor Larry Kessler and have presented them in Annex G, focusing on specific categories from the Cocco et al. review. These descriptions are not an attempt to define features but an effort to arrive at 'working explanations' to use in further consultations in stakeholder workshops. Any attempt to formally define features would be a project in itself requiring additional expertise. However, future work could address this opportunity by developing clear and consistent overarching frameworks to guide future TPP development efforts.

It was brought to our attention during a recent workshop (not part of the scoping work) that The Clinical and Laboratory Standards Institute (CLSI) is working on a database of internationally accepted terminology for laboratory science and the healthcare industry more generally. However, it is not TPP or oncology-specific. We have consulted on features which are less straightforward to define, but many of the terms are not yet covered in the database. However, it suggests an area where further research could map this and other initiatives developing terminology to establish a standard-terminology reference set for work.

Table 1. Features appearing in half or more of the TPPs (n=44) that informed the Cocco et al. (2020) review[1] (adapted from source information and supporting documentation supplied in the review)

Category of features in the Cocco et al. (2020)[1] typology	Features that appear in ≥ 50% of the 44 TPPs that informed the Cocco et al. (2020) review
Unmet clinical need	• Intended use (98%) • Target level of health system (86%) • Target population (75%) • Target user (75%)
Analytical performance	• Sample type (95%) • Time to test result (86%) • Manual sample/specimen preparation (77%)
Clinical utility	• n/a
Clinical validity	• Diagnostic/testing sensitivity (70%) • Diagnostic/testing specificity (64%)
Human factors	• Training and education (75%)
Infrastructure requirements	• Storage conditions and shelf life (70%) • Temperature and humidity (61%) • Power requirements (52%) • Stability during transport (52%) • Waste disposal (50%)
Costs	• Price/cost of individual test (61%)
Regulatory requirements	• n/a (no feature in this category appeared in half or more of the TPPs in the Cocco et al. review)
Environmental impact	• n/a (no feature in this category appeared in half or more of the TPPs in the Cocco et al. review)

3) Feature attribution to specific conceptual categories

Observation: **How features are conceptualised, i.e. the overarching TPP category within which they are classified, can also vary across TPPs: there is no standard protocol specifying which feature should belong in which category of a TPP structure.** Therefore, it is unsurprising that the Cocco et al. (2020) conceptualisation for TPP feature categories includes some features that appear in more than one category in their framework's underlying dataset. For example, the feature 'result' appears in both the 'analytical performance' and 'human factors' categories in their framework. Some features also appear across all nine general categories reported by Cocco et al. (2020) (see Figure 2 below).

Figure 2. Features repeated across multiple categories in the Cocco et al. (2020) review[1]

Categories:	Features:
Analytical Performance and Human Factors	• Result
Analytical Performance and Clinical Validity	• (Generic) sensitivity • (Generic) specificity
Human Factors and Infrastructural Requirements	• Supplies needed • Service and support • Assay packaging • Materials used
Clinical Validity and Clinical Utility	• What is the risk of an inaccurate test result?
Analytical Performance and Infrastructural Requirements	• Reagent kit (transport, storage and stability, supplies not included in kit)
Human Factors and Infrastructural Requirements	• Biosafety requirements/Safety precautions (biosafety requirements)
Analytical Performance and Clinical Validity	• Precision/concordance

Implication: **Efforts to develop overarching/unifying frameworks for TPP development should consider where specific feature types best fit (i.e. classified under which overarching category) according to relevant stakeholders' 'mental models' and views, especially those developing or commissioning and paying for novel tests**. Having some consensus on these matters can help avoid duplication and unclear signals for future TPP development efforts in the oncology space. However, any decisions on feature attribution to a category should serve as guidance only and not be enforced, as there may be good reasons for 'departing from the norm' in specific TPP development projects.

4) Feature specifications

Observations: **Our review of the TPP sample we conducted as part of our scoping phase revealed significant diversity in whether TPPs list optimal *and* minimal requirements in specifications for TPP features and whether the reasoning is given.** For example, the TPP for 'Simplified Blood Culture to Enable Widespread Use in Resource-Limited Settings'[2] details optimal and minimal requirements for all 27 features that it lists, while only 46% of features in the TPP for a mobile app to read rapid diagnostic tests to strengthen infectious disease surveillance[8] have requirements for optimal and minimal requirements. However, the latter is an exception, with the rest of the reviewed TPPs listing optimal and minimal requirements for at least more than 60% of their features.

Implication: **When developing TPPs, stakeholders will likely benefit from upfront consideration about whether to specify optimal and minimal requirements and reasons for specifications for all features in a TPP relative to importance and feasibility.** For example, feasibility may be affected by the nature of the existing evidence base, the clarity of need for improving existing tests' features and gold standards, and any

uncertainty about which specifications are achievable within a field's current scientific and technological advances. **Balancing unambiguous specifications with adequate flexibility and innovation is critical.** According to a discussion with one of our advisors, it may also be more feasible to specify optimal and minimal requirements for some features than others, and qualitative information may be needed where quantitative information is tricker to access. Arriving at minimal and optimal requirements for some quantitative specifications may require going beyond desk research and stakeholder consultation to engage in modelling and comparisons to an existing gold standard.

Finally, those developing TPPs must consider whether they signal a short-, medium- or long-term need (as far as possible) and the timeframe over which the TPP will be considered valid (before it expires). A short-term need may be less incentivising to industry innovators and influence the attractiveness of a TPP as a demand signalling tool, even considering a TPP's dynamic nature as a living document.

5) Multi-component diagnostic tests (through the lens of an example involving digital technologies)

Observations: We also considered an example of a multi-component test TPP in our scoping work. We deliberately chose a test involving digital technology, as these test types are more novel and may not have featured as prominently in previous literature on diagnostic TPP development and features. Although it is beyond the current project's scope to examine complex digital diagnostics in detail, we wanted to explore any unique insights in this space and lessons related to multi-component tests. To this end, we examined an example TPP for a test based on an electronic clinical decision support algorithm (CDSA) that incorporates point-of-care diagnostic tests as part of a multi-component kit.[5] This was one of our sample TPPs. Recently, there has been growing interest in applying digital technologies to cancer diagnosis, particularly in the context of digital imaging and screening combined with clinical decision support software and/or artificial intelligence (AI) and machine-learning capacities. Therefore, we felt it could be helpful to explore key learning transferrable to future efforts to develop TPPs for tests involving digital and AI technologies (within the limits of the current project's scope).

This particular TPP defined a toolkit comprising a clinical decision support algorithm and point-of-care tests to support evidence-based clinical decisions by capturing patient, clinical and contextual data and diagnostic test results to arrive at diagnosis and patient care needs recommendations. The algorithm integrates the diagnostic test results with the other relevant information, all embedded in an app (Pelle et al., 2020).

The group developing the TPP used the FIND and WHO framework of features to consider as a starting point. However, the final TPP had a bespoke structure and reflected several unique considerations:

- The first category of features covered **characteristics describing the general scope of the test**, which included features typical of many TPPs (classified in the Cocco et al. review under 'intended use'), including specifications regarding the test's intended use, target population, setting and target end user.
- The second category covered characteristics describing the test kit's **core components**, which comprised an algorithm, associated point-of-care testing tool, compatible devices for the app to function on and associated operating systems in this case. This category of features focuses primarily on component characteristics essential for an accurate and clinically useful test. These characteristics broadly correlated with general features in the Cocco et al. (2020) framework relating to analytical performance (accurately measuring/capturing the required patient, clinical, diagnostic and contextual data), clinical validity (measuring/capturing and conveying the appropriate information) and clinical utility (informing the correct clinical decision-making for desired patient care outcomes), as well as regulatory requirements. Given the test in question, this concerned features such as algorithm access format/design (e.g. access via an app), content informing the algorithm (e.g. underlying data-input requirements to supply the algorithm with credible, clinically valid information), information related to treatment recommendations

(e.g. compatibility with national guidelines to provide appropriate clinical validity and utility), information on compatible/additional associated diagnostic tools to be used/prompted by the app (e.g. point-of-care tests to support clinical validity), regulatory considerations for diagnostic tools, information on compatible device requirements (e.g. tablets, phones and laptops) and compatible operating systems.

- The TPP then presents different test kit components as an overarching category and discusses diverse features for each category. Thus, the TPP includes separate categories for the **clinical decision support algorithm element, point-of-care tool element, app component/device element and data component/element**. Each includes diverse features, many of which broadly correlate with categories in the Cocco et al. (2020) framework, even if not organised this way. For example, it includes features compatible with indicators covered in Cocco et al.'s framework under 'analytical performance and associated operational requirements', 'clinical utility', 'human factors', 'infrastructural requirements' and 'regulation'. Although 'procurement' is also mentioned, no further details are provided.

Implications: **Our key learning point from scoping this TPP is that efforts to develop TPPs for digital and AI-involving diagnostic devices and multi-component tests may need to consider the same overarching categories of features as other test types. However, they will likely involve additional complexity and require attention to unique parts/individual elements of multi-component tests** (e.g. the AI algorithm and software, digital hardware device, associated app, and potentially other associated tests). Thus, and for illustrative purposes only, we conclude that while some overarching feature categories covered in the Cocco et al. framework may apply to the overall diagnostic device (e.g. intended use, environmental impact, and cost/health-economics-related information), there may be a need to consider other feature categories at the level of individual test components (e.g. the digital imaging device component, the AI/machine-learning and decision-support software components, and the app component). For example, this might apply to features related to unique analytical performance (whether the component is accurately capturing and measuring the required data), clinical validity (whether the component is measuring and capturing the correct information), clinical utility (whether the component is determining an intended care pathway and patient outcome as intended), human factors (e.g. training, instructions for use, result format/visualisation and interpretation), and infrastructure and regulatory requirements, etc.

What constitutes common and unique elements across a test's components and TPP may vary on a case-to-case basis.

In addition, the regulation of software as a medical device is evolving; thus, regulatory specifications may present additional layers of complexity for such TPP specifications, depending on regulatory jurisdictions.

2.1.2. TPP features in light of the CanTest framework for evaluating diagnostic strategies

Alongside insights from research exploring and developing TPPs, we considered recent efforts to develop a framework for diagnostic cancer-test evaluation. This process involved specifying test requirements that should, in principle, align with what matters from a test evaluation perspective; hence, we also wanted to explore the latter. This matters because, as well as guiding development efforts and signalling needs to innovators, TPPs are also meant to help innovators understand what they must demonstrate for evaluation and assessment purposes. While we looked into this further as part of work to draw out evidence requirements and sources for cancer diagnostic tests from an HTA perspective, we specifically considered recent learning from the work of Walter et al. (2019).[12]

Walter et al. (2019) conducted a systematic review of diagnostic-test evaluation frameworks and consensus research to develop a novel framework – the CanTest framework – that can incorporate more diverse diagnostic-test uses than previous frameworks and considers the non-linearity of test development processes, e.g. not just for diagnosis and surveillance but also to assist triage and as part of broader or more complex testing strategies involving other tests. The authors considered different test development phases, the important indicator types, and

potential evidence sources for different indicators.[12]

Reassuringly, many of the indicator types Walter et al. (2019) refer to in the CanTest framework align relatively well with the indicator types Cocco et al. (2020) identified in TPP development efforts, although the terminology used to describe specific feature types varied.[1,12] For example, Walter et al. (2019) refer to relevant concepts also reflected in the Cocco et al. framework, including:

- Technical performance (e.g. diagnostic accuracy, analytical validity/test reproducibility in intended settings and quality).
- Operational requirements and processes (e.g. sampling, sample processing, quality control), including human factor considerations (e.g. staffing, clinician acceptability, considerations related to appropriate test use and clinical interpretation and safety precautions/ measures).

However, the CanTest framework also highlights the importance of a broader set of human factors often neglected/under-considered in TPP specifications.[12] Examples include:

- Patient acceptability.
- Implications on downstream clinical workflows and patient care pathways, possible requirements for test follow-up, potential effects on triage, and fit with/ incorporation into broader diagnostic strategies (although this could potentially also fall within the scope of features covered within TPPs such as intended use, medical decision to be determined, medical need or test rationale). Some TPPs consider alignment issues with clinical workflow and care linkage, but this is relatively rare.

There is also scope for more thorough consideration of accessibility in TPPs, particularly regarding inequalities. The Cocco et al. (2020) review[1] covers some considerations related to market access and routes to market in identified features, such as expected manufacturing scale, potential market, market segmentation/channels to the market, region(s) of commercialisation and competitive landscape. However, these are not the same as accessibility. Features related to infrastructural requirements can also consider access and accessibility in the context of a test's suitability for remote conditions and resource poor-settings, and pricing considerations can also reflect access requirements. However, we found no accessibility considerations related to inequalities and specific user groups (e.g. those with disabilities, underserved communities and communities with specific cultural sensitivities) in the TPP sample we analysed or in the supportive information underlying the Cocco et al. (2020) typology. However, The CanTest framework considers effects on inequalities (and accessibility may conceptually be related to considerations of inequalities).[12]

The CanTest framework also flags the importance of clinical effectiveness and cost-effectiveness considerations.[12] These insights have implications for efforts to consider how health economic modelling can inform TPPs:

- Some but not all TPPs specify intended clinical outcomes (clinical effectiveness or other intended patient outcomes such as quality of life). In addition, while analytical performance is discussed in many TPPs and can capture indicators related to performance compared to a control/ reference method, Walter et al. (2019) link this indicator to a diagnostic strategy's effectiveness compared to a reference strategy/standard testing processes – implying comparative effectiveness in a real-world clinical context, as opposed to in the controlled lab conditions that tend to be covered in TPP specifications (to the best of our understanding).
- Regarding cost-effectiveness, many TPPs provide specifications related to costs per test. However, we have not encountered wide-scale use of cost-effectiveness specifications or EEM for TPPs.
- We did not encounter any TPPs in our scoping work that considered broader effects on health systems, e.g. regarding utilisation, referral patterns, and system costs. However, these are identified in Walter et al.'s 2019 CanTest framework and have implications for issues needing consideration in EEM.
- Similarly, the CanTest framework raises considerations of population-level health

outcomes (such as mortality, inequalities and disease stage at diagnosis) that are not mirrored in TPPs and could likely only be accommodated through early health economic modelling.

In summary, the Walter et al. (2019) CanTest framework highlights some feature considerations that do not appear prominent in current TPPs developed to signal areas of need to innovators, particularly on matters related to:

- **Non-economic impacts on patients, such as patient test acceptability and experience** (others, such as impacts on quality of life and survival, could be considered under 'clinical utility'. Impacts on carers/significant others also merit consideration for some tests).
- **Addressing inequalities** (though this is implicit in some TPPs developed to address the specific needs of resource-poor settings and communities in the global health space and directly reflected in their unmet need rather than a distinct consideration, which could be important for oncology and a UK context).
- **Downstream effects on care pathways and processes (e.g. the implications of test findings on further care needs and processes or interactions with other tests).**
- **Capturing downstream impacts on health systems and population-level outcomes (potentially as part of/through economic modelling).**
- **The importance of meeting real-world performance needs, not just performance in laboratory conditions.**
- **The importance of cost-effectiveness alongside cost considerations in health economic modelling.**

Walter et al. (2019) also provide some insights into the research types that can provide acceptable evidence to inform evaluations, commenting on designs spanning case series and case-control designs, cohort studies, qualitative research natural experiments, randomised control trials, routine data analysis and health economic modelling.[12]

2.1.3. Efforts to develop diagnostic TPPs in oncology

Developing diagnostic TPPs as demand signalling tools in oncology is nascent. We know of one ongoing effort (CRUK-funded) in the ovarian cancer space involving some of our project advisors (Professor Larry Kessler, Professor Bethany Shinkins and Dr Brian Nicholson). There is also an effort to develop a diagnostic test for HPV (linked to cervical cancer) involving one of our project advisors (Professor Mike Messenger). However, information on this is not publicly available at the time of writing.

We have seen an early draft of the work in progress for the diagnostic TPP for ovarian cancer, and there are a few points that stand out regarding development process practice and features considered in previous TPP efforts**:**

- **The importance of a test's fit with existing HTA/NICE guidelines and regulatory (MHRA) requirements.** However, it is worth noting that industry does not produce products for the UK market alone. A TPP focused solely on UK regulatory and HTA requirements or UK-based unmet needs could be less of an incentive for some industry players.
- **The capacity for post-market surveillance data collection** given regulatory requirements) – extending beyond adverse events reporting to monitor real-world performance, too.
- **The developers of the TPP also recognise the complexity of cost and quality-adjusted life year (QALY) assessments for diagnostics, which lead to QALY impacts indirectly and further down the line**. In line with this, the TPP semes to adopt a nuanced approach to understanding clinical utility – looking at contribution to ultimate patient outcomes through proxy measures such as overall reduction for the interval from presentation to diagnosis, reduction in further diagnostics needs, and contribution to earlier stage diagnosis.
- **Health equity issues seem to be considered, often overlooked in some TPP development efforts.**

- **Under costs, the TPP considers some implementation costs in terms of clinician and patient time (in addition to test and instrument costs).**

How these considerations will translate into TPP specifications and how far these features are amenable to quantification is yet to be seen, given that TPP development remains a work in progress.

2.2. Insights into TPP development processes

2.2.1. A summary of TPP development processes

TPP development efforts typically start with a set-up phase to establish a core working group, agree on governance and oversight arrangements and specify a plan of action. This can include establishing the core working group's terms of reference and ensuring no conflict of interest.

According to a recent systematic review of TPPs for medical tests by Cocco et al. (2020),[1] this is followed by three general implementation phases: a scoping phase, a drafting phase and a consensus phase.[1] In this section, we first briefly summarise and then elaborate on each phase based on examining a sample of TPPs[2-9] and insights from Cocco et al.'s (2020) systematic review.[1] We briefly outline each phase below:

- **The scoping phase aims to overview the relevant disease area and the limitations of existing technologies/products to understand unmet needs, key desired features and the case for a novel test.** This phase also involves identifying stakeholder and organisation types to involve in further TPP development phases.
- **The drafting phase involves detailing the desired features based on the scoping phase's insights and considering specifications relating to the new test's desired features.** Clearly defined optimal and minimal specifications (where possible and a priority) are essential for this phase. Building on insights from the scoping phase, the core working group can develop and iteratively revise an initial draft TPP via consultation with external experts. Since the drafting and consensus phases are not always sequential, various consensus methods can be employed concurrently in some TPP development projects and more sequentially in others.
- **The consensus-building phase seeks to reach a consensus regarding the TPP's features and specifications (e.g. optimal and minimal specifications).** This phase aims to find agreement on TPP features and specifications to finalise the TPP.

Below, we outline some overarching insights based on our analysis of the TPP development approach, processes, and methods.

The nature of the development process:

- **Scoping, drafting and consensus phases are often iterative, non-sequential and interconnected**. Instead, their boundaries are frequently blurred; TPPs in development often involve iteration between the drafting and consensus-building phases, for example, refining one draft to create the next version of the TPP.
- **Diverse stakeholders are involved in TPP development efforts in a staged way. Usually, a key group leads/drives the effort and engages a broader set of expertise from diverse groups as the development effort evolves. However, the stakeholders involved vary across TPPs:**
  - » **There seem to be 'fit for purpose' pragmatic considerations about who to engage feasibly and efficiently and at which stage of the process, i.e. not all potentially relevant stakeholders are involved in each phase, and key groups' involvement seems to scale up in later development stages.** The process usually begins with a smaller core group of experts, e.g. leads from organisations funding and/or overseeing the TPP development effort and a smaller number of external technical and research experts and/or consultants). This core group is involved in scoping and earlier drafting phases, sometimes complemented by interviews with external experts to help refine thinking. A wider pool of stakeholders, e.g. diagnostics/life sciences experts, clinical/healthcare professionals, and patients/public, typically input into

the drafting and consensus-building phases as participants in surveys, meetings, workshops, additional interviews or by commenting on drafts. Patient and policymaker involvement was not always made explicit in the TPP sample we analysed.

» **The evidence base should include more information on the involvement of healthcare professionals, industry, regulators, HTA agencies, policymakers, payers, and patients/ the public to some extent.** Given that diagnostic TPPs aim to guide the diagnostic test development and evaluation process, it seems crucial that their features and specifications align with the expectations of end users (e.g. clinicians and patients) and those developing diagnostic tests (e.g. industry and/or researchers in academia/ research institutions) and with regulatory, HTA agencies and policymakers. However, as much of the literature does not detail these groups' involvement, further insights are needed about how, when and in what to involve them.

» Regarding the patient/public voice, while it is important to pursue diversity in understanding unmet needs from a patient or carer and broader public perspective, it is challenging (if not impossible) to achieve full representativeness (as discussed in a conversation with a project consultant).

Nature of the evidence base on TPP development:

- **The degree of publicly available information and detail (transparency) on a TPP development process and, thus, on what informed final specification decisions varies significantly across TPPs. Future efforts could focus on ensuring more comprehensive, transparent information to support clarity on the process and assist future efforts, including TPP refinements over time as new information becomes available.** TPPs were either published as reports by the lead organisation developing the TPP or in an academic journal and tended to include the most detail on methods. There is scope for academic publications on TPPs to provide further details as part of supplementary material.
- **There is also limited detail and inconsistent clarity about how qualitative methods inform TPP development processes.** TPPs often mention using stakeholders' comments or free-text inputs during drafting or consensus efforts or interviews in the scoping and drafting phases but rarely give details on the nature of questions asked, the issues explored (e.g. interview protocols) or the methods they used to analyse and incorporate those findings. This ambiguity limits our ability to learn from previous TPP development efforts and points to potential limitations in rigorous processes and transparency versus pragmatism. This aspect is likely to vary across different TPP development initiatives.
- In the following content, we provide further information on each phase based on what we have learnt from our own 'scoping' of the issues. Figure 3 provides a visual summary of the key TPP development phases.

2.2.2. Insights on the scoping phase

Aims: Drawing on our analysis of TPPs and of the systematic review by Cocco et al. (2020),[1] the main aims of a scoping phase tend to include:[2-9]

- **Assessing the need for a TPP for improved diagnostic tests** (i.e. the rationale for developing a TPP).
- **Specifying the TPP's scope regarding key features such** as clarity on unmet need, intended use, target populations, health-system level and users, and the broader features required in a novel test.
- **Identifying the stakeholders and organisations who should be involved in further phases of TPP development, e.g. drafting and consensus-building.**

Key activities: the scoping phase can involve activities to support the above aims, including:

- **Identifying existing types of tests and technologies on the market and their limitations.**
- **Analysing the broader evidence base to understand the unmet needs and a TPP's desired scope and features.**

These activities are sometimes referred to as conducting a 'needs assessment',[2,4,7] which

Figure 3. A summary of the TPP development phases

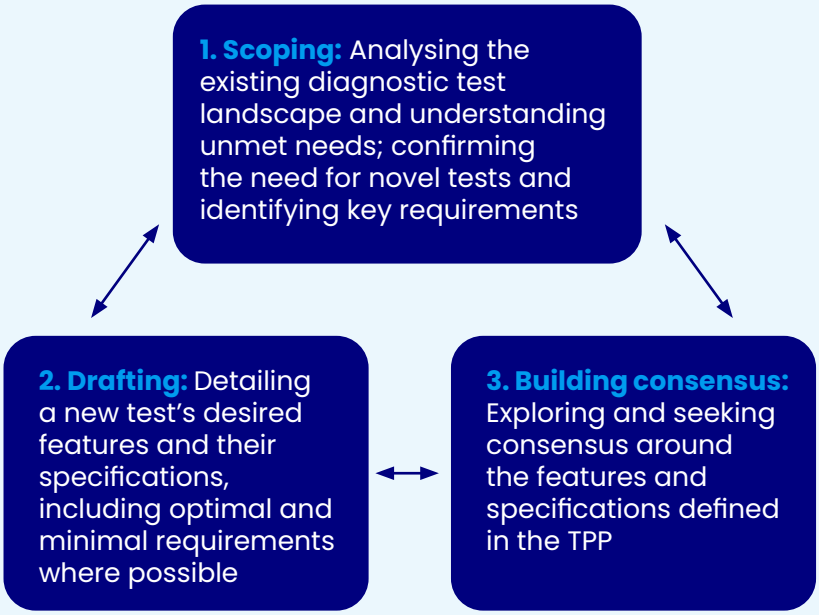
Core phases in developing a diagnostic TPP: Learning to date

Inception/preparation:

- Establish governance and coordination arrangements
- Assemble the core working group, deciding on its composition, size and constituent roles
- Agree on a plan of action and expected timelines
- Identify stakeholders and organisations to engage during the process

Implement development (iterative processes that may overlap):

- Scoping
- Drafting
- Building consensus



Scoping phase: Key insights to date

PURPOSE	Arrive at an output that: • Assesses the need for a novel test • Specifies the scope (e.g. unmet need, intended use, target populations, use setting, target users, and key desired features, e.g. technical, care pathway and patient access) • Identifies the stakeholders and organisations to involve in later TPP development
PEOPLE	• A core working group of experts leading TPP development, possibly divided into subgroups leading different tasks (e.g. unmet need, scope, diagnostic landscape) • A limited number of external experts to consult about unmet needs, key desired test features or the existing diagnostic landscape, e.g. researchers, clinicians, industry, or (occasionally) patient/public voice representatives or intermediaries • Minimal HTA, regulator, payer or policymaker involvement observed
ACTIVITIES AND METHODS	• Analysing the evidence base to understand the unmet need, appropriate TPP scope and key test features required (e.g. systematic literature reviews or rapid evidence assessments of academic/grey literature and guidelines) • Identifying available tests on the market and their limitations to confirm unmet needs via desk research of publicly available information or commercial data repositories • Initial stakeholder consultation via interviews or workshops with a limited number of individuals to refine and/or confirm understanding (variable scope and rigour) • Core working group meetings to discuss insights

Drafting phase: Key insights to date

PURPOSE

- Bring together insights from the scoping phase and produce an initial draft TPP with specifications for diverse features, and evolve it into more refined and advanced drafts over time
- Identify gaps, i.e. features that need specifying but for which the scoping phase did not provide sufficient information, and thus further stakeholder consultation or desk research is needed
- Explore acceptable evidence levels for demonstrating that a novel test meets TPP requirements, e.g. trials/study types – though this is not always done

PEOPLE

- The core working group leads and develops an initial draft
- A broader set of stakeholders are involved to help specify features in the initial draft and evolve further drafts, e.g. research and technical experts in the diagnosis and disease area, healthcare professionals and industry, patient and public representatives

ACTIVITIES AND METHODS

- Working group desk research and meetings to specify different features' optimal and minimal requirements and associated comments/reasonings for the choices made.
- Broader stakeholder consultation through various methods, ranging from invitations to provide written feedback on drafts (e.g. free text comments) to interviews to using (Delphi-inspired) consensus surveys, with the additional potential for workshops
- Modelling is sometimes needed to finalise specifications for more challenging technical performance features (it is not always possible to rely on literature or expert opinion for all feature specifications)
- Multiple drafts are possible; two to three are common, but there can be more
- Cost-effectiveness or early economic modelling are barely mentioned
- Acceptable evidence levels are rarely clarified; some consider it outside a TPP's role

Consensus phase: Key insights to date

PURPOSE

- To explore and arrive at a consensus on a final draft of a TPP

PEOPLE

- The core working group of experts leading the TPP development
- Broader stakeholder engagement, featuring considerable variety in who engages in consensus building and how (no golden rule)

ACTIVITIES AND METHODS

- Initial draft-focused consensus meetings in the core working group, comprising a mix of virtual and face-to-face meetings
- Wider consensus exploration:
  - » Surveys (Delphi), which can include healthcare professionals, clinical academics, industry representatives, and sometimes product developers. There is consideration variety in who participates and a lack of clarity on this in the evidence base
  - » Delphi-inspired consensus workshops with specific stakeholder groups and inviting external experts into consensus meetings
  - » Opening up the later/pre-final or final drafts to wider public commentary/consultation
- Consensus thresholds vary; they are generally 75% or more but occasionally 50% (they can also be lower in earlier drafts and higher in the final draft if multiple consensus rounds are necessary)
- Consensus can be sought on specified features first, followed by optimal and minimal specifications, or just the latter
- Relevant experts can be interviewed to discuss features for which consensus is low

often (but not always) includes a 'landscape review'.[4,9] This helps to understand the unmet need and the case for developing a TPP, including funding, sponsoring, and overseeing organisations' goals. It also helps to specify the TPP's scope and remit.

Methods supporting the activities: Scoping can involve one, all, or varying combinations of the following:[1-9]

- **Desk-based reviews of academic literature**, e.g. systematic reviews or lighter-touch literature reviews, or **grey literature**, e.g. policy documents, guidelines for developing prior TPP efforts and technical reports.
- **Mapping of commercially available tests** (e.g. repositories of existing products).
- **Stakeholder consultations** to understand the existing landscape and unmet needs.

Desk-based research (e.g. literature review, other desk-based research and data repositories analysis) seems more common in the TPP scoping phase than stakeholder consultation. However, we acknowledge that this may also be an artefact of the sample we analysed. The few TPPs in our sample that incorporated stakeholder consultation in their scoping phase[2,4,9] often did so through structured interviews exploring context and gaps in the relevant diagnostic space.

Stakeholder and expert interviews were often used to identify gaps in relevant diagnostics regarding unmet needs and help identify available commercial products that those developing TPPs could consider in their landscape reviews. When used in isolation, interviews with stakeholders helped identify the gaps that needed addressing. When combined with desk-based scoping efforts, interviews were used more to verify and refine findings.

Stakeholders involved in the scoping phase:

The core group of individuals steering the TPP development process and scoping activities tended to include representatives of the organisation leading, overseeing and coordinating TPP development (e.g. international not-for-profit organisations such as PATH, FIND, WHO, Médecins Sans Frontières) and a small group of technical experts as consultants brought into the core group driving the effort (usually comprised of consultants and clinical academics).[2-9]

Wider stakeholders consulted as part of scoping stage work, such as through stakeholder interviews, could include healthcare professionals, researchers, public health experts and sometimes the public through a public consultation.[1,2,4,9]

Overall, the stakeholder types involved in the scoping phase tended to span relevant technical and topic area experts, public health professionals, international non-profits overseeing the overall development effort, and other academics/researchers to help assess priorities, gaps, and needs in the diagnostic area of interest.[2,4,9] Healthcare professionals were also sometimes brought in to assess the needs and perspectives of diagnostic end-users.[2,4,9] Few TPPs introduced a public perspective early on by consulting with the public at this scoping stage. However, there is limited information in the evidence base we consulted about how and regarding what the public was consulted.[9] Many sample cases purposely chose stakeholders based on their experience in relevant geographies or specific clinical facilities.[2,4,9] Although industry representatives can be involved in scoping activities,[1] this happens inconsistently across TPP development efforts.

Insights about the scoping phase based on the sample we examined: **How much detail documents describing TPP development efforts provide on the activities and methods used for the scoping phase varies significantly**. For example, some TPPs provided no information on an explicit method for the needs assessment or about conducting a literature review or landscape analysis/review.[3,5,6,8] These TPPs were often published as academic articles, where the need for the TPP was presented as part of background/contextual information in the paper's introductory sections rather than describing a formal 'scoping phase' to understand unmet needs and make the case for a TPP.

Other TPPs describe the scoping phase as an integral task in TPP development and provide considerable detail about the specific research methods used to inform this phase (e.g. literature reviews, other desk-based research and expert consultation) and step-by-step descriptions of landscape review processes.[2,4,7,9] These needs assessments were published separately and referenced in the TPPs. The methodological transparency and detail supported a compelling case for developing a

TPP.

2.2.3. Insights on the drafting phase

Aims: **This phase aims to arrive at an initial TPP draft and gradually evolve it into more refined and advanced drafts.**

Activities: **Key activities involve developing an initial draft and gradually advancing it into more mature ones via stakeholder consultation.**

Methods: **Methods adopted in the drafting phase vary from opportunities to provide written feedback on drafts (e.g. free-text comments) and stakeholder interviews to consensus surveys (illustrating the often iterative and non-sequential nature of drafting and consensus stages).**

Core working groups driving a TPP's development (e.g. with representation from organisations in TPP oversight/governance roles and funders in consultation with technical experts from various sectors such as academia/research and healthcare professionals) usually lead the drafting of TPPs.[1-9] This commonly involves a draft version '0' TPP[2-9] based on the scoping phase's insights and utilising outlines and examples from specific funder or governance/oversight organisations (e.g. WHO, FIND) in previous efforts to develop TPPs (i.e. structures and presentation formats used in the past for other TPPs),[2,3,5-9] This implies a degree of path-dependency but also value in overarching and guiding frameworks and approaches.

Version '0' typically follows a table format with optimal and minimal specifications for desired features. Some TPPs also include comments defining or explaining features or specifications in more detail or a rationale for a feature's optimal and minimal specifications.

After the initial draft, those developing a TPP tend to elicit stakeholder feedback through general ask-for-comment or semi-structured interviews or surveys. This phase sometimes overlaps with consensus exploration via Delphi-based methods (see consensus phase for further detail). How general stakeholder comments were utilised to make changes was often unclear in the evidence base we analysed.[2,5,6,9] TPPs that reported using semi-structured interviews tended to provide information on the general topics covered, such as questions on current tools, users and screening/diagnosis challenges, the performance/product specifications of a new test (to understand how this relates to version '0''s content), and price-related information and channels to market.[3,4,6] This information was then fed into the core working group's efforts to refine the draft protocols. When public consultations were used, data collection often involved surveys or sharing TPP drafts on public portals/websites to allow for comment/consultation.[6,9]

There are generally multiple drafts (often 2–3 based on the TPPs we looked at) before a final TPP is drafted and published.[1] Each draft is refined based on consultations with stakeholders, presenting them with the latest version and inviting feedback.[2-9]

Stakeholders involved: **There was significant variation in the composition of stakeholders, the draft they were consulted on, the consultation method, and the utilisation of their feedback among the TPPs we analysed. However, one common feature was a core group of individuals driving the drafting (i.e. the overall leads of TPP development), complemented by broader stakeholder consultation.**

Earlier TPP drafts are usually drafted by the core group leading the TPP development effort,[1-9] often including stakeholders from organisations leading the TPP development effort, i.e. those overseeing and/or funding the process (e.g. leads at WHO, PATH or FIND TPP in our sample), and other external experts/consultants invited to join. Further drafts are then refined and developed in consultation with 'outside' experts, spanning experts from academia,[3-5] public health,[4,5,7] healthcare,[2-4,7,9] research,[2,4,5] industry,[3,5,8] clinical end-use[3-5] and the public.[6,9] It is worth flagging that the documents we reviewed often used the term 'technical experts' loosely.

Stakeholders brought in during the drafting phases differed between TPPs and between different drafts of the same TPP. Some TPP development efforts utilised the same group of stakeholders throughout all TPP drafts (scaling up the number of individuals inputting at different stages); others used different stakeholder input to evolve different drafts. For instance, one TPP had an expert panel working group comprised of various

stakeholders representing industry, researchers and healthcare professionals.[6] In another TPP development effort, the first draft was drafted by the TPP development group (core working group) of technical experts, while later drafts were refined through other stakeholders' verification, comments and wider inputs or verification of the appropriateness of prior drafts, such as the public, clinical end users, and broader pool researchers.[9]

Insights on the nature of the evidence, based on our sample: There was no consistent granularity in the level of information TPPs provided on stakeholder identities in our sample. While some provided tables with each stakeholder's name, title and affiliations,[3-5] others only described them collectively (as 'external stakeholders') or used vague terms such as 'experts.'[6,9] TPPs also described potentially similar or overlapping stakeholder groups differently, e.g. one TPP described a 'private-sector'[5] stakeholder and another described an 'industry'[8] stakeholder – the former could include the latter.

2.2.4. Insights on the consensus phase

Aims: **The consensus phase (sometimes concurrent with drafting) aims to explore and arrive at a consensus for a final TPP.**

Activities: **Consensus is often sought to strengthen each TPP draft's refinement.[1] However, a minority of TPPs we analysed did not seek consensus**.[4] Some TPPs embedded consensus in every draft,[3,5,6] while others only pursued consensus-exploration-and-building for some drafts.[7] Some TPPs only applied consensus methods to the final phases[2,8] if at all, or it was unclear if consensus was involved.[4,9] For example, PATH TPPs in our sample did not have a consensus phase in their TPPs, conducting the scoping phase via a literature review and desk research in conjunction with stakeholder interviews.[4]

Methods: **Consensus usually took the form of a Delphi (or Delphi-inspired) methodology.[1-3,5-8]** The consensus phase varied depending on the number of consensus rounds, the threshold for defining consensus, whether consensus-building included qualitative narrative data, and which stakeholders participated in the consensus phases. Interviews were sometimes conducted to discuss features where consensus was low.

A consensus threshold of 75% or more stakeholders agreeing or strongly agreeing to a particular characteristic's inclusion in a TPP was the most common threshold in our sample. However, there was some variance (e.g. some TPPs used a 50% threshold). Consensus thresholds seemed higher when consensus was only sought once, usually before the final TPP was drafted (70–75%).[5-8] Where consensus was sought for multiple drafts of the same TPP, the first consensus threshold was often lower (e.g. > 50%)[6,8] and focused on the nature of the characteristics under consideration. In contrast, later drafts used higher consensus thresholds (e.g. > 70%) focused on specific criteria for minimal and optimal specifications.

Based on our sample, Delphi surveys often used Likert scale responses,[2,3,5-8] and a few TPPs included the opportunity to comment if stakeholders said they disagreed on a feature or a free-text comment at the end to add their thoughts if they felt something important had not been covered.[5,6] Exactly how the free text was incorporated into consensus building was not specified, however. Qualitative narrative data was also elicited in some TPP later drafts regarding high-priority characteristics, which stakeholders discussed in more detail during the Delphi process.[2,3,6,8] Those leading TPP development efforts often dealt with the most controversial characteristics and/or specifications (i.e. those with less consensus) by engaging with individual, relevant topic experts and conducting interviews with them.[1-3,6,8]

According to our sample observations, consensus meetings were used to consult with relevant field experts on specific characteristics that did not meet consensus thresholds to discuss and decide on their inclusion/exclusion.[6,8] Inclusion was usually decided based on votes within the specific field expertise group for a given characteristic.

Stakeholders: The stakeholders consulted as part of consensus building (whether via Delphi or more qualitative means) varied among TPPs and across different consensus phases. In many TPPs, consensus was initially sought internally amongst the core experts and funding and/or overseeing body of the TPP development process (akin to the core working group) through a workshop or meeting.[1-3,5-8] Wider consultation with broader stakeholders, e.g. healthcare professionals, industry

representatives, and/or the public, tended to happen for later drafts.[2,3,5-8] Healthcare professionals, academic clinicians, industry representatives and product developers typically partook in Delphi surveys[2,3,5-8] whereas the public was consulted via a public questionnaire or survey (the survey/questionnaire type was unclear).[6,9]

2.3. The governance of TPP development processes

Not-for-profit organisations appeared to lead and provide oversight of TPP development efforts when TPPs were used as demand signalling tools: Our sample of TPPs[2-9] showed the prominence of international organisations like the WHO, FIND, Médecins Sans Frontières and PATH in leading the development of TPPs (and being on the author lists of TPP reports). This prominence is not surprising given that our sample was primarily drawn from fields in which TPP development is most evident, i.e. neglected infectious diseases and global health. Technical units within these organisations can initially establish the need for a TPP development (given their organisations' broader remit, knowledge and experience) and produce a brief document outlining a novel test's need, scope and purpose (an initial short outline) to secure internal clearance to embark on a TPP development effort (based on communication with an advisor).

These organisations often cited their in-house guidelines and previously published TPPs as models for TPP development.[3,5-9] FIND and WHO both follow standard in-house guidelines (not publicly available), while PATH publishes literature reviews and other material as part of the TPP development on their website.[4] These publications are a useful resource for informing future efforts.

Organisations overseeing and leading TPP efforts (e.g. FIND and WHO) often link the TPP development rationale to their organisations' broader remit, citing its previously established specific public health goals.[2-9]

Such organisations typically utilise in-house expertise during the first round of scoping and drafting and complement this with a select number of external experts/consultants brought into the core working group or to advise. They design the development process to engage a much more comprehensive range of expertise (e.g. including industry and healthcare professionals)[1-9] and the public (to varying degrees)[6,9] as participants contributing to and informing the development process via consultation.

While organisations providing overall leadership in developing TPPs (e.g. PATH, FIND, WHO and Médecins Sans Frontières) offer guidance and authorship, they are not always the sole funders of the TPP development. **TPP development has also been partly or fully supported (i.e. funded) by research funding bodies like Wellcome, Fondation Botnar and the Bill and Melinda Gates Foundation, charities like the International Committee of the Red Cross, and governmental departments or programmes like UK Aid, Swiss Agency for Development and Cooperation, UK Foreign, Commonwealth and Development Office (formerly the Department for International Development).**

Organisations leading TPP development were committed to publishing their outputs in the public domain, though formats vary. FIND tended to publish their TPPs through academic journals, often meaning less detail on some aspects of methodology than others, such as WHO TPPs published in longer reports.[6,9]

2.4. Insights from the interviews

This section describes key insights from the exploratory scoping interviews conducted. We discuss interviewees' insights regarding the approach to and process of developing diagnostic TPPs, key feature-related considerations, and enablers and barriers to TPP development and use. We also explore learning from experts involved in TPP development as demand signalling tools internationally and views from UK oncology experts (academic and test evaluation, industry and charity). For purposes of informed consent and preserving anonymity, we use interviewee codes.

2.4.1. Insights on the approach and process of developing diagnostic TPPs as demand signalling tools

There is no single 'correct' way to develop a TPP, and many efforts focus on 'rightsizing' the approach to available resources (Int. 7). Different methodological options can have trade-offs. However, it is vital to balance what is ideal with what is feasible in a given context (Int. 4 and 7). However, the processes must also be robust, balancing gold standards with pragmatism (Int. 5).

Relying on expert opinion alone is insufficient (Int. 2 and 5), and decisions on feature specifications must go beyond expert consultation. Any numbers stated in a specification must be evidence-based; thus, specialist opinion should be complemented with desk research, and modelling may also be needed to determine a quality specification for some features (Int. 2), particularly in the absence of a gold standard. Access to modelling experts early in the process can be beneficial (Int. 2 and 5) to help identify and specify an unmet need.

It is essential to engage representatives of all relevant stakeholders in the TPP development process (Int. 2, 3 and 4), and methods can vary. Interdisciplinary perspectives are needed, including early on (Int. 4). However, not all groups and organisations need to be involved in every stage of the process in the same way (e.g. not everyone can complete a technical Delphi survey) or every issue. Some groups will be better able to engage with specific stages of the process (e.g. commenting on detailed specifications) than others (Int. 4). It is important to include different stakeholders' voices early and not just once the project is mature (Int. 3):

- The types of **research and clinical/health care professionals** to engage will depend on the tumour site, use case and test type, e.g. types of oncologists, specialists in different diagnostic test types, such as imaging or liquid biopsy or tissue biopsy or endoscopies, pathologists and path lab experts and primary care professionals (Int. 6). **Health economics expertise** also matters (Int. 6). Since clinicians are also end users it is vital to involve them in TPP development (Int. 3 and 7).
- When considering end-user voices, **healthcare management** roles are also critical, as they have insights into healthcare provider organisation budgets and resources (Int. 1) that can affect the likelihood of novel diagnostic uptake.
- There is a general need to improve how **patient and public** voices are integrated within TPP development to better understand their perspectives and experiences of diagnosis and unmet needs (Int. 2, 3, 5 and 7). There may be value in future work to develop a framework for how best to engage patients and the public in TPP development, considering when, where and how patient and public involvement can add the most value (Int. 4). Applying such a framework to bespoke efforts may vary depending on the test's risk level and invasiveness among other factors (Int. 4). While eliciting direct input from patient and public representatives is ideal, some TPPs have sought their perspectives via intermediaries such as healthcare professionals working with patients on the ground with knowledge of patient views (Int. 2).
- A **regulatory** lens also matters (Int. 1 and 3), including that of **HTA agencies** (Int. 7). However, these voices are not always heard directly but through intermediaries aware of HTA requirements, even if they don't work in HTA agencies (Int. 7). Members of guideline committees, including from NICE and HTA agencies, are also relevant (Int. 6).
- **Industry** involvement is also crucial as without it, there is a risk that a TPP will not prove feasible or of interest to the industry (Int. 3, 5, 6 and 7). The involvement of small and medium-sized enterprises (SMEs) as well as large companies is essential (Int. 3) to ensure a diverse mix. Industry provides a viability 'hat' and can bring regulatory awareness (Int. 5).
- It is also vital to bring relevant stakeholders from the broader **policy and funding** landscape into the process (Int. 4). Funders can include government agencies, arm's length bodies and research charities.

The time it takes to develop a TPP also varies widely, partly because those leading development must fit the approach and methods to the available resources and output urgency, and depends on the granularity required (Int. 2, 4 and 7). It can take a few months (Int. 2 and 7) or longer. According to one advisor, the process often takes approximately

6–12 months. In a UK oncology context, the first efforts to develop a demand signalling TPP will likely take longer as it will be a novel and likely comprehensive process (Int. 4).

According to one interviewee in a UK oncology context, there is a need to improve the pipeline of novel diagnostic tests emerging, particularly for use in primary care and screening (Int. 4). There is also a related need to tackle the disconnect between the technology available or becoming available and unmet needs and demand areas (Int. 4), an area in which TPPs have a role to play. Not-for-profit organisations such as cancer charities could help steer, facilitate or coordinate the development of TPPs, but it is not a job for any one organisation. Success will depend on collaboration with the broader landscape of stakeholders who set policy and practice or implement it in oncology care, meaning a broader set of bodies will need buy-in and involvement (Int. 4). Although not entirely clear yet who would need to have an overseeing role or a degree of ownership, examples might include the National Health Service (NHS) England and organisations across the devolved nations, the Office of Life Sciences, Innovation Consortium in Scotland, NICE, Innovate UK, NICE and MHRA.

In terms of specific TPP development stages (inception, scoping drafting and consensus building):

- In the TPP inception stage, it is important to establish an expert group to lead the effort, including diverse expertise relevant to the disease area and its diagnosis as a starting point (Int. 2).
- In the scoping stage, the key is identifying and understanding the unmet need and use case (Int. 2 and 5) and the existing diagnostic landscape (Int. 2). In this context, identifying who is going to use the test and for what is critical for the TPP to have any impact. A clear use case and value proposition are essential (Int. 7). Sometimes, TPP development efforts use a proportion of the core working group members to focus on identifying the unmet need and specifying the use case and key test features, while others map the existing diagnostic landscape and its limitations. Individuals working on these related scoping aspects can then come together to confirm alignment (or misalignment) between unmet needs versus the existing diagnostic test landscape (Int. 2). Some efforts begin with high aspirations and then scope down to what is realistically achievable (Int. 6). Consultation with developers (industry), end users (healthcare professionals and patient voice) and payers can help foster realism (Int. 6).
- Scoping can involve various methods, including literature reviews of different scopes and scales, stakeholder consultation and diagnostic-landscape mapping. It is also essential to provide a sense of the estimated market size and revenue potential, either using economic modelling and information on what price customers would be willing to pay or what they pay for similar or related products or engaging health economists to develop a revenue model (Int 6). Cost-effectiveness matters for HTA (Int. 5). However, effective health economics depends on good quality evidence; estimates, projections and models are only as valid as the underlying data on which they are based (Int. 1, 5, 7) and the assumptions and understanding of how new diagnostics would be used in clinical care pathways. This area can be challenging in the early development phases because of many uncertainties (Int 5).
- Regarding TPP drafting, it is helpful to decide in advance which features and specifications are essential, and which are 'nice to have' but not absolutely necessary, and the same for specifications and the reasons behind them (Int. 6).
- Regarding consensus, achieving alignment between diverse stakeholders can be challenging, and thus, a clear method and process need to be established and communicated (Int. 2). However, one interviewee suggested that high levels of disagreement are rare if people are consulted throughout the process, though there may be some individuals who do not agree with the majority – usually (though not always) resolved through dialogue (Int 2). Having processes and lines of responsibility for ultimate decision-making is essential, as the timeline and resources for developing a TPP are not infinite (Int. 2). Although Delphi processes are considered the gold standard, implementing them in

full can be expensive and depends on the resources available. Hence, Delphi-inspired or 'Delphi-light' approaches are also used (Int. 7), which adopt Delphi principles but may not include full Delphi surveys (e.g. consensus workshops). The resulting TPPs flag up information relevant to issues of consensus or uncertainty.

It is also essential to recognise that TPPs are living documents, not static ones (Int. 1, 3 and 7). Since TPPs can be updated, a clear plan should be in place to ensure their continued relevance and quality over time (Int 1). However, it is expensive to develop novel tests, and the industry is astute at spotting a TPP that is likely to go out of date before it can get a product on the market (Int 7).

One interviewee mentioned that considerable knowledge about TPP development in infectious diseases is transferrable/adaptable to other areas and product types in their experience, flagging that there is a lot of cross-pollination (Int. 7).

2.4.2. Insights related to feature-related considerations

TPPs must optimise clarity without overly compromising flexibility and experimentation for innovators responding to them. Therefore, TPPs should seek to prioritise and clarify (a) which features must be specified and which can be left to the innovator's discretion (Int. 1) and (b) which features can be specified with minimal and optimal requirements and which cannot (Int. 2, 4, 5 and 7). Ultimately, the key value proposition must be clear and evidence-based (Int. 1 and 5). Uncertainty in science is a major issue when accurately specifying analytical and clinical performance (Int .1). When deciding where to specify unequivocally and where to allow flexibility, mindfulness about TPP turnaround and diagnostic test development times is vital given NHS needs (Int. 3).

According to one interviewee, there are risks that developers will aim for the minimal requirements even if optimal requirements are also specified (Int. 5). A core working group leading the effort could be well placed to steer and provide insights on key requirements needing specification (Int. 4) while being mindful of what a feasible/viable TPP for industry to respond to looks like (Int. 3, 5 and 7). It is also important that TPPs are careful to create a level playing field between smaller SMEs and larger innovators with different resource levels (Int. 3).

There will also be trade-offs between different features' specifications (e.g. lower sensitivity versus higher accessibility), and these must be considered early on when identifying the unmet need and prioritising key novel test requirements (Int. 1, 2 and 7). For example, accuracy is not always the most important feature of a given diagnostic; rapid results, better accessibility or less harm/invasiveness (e.g. a biopsy versus a blood test) and quality of life or clinical utility can also be important considerations (Int 1). Ethical considerations about data and information also warrant consideration in TPPs.

It is essential to specify technical performance features, e.g. specimen type, specimen handling and transport, specimen stability, test analytical specificity and sensitivity, clinical diagnostic specificity, cross-reactivity, and negative and positive predictive value (Int. 5 and 6). However, wider, non-technical features also matter. Interviewees flagged diverse illustrative examples, including who the target population and users who will conduct the test are (Int. 5), economic and price considerations (Int. 3, 4, 5 and 6), what the NHS is willing to pay; information on the practicalities of integrating a test into care pathways, e.g. timelines and staff training (Int. 6 and 7), wider human factors (Int. 5) and issues related to general access (Int. 1, 3 and 7) that can impact the value proposition. Inequality considerations are a part of that (Int. 5 and 7), including considerations related to minimal and desirable features when testing validation in diverse populations (Int. 5). In terms of requirements, it is vital to be clear that the pathology laboratory's core needs (e.g. technical performance, convenience) may not be the same as the patients' (Int. 6). Regulatory considerations also matter (Int. 5) as do aspects related to safety more widely (e.g. contamination issues and quality assurance).

In the UK, the regulation and evaluation of diagnostics are less established and advanced than for drugs (Int. 1 and 2). Understanding evidence requirements for diagnostics from a regulatory and HTA perspective is essential to inform TPP development efforts. According to one interviewee, diagnostic R&D sometimes focuses on extreme positive/negative cases

to the detriment of in-between cases because extreme cases are easier to sample. Therefore, understanding an appropriate sample (one that constitutes acceptable evidence) is important for quality diagnostics (Int 1). According to this interviewee, TPP development should not be just about specifying features but also evidence requirements associated with those features. They suggest there is often a lack of clarity regarding acceptable-quality evidence (Int. 1). From a regulatory perspective, this can be a barrier/disincentive for companies even if the TPP is focused on a UK context if international regulation and alignment with it is not considered (Int. 1).

A TPP should also consider pathways for reimbursement for any tests that would be developed in response so that the TPP has utility and traction and can support effective interfaces between the industry and the NHS (Int. 3). The strongest signal to industry is a funding announcement for products – a credible signal that there will be a market (Int. 7).

2.4.3. Reflecting on key enablers and challenges

In reflecting on the insights discussed, we identified a series of key enablers and challenges related to TPP development or use (or both):

Key enablers include:

- A clear plan for TPP development (Int. 2) tailored to the resources available (Int. 2 and 7).
- Adequate funding for TPP development (Int. 7).
- Balancing the ideal with the pragmatic (Int. 4 and 7) regarding the TPP development approach and process and what can be specified in terms of features.
- Ensuring appropriate stakeholder engagement in TPP development (Int. 2, 4 and 7) to support the development of fit-for-purpose TPPs and increase the likelihood that the TPPs will be used. This includes ensuring that the TPP will resonate with those who will use and pay for any resulting tests (Int. 6).
- Keeping the use-case front and centre of the TPP development effort (Int. 4).
- Prioritising specifications for key features central to potential impact (Int. 2).
- Engaging modellers to help specify features where this is not easy to do from existing evidence (Int. 7).
- Being realistic in terms of what will actually incentivise industry and what can be feasibly delivered (Int. 6).

Key challenges in previous TPP efforts arose where many of the above-mentioned enablers were absent, including:

- Ensuring effective TPP take-up:
  - » Ensuring that a demand signalling TPP will be used to help bridge the translation gap and that there is sufficient oversight to enable focus on appropriate use cases and TPP take-up (Int. 4).
  - » Ensuring and mobilising buy-in from relevant stakeholders, including those who will pay for the test and use it. Going through NICE and being cost-effective does not guarantee adoption even if the specifications are UK-based (Int. 5), so efforts to develop TPPs must consider market viability upfront.
  - » Socialising the need for a TPP and novel tests in the wider landscape (Int. 4).
- Developing a feasible TPP for innovators to respond to:
  - » Introducing appropriate economic analyses into TPPs; there is limited research and certainty around the actual costs of tests (Int. 4). Sometimes estimated costs do not match industry views on what is feasible (Int. 7), and it is essential to address cost and usability considerations early in TPP development (Int. 3).
  - » Ensuring alignment with industry feasibility: Industry will not engage with a TPP that is considered unfeasible or unrealistic on any ground, be it technical or economic, and this can be a challenge with TPPs developed by the public sector (Int. 6, 7).
- Arriving at accurate and appropriate feature specifications in a TPP (Int. 7).
- Prioritising which TPP to develop first in the UK given multiple needs:
  - » Prioritising which TPP should be developed

first (Int. 4) in light of the potential for impact and for mobilising streamlined funding around that priority.

- Ensuring the TPP remains relevant long enough for it to be an attractive signalling tool for industry (Int. 1, 7):
  - » TPPs may respond to a current need, but there needs to be enough incentive for industry to respond to it as a viable market and in consideration of diagnostic test development timeframes. Industry is good at foreseeing whether the need and market will change in a few years (Int. 7).
- Creating a level playing field between smaller and larger companies in terms of input into TPP development and the ability to respond to a TPP:
  - » Getting companies to engage can sometimes be tricky, especially smaller companies (Int. 3).
  - » Involving industry in developing a TPP to reflect diversity and mitigate against biased output (Int. 3) and getting independent commercial advice.

2.5. Insights from the early economic modelling (EEM) scoping

In developing a TPP, it is essential to consider the perspective of the final decision-maker. NICE considers the cost-effectiveness of the health technologies it appraises by examining their impact on healthcare costs (including health service resource use) and health. It uses this information to determine whether the health technology offers good value for money, which is critical. If the technology does not provide good value for money, its adoption will displace more health benefits than it generates (i.e. any money spent on the new technology is unavailable to spend on cost-effective activities already in place).

There is a simple case for including EEM in a TPP. By considering the final decision-maker's perspective, we can assess the characteristics a diagnostic test must comprise for the payer to evaluate its cost-effectiveness favourably.

In this scoping exercise, we summarise the key literature related to the early economic evaluation (EEE) of diagnostic tests and the evaluation routes through NICE, considering which should guide our approach to EEM for TPPs in England.

2.5.1. A summary of key findings from the scoping exercise on health economics: a summary

Diagnostic technologies are distinct from other technologies in terms of **value generation**. The test's impact on clinical decision-making significantly influences cost-effectiveness, often more than the accuracy of the test itself.[13] This is due to the downstream impacts on health outcomes and costs associated with the change in the clinical decision-making that the diagnostic test informs. This difference should be considered in a new test's value proposition.

Clinical care pathway mapping is vital for establishing new diagnostic technologies due to the way value is generated. However, there can be difficulty benchmarking where no pathway previously existed, yielding greater uncertainty and difficulty demonstrating value.

TPP development comprises three phases with different elements of EEM having the potential to feed into each phase.

Key EEM analytical approaches include:

- **Care pathway analysis**: This can be undertaken during the scoping phase of the TPP development to establish the clinical care pathway. There will likely be uncertainty in the precise placement of the test in the pathway at this stage; therefore, the analysis could be supplemented with scenario analysis to test different placements of tests in the pathway.
- **Headroom approach**: This provides a maximum reimbursement price and can be used as an initial gauge of the likelihood of a device's viability in the healthcare market. This approach could be used in the second phase of TPP development when the care pathway and test specifications are better defined.
- **Scenario/deterministic sensitivity analysis**: This can be used in the second phase of TPP development to test the factors affecting a test's sensitivity and specificity. For example, it could help explore uncertainty about a test's turnaround time and how this affects

clinical cost-effectiveness.

- **A value of information (VOI) analysis**: This can be undertaken during the final stage of TPP development when the evidence is more mature to highlight the potential benefit of investing in further research to resolve some of the remaining uncertainty.

Diagnostics can go through two routes at NICE: the diagnostics assessment programme or the medical technologies evaluation programme. Clinical guidelines from NICE (and similar organisations) can provide good evidence for the care pathway.

2.5.2. Early Economic Evaluation (EEE) of diagnostic tests:

The way value is generated for diagnostic technologies is more complex than most treatments. The value goes beyond a diagnostic test's accuracy (i.e. sensitivity and specificity).[13] Instead, its main value accrues from identifying patients expected to benefit from distinct treatments,[14] with the test itself having no direct influence on long-term health outcomes.[15]

Therefore, diagnostics is fundamentally distinct from treatments and principles of patient benefit should be incorporated when specifying the new test's decision influence, clinical pathway and value proposition.[14] Ideally, new tests should only be incorporated into clinical practice if there is evidence suggesting a higher probability of improving patient health than with existing tests.[13]

Establishing the clinical care pathway is critical to this process. The clinical care pathway involves the interactions a patient experiences as they move through the health system with a particular medical condition.[16] This is particularly important to establish for diagnostic technologies due to their indirect effect on clinical utility.

However, it can be challenging where no clinical pathway previously exists[17] as the consequences of introducing a new test are difficult to benchmark and highly uncertain. EEE would allow for flexibility where the value of the technology hasn't been fully characterised, enabling exploration of a range of outcomes (both interim and final).[17]

2.5.3. Learnings from Cocco et al. (2020, 2021)[1,10] regarding their implications for EEE

Economic health evidence is often lacking and has been identified as one reason diagnostics fail to achieve market access.[15] Cocco et al. (2020, 2021) have identified areas in the development phase of TPPs where EEE has a role.[1,10]

As discussed earlier in this working document, there are three main development phases of a TPP: scoping, drafting and consensus-building.[1,10] EEE has a different role to play in each phase.

Scoping phase: early economic analysis implications

Early HTA is key in establishing the clinical pathway through care pathway analysis.[10] Establishing the care pathway is particularly important when evaluating medical tests since they do not directly improve patients' health. Instead, they may save costs by diverting resource use or improving downstream health outcomes via a more timely or accurate diagnosis. Currently, oversight of a test's true value and clinical utility is not sufficiently well addressed in the reporting of TPPs for diagnostics,[1] with the main focus being on generating evidence on the analytical performance of a new test. In demonstrating a test's economic value to decision-makers, a highly accurate test won't guarantee improved patient health.[1] The test's downstream effects on clinical decision-making will play a central role. Therefore, this issue must be considered early in technology development.

Scenario analysis can also be performed at this stage by testing different test placements in the clinical pathway to establish where it is likely to be most cost-effective.[10] Although there is potentially a key role for EEE here, this relies on the pathway being well characterised, which may not always be the case and is likely to differ by therapy area.[17]

Drafting and agreement of test specifications: EEE implications

In the next stage, test specifications are drafted and agreed upon at different levels, usually 'desirable' and 'acceptable'.[10] Many factors could impact test specifications, which can be tested through deterministic sensitivity analysis and scenario analysis in EEE. This process involves testing different scenarios to see their impact on cost-effectiveness. For example, it could explore uncertainty in a

test's turnaround time to see how this affects the cost-effectiveness and identify whether it should be a key factor to include in the test specifications (and if any particular level should be recommended). Threshold and headroom analysis could also be undertaken using the willingness to pay (WTP) threshold to back-calculate the maximum costs and minimum specifications for the test to be cost-effective.[10]

The headroom approach provides a connecting analysis thread at different stages, providing a 'rule of thumb' early on, probabilistic analysis through the development process and pricing guidelines in preparation for launching the project.[18] Commercial headroom is the net benefit that the healthcare provider would recognise if the device were supplied to the health service free of charge. This figure is the maximum reimbursement price, i.e. the maximum a manufacturer could charge while securing funding.

Updating TPP specifications: EEE implications

There is no standardised approach to updating EEE-related TPP specifications in the final stage of TPP development, although EEM could be incorporated. The model can be designed flexibly to include evidence as it is updated.[10] It is essential to ensure flexibility, with the initial TPP focusing on defining the unmet need and seeing whether the test could be plausibly cost/clinically effective. The early economic model can be updated to incorporate more precise specifications when evidence becomes more mature.

A value of information (VOI) analysis estimates the value of conducting further research, assuming the information generated would reduce the uncertainty around an estimate.[19] VOI could be used to highlight the potential benefit of resolving the uncertainty that remains in the analysis.[10]

An example of a VOI analysis is calculating the expected value of perfect information (EVPI). An EVPI value that exceeds the cost of conducting further research would mean the research would be considered worthwhile.[20] This provides an informative upper limit threshold for expenditure on further research.

Further sensitivity analysis could also be undertaken to highlight the most sensitive variables (so long as the model is sufficiently comprehensive).[17]

2.5.4. The CanTest framework and the role of EEE

Most diagnostic tests fail due to inadequate performance in real-world settings, where there is often a low disease prevalence (in primary care and general community populations).[12] The CanTest framework aimed to develop a development framework for tests in low prevalence populations, where tests are often applied for triage testing and within a broader diagnostic strategy.[12]

The CanTest framework outlines five phases a new test should follow before integration into routine practice. Phase 3 and Phase 4 both recommend the use of health economic modelling. Phase 3 relates to the **'impact on clinical decision-making and health outcomes.'** Phase 4 defines the **'effectiveness of new diagnostic strategy on clinical outcomes'** and the clinical and cost-effectiveness of the new diagnostic strategy compared to existing methods.

The role of economic evaluation in these stages is clear and has been outlined by the authors.[12] However, EEE arguably has a role in pathway mapping, sensitivity analysis and headroom analysis, which can be performed alongside a test's early development rather than performing an economic evaluation once the evidence is more mature. The data in the early model can be updated as evidence is generated.

Regarding evidence generation, the authors focused on the diagnostic test's accuracy, impact, implementation and cost-effectiveness.[21] These are key factors in the cost-effectiveness of the test and should be performed early in the development.

2.5.5. Diagnostic technology assessment through NICE

Diagnostic technologies can go through two routes at NICE[11] (see Table 2 below for NICE evaluation routes by health outcome):

- The diagnostics assessment programme evaluates diagnostic tests/technologies when such evaluation is complex.[11] This involves diagnostics that have the potential to improve health outcomes but whose introduction is likely to come at a higher cost to the NHS (or similar health outcomes

at less cost or improve health outcomes at an equal cost).

- The medical technologies evaluation programme aims to help the NHS adopt **efficient, cost-saving** medical devices and **simple diagnostics** more rapidly and consistently.[11] A cost-minimisation approach is used to assess products, considering the cost/resource consequences resulting from or associated with the technology under evaluation.

Table 2. NICE evaluation routes

Health outcomes	Cost	NICE evaluation route
Improved	Higher	DAP
Similar/same	Lower	DAP
Improved	Similar/ same	DAP
Improved (or similar/same)	Lower	Med Tech Evaluation*

*This means that the technology can go through the med-tech evaluation route, but it does not necessarily mean it will.

Evidence requirements

For diagnostic technologies, end-to-end studies are preferred.[11] When there is insufficient evidence from these studies, a linked-evidence approach should be taken that combines evidence from different study designs, data sources or methodologies. For diagnostic evaluations, linked-evidence modelling is usually needed to measure and value health effects because end-to-end controlled trials with follow-up through the care pathway are uncommon.[11]

Clinical expert opinion or expert elicitation is likely important to help resolve the uncertainty associated with the clinical pathway for a diagnostic technology.[11] These expert opinions can also help inform necessary sensitivity/ scenario analyses.

Clinical guidelines from NICE (and similar organisations) can provide good evidence for the care pathway.[11] Diagnostic before-and-after studies also provide useful information on any change in management after introducing an index test to clinical practice.

Methodological approaches:

QALY is the length and quality of life as measurements of the value of health outcomes for people. For diagnostics, a QALY weight for severity based on absolute (i.e. total future health loss due to condition) and proportional (i.e. proportion of future health loss relative to remaining life expectancy) QALY shortfall is unlikely to reflect the societal value and severity of disease in a way that is relevant to the diagnostics context.[11] Therefore, the severity modifier will not usually be applicable in diagnostic evaluations.

The value of benefits that may indirectly affect health, e.g. a diagnostic improving the speed of correct diagnosis, should be considered.[11]

The analysis should include all relevant patient outcomes that change in the care pathway due to the diagnostic test or sequence of tests.[11] The nature, severity, time and frequency of occurrence and duration of the outcome may all be important in determining the effect on quality of life and should be considered.

References

1 Cocco, P., Ayaz-Shah, A., Messenger, M. P., West, R. M. & Shinkins, B. 2020. 'Target Product Profiles for medical tests: a systematic review of current methods'. *BMC medicine* **18**, 1-12.

2 Dailey, P. J. *et al.* 2019. 'Defining System Requirements for Simplified Blood Culture to Enable Widespread Use in Resource-Limited Settings'. *Diagnostics* **9**, 10.

3 Vetter, B. *et al.* 2021. 'Development of a target product profile for a point-of-care cardiometabolic device'. *BMC Cardiovasc Disord* **21**, 486 (2021). As of 22 March 2024: https://doi.org/10.1186/s12872-021-02298-7

4 Program for Appropriate Technology in Health. 2018. *Diagnostics Instrument* –Target Product Profile. Diagnostic Instrument: Hemoglobinometer. Seattle, WA, USA.

5 Pellé, K. G. *et al.* 2020. 'Electronic clinical decision support algorithms incorporating point-of-care diagnostic tests in low-resource settings: a target product profile'. *BMJ Global Health* **5**, e002067 (2020). As of 22 March 2024: https://doi.org/10.1136/bmjgh-2019-002067

6 World Health Organization, FIND and Medecins sans frontieres. 2020. *A Multiplex multi-analyte diagnostic platfor*m.

7 Mather, R. G., Hopkins, H., Parry, C. M. & Dittrich, S. 2019. 'Redefining typhoid diagnosis: what would an improved test need to look like?' *BMJ Global Health* **4**, e001831.

8 Kadam, R. *et al.* 2020. 'Target Product Profile for a mobile app to read rapid diagnostic tests to strengthen infectious disease surveillance'. *PLoS One* **15**, e0228311.

9 World Health Organization. *Target product profiles for antibacterial resistance diagnostics.* World Health Organization.

10 Cocco, P., Messenger, M. P., Smith, A. F., West, R. M. & Shinkins, B. 2021. 'Integrating Early Economic Evaluation into Target Product Profile development for medical tests: advantages and potential applications'. *Int J Technol Assess* **37**. As of 22 March 2024: https://doi.org/10.1017/s0266462321000374

11 National Institute for Health and Care Excellence. 2023. *NICE health technology evaluations: the manual*. As of 22 March 2024: https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741

12 Walter, F. M. *et al*. 2019. Evaluating diagnostic strategies for early detection of cancer: the CanTest framework. *Bmc Cancer* **19**, 1-11.

13 Ferrante di Ruffano, L., Hyde, C. J., McCaffery, K. J., Bossuyt, P. M. & Deeks, J. J. 2012. 'Assessing the value of diagnostic tests: a framework for designing and evaluating trials'. *BMJ* **344**, e686. As of 22 March 2024: https://doi.org/10.1136/bmj.e686

14 Soares, M. O., Walker, S., Palmer, S. J. & Sculpher, M. J. 2018. 'Establishing the Value of Diagnostic and Prognostic Tests in Health Technology Assessment'. *Med Decis Making* **38**, 495-508. As of 22 March 2024: https://doi.org/10.1177/0272989X17749829

15 Oosterhoff, M., van der Maas, M. E. & Steuten, L. M. 2016. 'A Systematic Review of Health Economic Evaluations of Diagnostic Biomarkers'. *Appl Health Econ Health Policy* **14**, 51-65. As of 22 March 2024: https://doi.org/10.1007/s40258-015-0198-x

16 NIHR Newcastle In Vitro Diagnostics Co-operative. 2023. *Care pathway analysis to identify the value proposition.* As of 22 March 2024: https://newcastle.mic.nihr.ac.uk/expertise/evaluation-themes/early-stage-evidence/care-pathway-analysis/

17 Abel, L. *et al.* 2019. 'Early Economic Evaluation of Diagnostic Technologies: Experiences of the NIHR Diagnostic Evidence Co-operatives'. *Med Decis Making* **39**, 857-866. As of 22 March 2024: https://doi.org/10.1177/0272989X19866415

18 Girling, A., Lilford, R., Cole, A. & Young, T. 2015. 'Headroom Approach to Device Development: Current and Future Directions'. *Int J Technol Assess Health Care* **31**, 331-338. As of 22 March 2024: https://doi.org/10.1017/S0266462315000501

19 Fenwick, E. *et al.* 2020. 'Value of Information Analysis for Research Decisions-An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force'. *Value Health* **23**, 139-150. As of 22 March 2024: https://doi.org/10.1016/j.jval.2020.01.001

20 Thorn, J., Coast, J. & Andronis, L. 2016. 'Interpretation of the Expected Value of Perfect Information and Research Recommendations: A Systematic Review and Empirical Investigation'. *Med Decis Making* **36**, 285-295. As of 22 March 2024: https://doi.org/10.1177/0272989X15586552

21 Verbakel, J. Y. *et al.* 2017. 'Common evidence gaps in point-of-care diagnostic test evaluation: a review of horizon scan reports'. *BMJ Open* **7**, e015760. As of 22 March 2024: https://doi.org/10.1136/bmjopen-2016-015760

Annex B: Scoping Document

Authors: Mark L Cabling, Jessica Dawney, Matthew Napier, Zuzanna Marciniak-Nuqui, Fifi Olumogba, Larry Kessler, Amanda Cole, Lotte Steuten, Sonja Marjanovic

1. Analysis of individual features and categories in Cocco et al. (2020)

Annex B is the second of eight annexes complementing the main Cancer Research UK-funded project's final report: 'Advancing the development and use of diagnostic target product profiles for cancer.' The not-for-profit research institute RAND Europe led the project in collaboration with the Office of Health Economics. The project has benefited from ongoing support and advice from Professor Larry Kessler (University of Washington), a key consultant on the work. This document provides detailed findings and analysis of the scoping desk research and preliminary stakeholder consultation to which the final report refers; thus, Annex B, like all the other annexes, is primarily meant to accompany the final report and is not meant to be read as a standalone document.

In the contents that follow, we reflect on the features and categories covered in the Cocco et al. (2020) review (in supporting material) to identify common features across TPPs and examine where there is some scope for streamlining individual features and resolving duplication related to diverse terminology used to describe conceptually similar (or the same) features in different TPPs. The aim is to arrive at a somewhat shorter and more manageable list of features to consult with stakeholders on as part of further work packages on this project where feasible. We have arrived at these decisions in consultation with the research team and Professor Larry Kessler. We also considered the most logical flow of features in specific categories, mainly sticking to the clustering done by Cocco et al. (2020) but adapting one category (the 'infrastructure' category) as we felt it offered a more natural flow for this project's purposes.

We also provide 'working explanations' for features to enable effective stakeholder engagement in the future (given that the Cocco et al. (2020) review does not provide descriptions of individual feature labels). To do so, we drew on desk research, consultation within the project team and comments from some advisors. For less clear features, we checked for clarity on the Clinical and Laboratory Standards Institute (CLSI) Harmonized Terminology Database.[1] However, only some terms feature in the database, which is not specifically TPP-focused. Note that we do not intend the explanations to be formal definitions.

Individual TPP development efforts will want to consider the granularity level associated with any feature. Therefore, some of the features in the lists below may warrant splitting into individual features in specific TPP development efforts (e.g. separating infrastructure features associated with physical equipment from those associated with supplies/consumables).

Given that some features appear in multiple categories, individual TPP efforts may want to consider which overarching category a feature fits best (which may vary by test type).

This analysis is based on the contents described above. It is important to note that some features may only relate to some test types (e.g. sample/specimen volume for IVDs) **and that some features may not be present due to a lack of evidence** (e.g. some features relevant for imaging tests that may need adding to future bespoke TPP efforts).

Annex G provides working explanations for individual features, to which we refer the reader for further information.

2. Unmet Clinical Need

Commonly used features in this category across TPPs

This category covers features related to the unmet need a diagnostic test responds to, understood as the information that helps define and specify the clinical need for the new product and its scope of application. Cocco et al. (2020) listed a total of ten features in this category based on features covered in the TPPs that informed their review (see Figure 1). Of these, four featured in 50% or more of the TPPs reviewed by Cocco et al. (2020):

- Intended use (98%)
- Target level of health system (86%)
- Target population (75%)
- Target user (75%).

Scope for streamlining

There are also some features currently appearing as distinct in the Cocco et al. (2020) database that could potentially be collapsed into one overarching feature because of their conceptual proximity and similarities, even though the terminology used to describe the features has varied across TPPs.

Bespoke guidance in any specific TPP document to be developed would need to provide information on aspects of a feature relevant to a particular test.

As part of an effort to streamline features to inform stakeholder consultations that took place in further work packages, we suggested that there is scope to combine 'medical need' and 'test rationale' into one category with the working label 'medical need'.

Figure A1. Features listed under the 'Unmet Clinical Need' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs that mentioned each feature

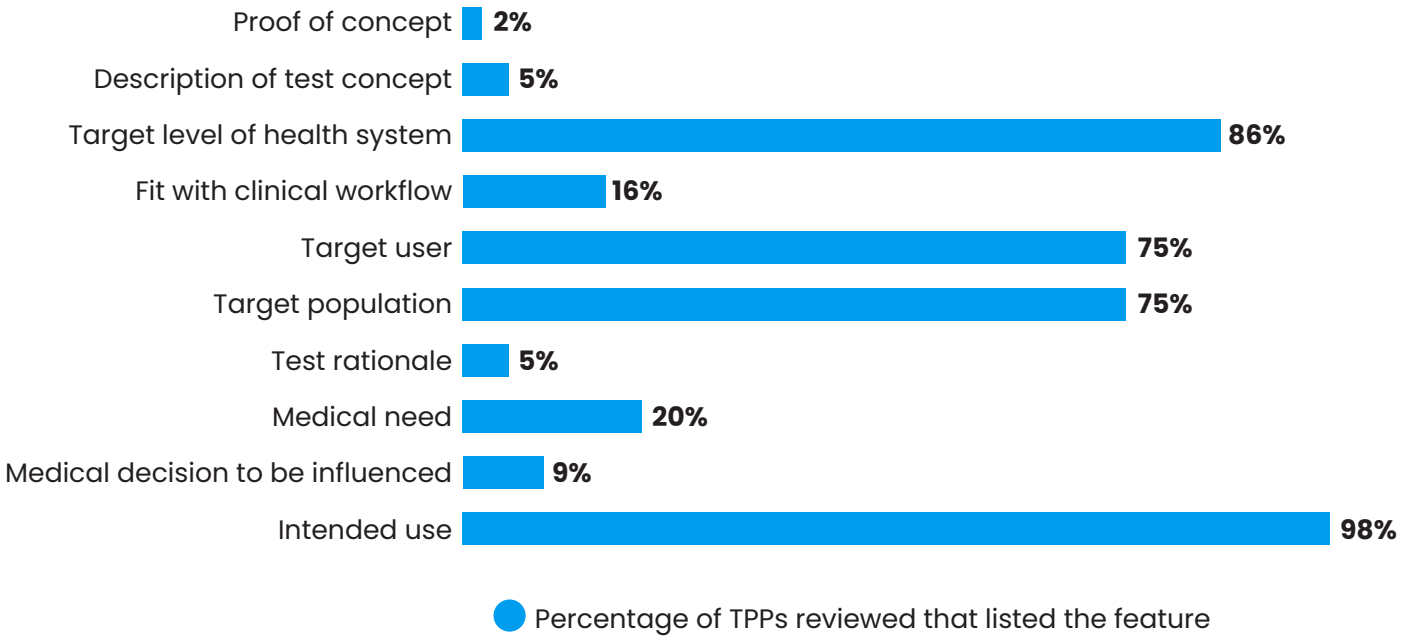


Table 1. Features listed under the 'Unmet Clinical Need' category in the Cocco et al. (2020) review

Feature (*denotes common features)	Explanations
Intended use*	At the highest level, this feature focuses on a test's purpose. According to the MHRA, it provides information on the 'function and intended use of the product, that is, whether the product 'may be administered….with a view to' achieving a medicinal purpose.' This should include information on what the test will be used for. Though by no means comprehensive, illustrative examples include screening (i.e. actively seeking/conducting detection processes in asymptomatic populations to support early diagnosis and timely treatment) and surveillance (e.g. monitoring people or hazards to support prevention or tracking people with a condition to see how it progresses).[2] Tests may also serve specific aims, such as supporting earlier or more accurate diagnosis of a disease stage. However, these are only illustrative examples.
Medical decision(s) to be influenced	Information on the type of medical decision(s) the test can help with, i.e. how the test aims to influence medical decision-making and how it will achieve its medicinal purpose. Examples include deciding whether a person has a disease or not or prioritising patients for further care pathway routes (e.g. whether to refer to secondary care or put on an urgent treatment list).
Medical need**	Information on why the test is needed and what unmet need it addresses, potentially including the scale of need and the rationale for developing a novel test given the field/science's readiness (e.g. relevant recent technological/scientific developments related to the test's feasibility and potential success). **A combination of two features listed in Cocco et al. (2020) – likely due to differing terminology between different TPPs ('medical need' and 'test rationale').
Target population*	Information on who the diagnostic will test (i.e. the eligible population, based on diverse features such as risk factors, symptoms, demographics, etc).
Target user*	Information on who will administer the test, e.g. a lab technician, a nurse or a patient alone at home. This may involve more than one person, e.g. one might collect the sample, and another conduct the test.
Fit with clinical workflow	Information about how the test fits into existing patient-care pathways and processes in the health system and whether it disrupts them/changes them in significant ways, positively and/or negatively, e.g. leads to a patient needing to see different types of healthcare professionals than in existing care pathways, means that healthcare professionals would need to assume new roles/responsibilities, or disrupts the care pathway by removing some patients from the system, etc.

Feature (*denotes common features)	Explanations
Health system target level*	Information about whether the test is designed for use in specific settings, e.g. primary care, acute care, community care or at home.
Description of the test concept	Information on the possible range of technologies that could help accomplish the TPP's aims, e.g. specific genomic sequencing of lab blood tests for chemical analysis. Alternatively, the TPP's developers may only present a high-level concept, such as, 'the test should rely on X type of biological sample' for which they cannot specify the exact analysis to undertake.
Proof of concept	Information on the test's scientific proof of concept, i.e. evidence demonstrating its feasibility based on prior research, trials, etc. For example, this feature might specify what evidence constitutes acceptable proof of concept, such as whether a randomised trial is necessary to demonstrate product efficacy. Related to this: • One advisor/consultant felt this might fit better in another category, such as analytical performance. However, we believe it works because it provides early, upfront signals of what constitutes acceptable evidence regarding, for example, underlying study designs and patient samples. • Another felt that it does not fit here or that the term is inappropriate, noting that while TPPs can guide the evidence needed to demonstrate clinical and cost-effectiveness as a separate exercise, proof of concept does not describe this well.

3. Analytical Performance

Commonly used features in this category across TPPs

This category covers features related to a test's ability to correctly detect and measure a disease analyte/marker, providing information that can support assessments of whether the test accurately measures what needs to be measured. Alongside 'output' features of analytical performance, it also covers features related to the requirements for appropriate analytical performance.

Cocco et al. (2020) listed 44 features[i] in this category (see Figure 2), three of which featured in 50% or more of the TPPs they reviewed. These were:

- Sample type (95%)
- Time to test result (86%)
- Manual sample/specimen preparation (77%).

Scope for streamlining

Some features that currently appear as distinct in the Cocco et al. (2020) database could be combined into one feature because of their conceptual proximity and similarities. As part of an effort to streamline features to inform stakeholder consultations in further work packages, we suggested a scope for combining the following into single (broader) features:

- 'Analytical sensitivity' and 'limit of quantification/detection'.
- 'Reproducibility' and 'reproducibility near clinical threshold'.
- 'Assay throughput', 'daily throughput (per module)', 'platform throughput' and 'sample/specimen capacity and throughput'.
- 'Robustness' and 'interferences'.
- 'Overall sample preparation', 'need for operator to transfer a precise volume of sample' and 'manual sample/specimen' preparation.
- 'Quality control', 'internal quality control' and 'external quality control'.

i Twelve features were listed under this category, with two listed twice: 'positive predictive value' and 'negative predictive value'. We have deleted the repetitions, leaving ten features in this category.

Figure A2. Features listed under the 'Analytical Performance' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs mentioning each feature

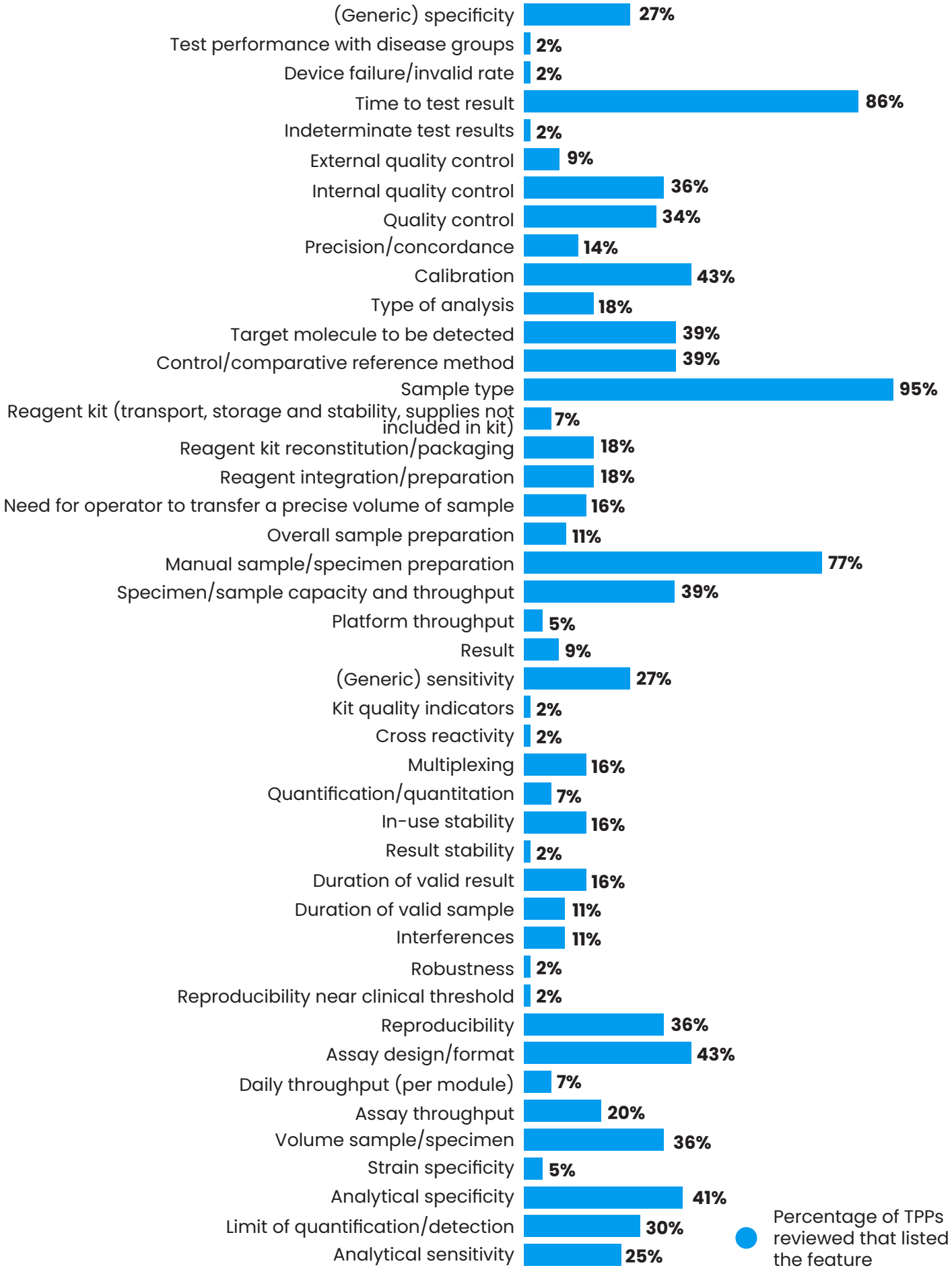


Table 2. Features listed under the 'Analytical Performance' category in the Cocco et al. (2020) review, with working explanations

Feature (*denotes common features)	Explanation
KEY TEST FEATURES AND ANALYTICAL PERFORMANCE INFORMATION	
Assay/test design and format	Information about the type of diagnostic test, e.g. molecular, serologic, antigen-detection, etc. Components of the test kit could also be considered to fall here, as no other feature describes a test's core components. However, no advisor commented on this aspect, and there is no definition for this term in the Clinical and Laboratory Standards Institute (CLSI) database or the Harmonized Terminology Database.
The target molecule for detection	Information about what specific molecule the test aims to detect.
Analytical specificity	Information about the test's ability to detect and distinguish the intended disease analyte/marker
Analytical sensitivity/limit of detection	Information about the smallest amount of substance necessary in a sample for accurate disease detection (a test's ability to detect low concentrations of the target analyte/marker). Armbruster and Pry (2008) note that 'Analytical Sensitivity, Functional Sensitivity, Lower Limit of Detection, Limit of Blank, Limit of Detection and Limit of Quantitation are terms used to describe the smallest concentration of a measurand that can be reliably measured by an analytical procedure. There has often been a lack of agreement within the clinical laboratory field as to the terminology best suited to describe this parameter.'[3] Definitions of analytical sensitivity and limit of quantification/detection are similar. For example, *analytical sensitivity* is defined as 'the smallest amount of substance in a sample that can accurately be measured by an assay,'[4] while the *limit of quantification* is described as 'the lowest analyte concentration that can be quantitatively detected with a stated accuracy and precision.'[5]
Strain specificity	Information about whether a test can accurately detect specific strains/ distinguish between different strains of a disease agent.
Cross-reactivity	In the most general sense, this refers to the extent to which different disease markers appear similar in the test's results (e.g. can be a source of false positives).

Feature (*denotes common features)	Explanation
Reproducibility*,**	Information on whether a test replicates the exact same result from repeated testing of the same person, reflecting 'the extent of agreement of a single person (observer) or different observers using the same diagnostic procedure in the same subject.'[6] Different types of reproducibility must be considered and nuanced in TPP development. For example, there may be an interest in reproducibility across test settings or whether reproducibility is possible near the clinical threshold. However, none of the consulted advisors commented on this point. **A combination of two features listed in Cocco et al. (2020): 'reproducibility' and 'reproducibility near clinical threshold.'
Robustness and interferences*,**	Information about the extent to which a test will be unaffected by changes in test conditions[7] as an indicator of its robustness, alongside information on when a test result may be falsely altered (interferences), which can compromise robustness. As Aubry and Weng (2015) note: 'a robust assay is one that will remain 'unaffected by small but deliberate changes in test conditions.'[7] Dimeski (2008) define interference as 'occurring when a substance or process falsely alters an assay result. Endogenous interference originates from substances present in the patient's own specimen. Exogenous interferences are substances introduced into the patient's specimen.'[8] **A combination of 'Robustness' and 'Interferences' from Cocco et al. (2020)
Control/ comparative reference method	Information about other test/ reference-method types against which the new test's performance must compare favourably, e.g. existing related tests and diagnostic methods.
Precision/ concordance	We understand concordance to be about how much the test results agree with the results of other tests applied to the same sample/ individual.[9] It is closely related to notions of validity of a novel diagnostic technique, which is related to a test's accuracy and reproducibility. This feature provides information about how the test or analysis's results compare against a recognised gold standard (where possible). Note: This feature also appears under the 'clinical validity' category, where it may be better placed. However, one consulted expert felt it belonged here.
Indeterminate test results	Information about whether a test can and does give invalid results and how/why that might occur. Some tests can provide neither positive nor negative results but inconclusive ones.

Feature (*denotes common features)	Explanation
Device failure/ invalid rate	Information referring to the cases/conditions/rates where the diagnostic fails to give a result. This differs from an indeterminate result, e.g. only showing part of a result.
Test performance with disease groups	Information about the test's performance for different patient profiles (e.g., patients with the disease but at different stages or severities or patients from different demographic groups).
Duration of valid sample	Information on how long the sample can be used once collected (e.g. blood, saliva, etc).
Time to test result*	Information on how long it takes from doing the test to getting the result.
Duration of valid result	Information on how long the reading is valid, e.g. once you get a result, how long will it remain accurate – not fade or otherwise change? This could apply to whether the test is stable or whether the disease changes over time, i.e. the result may not apply in X period. None of the advisors we consulted commented on or clarified this aspect. According to the advice sought, this feature describes two distinct properties which need separating. Test stability depends on the type of test, while how quickly the disease changes over time is relevant to all tests.
Result stability	We have understood this to mean diagnostic result stability, often defined as 'the degree to which a diagnosis is confirmed at subsequent assessments'.[10] Note: this feature is related to 'duration of valid result' but potentially concerns more than just time, covering information on other factors that might affect the stability of the result, e.g. environmental features or conditions affecting the test result, such as the room's temperature.
ANALYTICAL PERFORMANCE: OPERATIONAL AND ANALYTIC REQUIREMENTS	
Sample type*	Information on the sample type, e.g. blood, saliva, urine or other tissue.
Volume sample/ specimen	Information about the amount of specimen needed (e.g. volume of blood or saliva) for testing.
Sample/specimen preparation*,**	Information on the steps, processes and conditions involved in preparing the sample for testing, including a description of whether and what specific amount is needed. **A combination of three features listed in Cocco et al. (2020): 'overall sample preparation', 'need for the operator to transfer a precise volume of sample' and 'manual sample/specimen preparation.
Throughput	Meaningful throughput dimensions will vary across tests, with each specific test type needing bespoke consideration. This includes information on how many tests can completed in a specific period. Depending on the test type, this may depend on different aspects of throughput, such as (a) the specific platform tests are run on, (b) human capacity, (c) sample/specimen throughput and capacity, and (d) all of these aspects. *A combination of three features listed in Cocco et al. (2020): 'assay throughput', 'daily throughput (per module)' and 'platform throughput' and 'sample/specimen capacity and throughput.'

Feature (*denotes common features)	Explanation
In-use stability	Information on whether the material(s) the diagnostic test uses (e.g. reagents) remain(s) stable during use and produce an accurate result.
Type of analysis	Information on the analysis type to be conducted (e.g. qualitative or quantitative).
Quantification/ quantitation	Information on whether the test provides a quantitative measure/result, e.g. disease load, spread or severity.
Result	Information on how the test result is conveyed/displayed. Note: This information is repeated in the 'Human Factors' category, which may be unnecessary.
Multiplexing	Information on the process of detecting or identifying multiple biomarkers within a single diagnostic test.
Kit quality indicators	Information on test kit elements that indicate any degradation in test components' ability to do their job, e.g. where environmental conditions such as heat may affect the test's components, it is essential to have indicators that tell the testers or lab there may be a problem.
Reagent integration/ preparation**	Information on how to prepare and package the reagent(s) for use with the test. **A combination of two features listed in Cocco et al. (2020): 'reagent integration/preparation' and 'reagent kit reconstitution/packaging '.
Reagent kit (nature, transport, storage and stability, supplies not included in the kit)	Information on the nature of the reagent kit and how it should be transported, stored and kept stable, plus information on which supplies are not included in the kit and must be acquired externally. Note: this feature is also covered under 'Infrastructure'; it is probably unnecessary in both categories and might be better under one or the other.
Calibration	Information on how to set up the test so it is correctly calibrated for accurate use.
Quality control (internal and/or external)**	Information about quality control procedures and requirements associated with the test's use. This feature can cover different types of quality control, such as internal and external quality control, depending on the test type and context. Internal quality control procedures will likely refer to the method used, personnel and instruments, e.g. using controls in the lab. We did not know the difference between internal and external quality control and received no comments on this. **A combination of three features listed in Cocco et al. (2020): 'quality control', 'internal quality control' and 'external quality control' (not all TPPs cover internal and external quality control).

4. Clinical Utility

Commonly used features in this category across TPPs

This category covers features related to whether a test will positively affect intended outcomes, such as patient quality of life and longer lifespan, and how it would seek information on direct or indirect contributions to intended outcomes. Cocco et al. (2020) listed features in this category (see Figure 3), none of which featured in 50% or more of the TPPs they reviewed.

Scope for streamlining

None.

Figure A3. Features listed under the 'Clinical Utility' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs that mentioned each feature

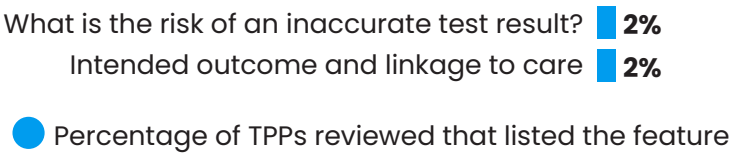


Table 3. Features listed under the Clinical Utility category in the Cocco et al. (2020) review, with working explanations

Feature	Explanations
Intended outcome and linkage to care	Information about the patient outcome a test will contribute to and how it will likely link to care pathways, e.g. is it intended to enable faster or earlier diagnosis or better triage, and will it ultimately support outcomes such as longer life or better quality of life? As mentioned by one expert we consulted, devising a narrative for this feature may involve a step-wise consideration, i.e. the decision the test impacts and how the affected decision-making impacts outcomes. It is vital to clarify outcomes of interest and the 'size' of improvement expected (where possible to articulate).
What is the risk of an inaccurate test result?	Information on the types of conditions (e.g. human, sample, operational or environment) that could engender inaccurate results and, if possible, information on the risk level for patients/results associated with different risk factors. Note: This indicator may fit better under the 'Clinical Validity' category (where it is currently repeated), as agreed by one advisor.

5. Clinical Validity

Commonly used features in this category across TPPs

This category covers features related to whether what is being measured correlates appropriately with a physiological condition, pathological process or state, i.e. is the test measuring an appropriate marker of the disease? Cocco et al. (2020) listed a total of 11 features in this category (see Figure 4), two of which featured in 50% or more of the TPPs they reviewed:

- Diagnostic/testing sensitivity (70%)
- Diagnostic/testing specificity (64%).

Scope for streamlining

None.

Figure A4. Features listed under the 'Clinical Validity' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs that mentioned each feature

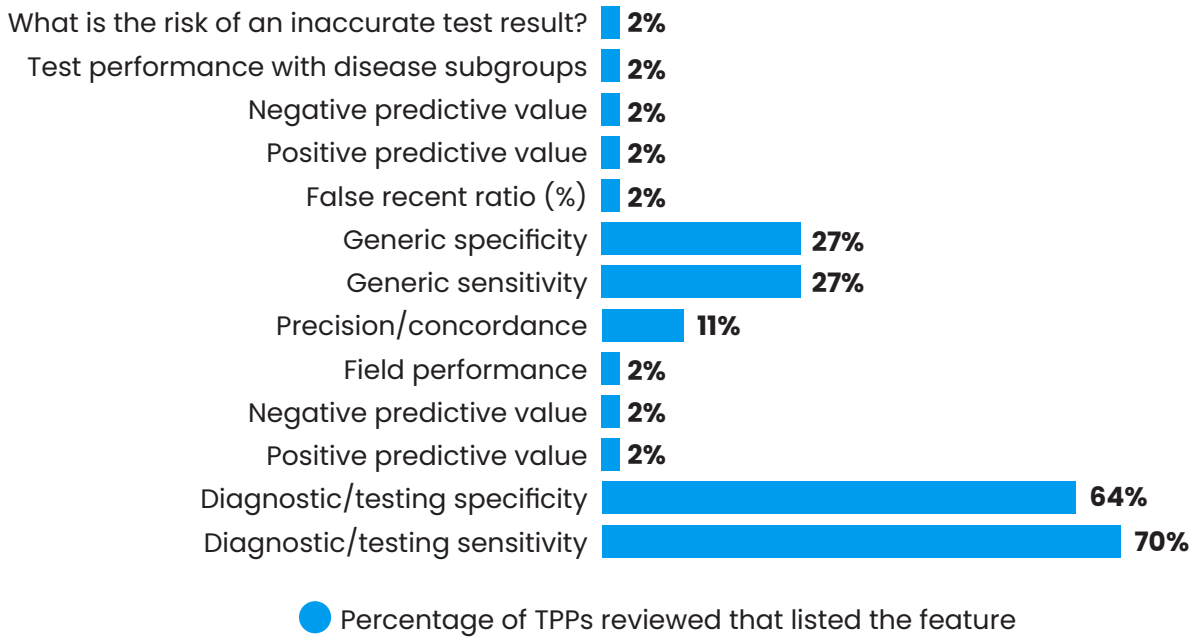


Table 4. Features listed under the 'Clinical Validity' category in the Cocco et al. (2020) review, with working explanations

Feature (*denotes common features)	Explanations
Diagnostic/testing sensitivity*	Information on diagnostic sensitivity is related to a test's ability to correctly identify individuals with the disease (i.e. without false negatives); it is the probability of a positive diagnostic test in a person with the illness.

Feature (*denotes common features)	Explanations
Diagnostic/testing specificity*	Information on the diagnostic specificity summarises the probability of a negative diagnostic test in an individual who does not have the illness, i.e. a test's ability to correctly rule out those without the disease (without false positives).
Positive predictive value	Positive predictive value (PPV) is the probability of a confirmed diagnosis among those with a positive test. PPV and negative predictive value depend on the prevalence of the disease tested for, whereas sensitivity and specificity are invariant concerning prevalence.
Negative predictive value	Negative predictive value is the probability of a confirmed diagnosis among those with a negative test result. PPV and negative predictive value depend on the prevalence of the disease tested for, whereas sensitivity and specificity are invariant concerning prevalence.
Field performance	Information on how well the test performs in the real world rather than a laboratory/controlled environment.
Precision/ concordance	We understand concordance as how much the test results agree with the results of other tests applied to the same sample/individual.[9] It is closely related to notions of validity of a novel diagnostic technique, which is related to a test's accuracy and reproducibility. This feature describes how the test's results or analysis compare against a recognised gold standard (where this exists). Note: one advisor commented that this feature fits better under the 'analytical performance' category.
False recent ratio (%)	Information about the proportion of diagnosed cases falsely classified/ misclassified as recent.
Test performance with disease subgroups	Information about a test's performance in different patient-profile groups (i.e. patients with different stages/severities of the disease or severities in different demographics). This feature provides information on performance across certain demographic groups.
What is the risk of an inaccurate test results?	Information on the conditions (e.g. human, sample, operational or environmental) that could cause inaccurate results and, if possible, the risk level to patients/results associated with different risk factors.

6. Human Factors

Commonly used features in this category across TPPs

This category covers features related to people's roles in healthcare provision, e.g. the effects of teamwork, tasks, equipment, workspace, culture and organisation on human behaviour and abilities and applying that knowledge in clinical settings. **Within a TPP/diagnostic test specification context, this feature concerns how individuals must interact with the test (e.g. when administering it to a patient, handling it, preparing it for use, interpreting and capturing results and any training needs)**. Cocco et al. (2020) listed a total of 31 features in this category (see Figure 5), of which only one featured in 50% or more of the TPPs they reviewed:

- Training and education (75%).

Scope for streamlining

Some features that currently appear as distinct in the Cocco et al. (2020) database could potentially be collapsed into one feature because of their conceptual proximity and similarities. A specific TPP's bespoke guidance would provide information on a feature's aspects relevant to a particular test. The features that can be combined into one are:

- 'Materials used' and 'supplies needed'.
- 'Result', 'Readout/reading system', 'Result documentation-data display' and 'Test outcome (nature)'.
- 'Unit size', 'Test size and weight' and 'Test size and portability'.
- 'Data analysis', 'Rate of errors in device interpretation' and 'Ease of test result interpretation'.

Figure A5. Features listed under the 'Human Factors' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs that mentioned each feature

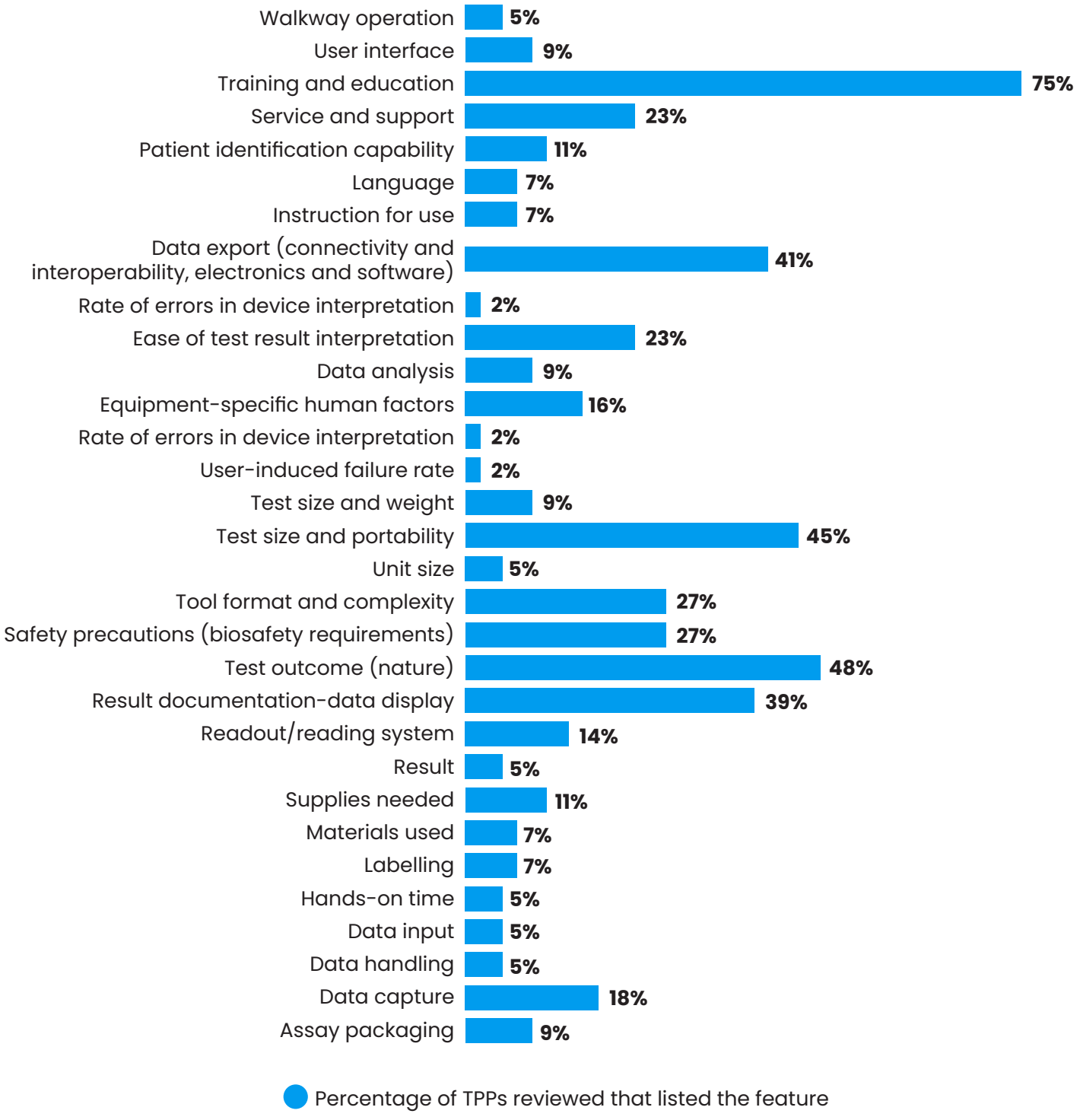


Table 5. Features listed under the 'Human Factors' category in the Cocco et al. (2020) review, along with working explanations

Feature (*denotes common features)	Explanations
GENERAL USE RELATED	
Assay packaging	Information related to assay packaging specifications, including safety considerations (e.g. safety seals to confirm the contents are authentic and have not been tampered with).
Supplies needed*,**	Information about the supplies needed for the medical test/product use. **A combination of two features listed in Cocco et al. (2020): 'materials used' and 'supplies needed'.
Test size and portability*,**	Information on the physical size and weight of the end product. This feature could include a medical test's sensitivity to external factors such as temperature, movement, storage and dampness; its portability and related implications for human use; and its size (included under 'human factors' because of the device's usability and portability). **A combination of three features listed in Cocco et al. (2020): 'unit size', 'test size and weight' and 'test size and portability'.
Equipment-specific human factors	Information about how an individual engages with the test and what they must do to correctly operate the associated equipment.
Patient identification capability	Information on a test's in-product capability for patient identification, i.e. correctly matching a patient to appropriate intervention/tests and communicating information about the patient's identity accurately and reliably throughout the care continuum. This includes the potential to add a patient/user code to ensure the right person is being tested, akin to the radio frequency identification codes or COVID tests.
Safety precautions (biosafety requirements)	Information about the biosafety requirements that must be in place for a test's safe operation.
Service and support	Information on the nature and level of human support and servicing needed to use and maintain a test. Note: this feature is also under the 'Infrastructure Requirements' category. While it can probably stay under 'Human Factors', it is not necessarily needed in both categories.
Use-induced failure rate	Information about the rate of human error in operating the device/product.
Ease of test result interpretation**	Information on the nature of human involvement in interpreting/analysing the test results, e.g. the analysis type and its complexity/simplicity). This can include information on the human error rate in interpreting test results, although this already exists as a separate feature. **A combination of three features listed in Cocco et al. (2020): 'data analysis', 'rate of errors in device interpretation' and 'ease of test result interpretation'.
Rate of errors in device interpretation	Information on the rate of human error in interpreting test results.

Feature (*denotes common features)	Explanations
TEST OPERATION RELATED (OTHER THAN DATA)	
Training and education*	The type of training and education the human user needs to have to effectively engage with any aspect of the test, whether preparing the kit, conducting sample collection from the patient, administering the test to the patient or interpreting the results.
Tool format and complexity	Information about the product's complexity level and how specialised the user has to be to use it correctly. This can include information about the nature of associated skill needs.
Hands-on time	Information about the time needed to conduct the test.
Labelling	Information on designing, reviewing, producing and attaching labels for the test.
Walkway operation	Information on whether the user must supervise assays closely or – as in the case of cultures, for example – can start the assay and leave it for hours/days before returning to complete it.
Instruction for use	Information on how the user should operate and use the test/test kit.
User interface	Information on the nature of the platform through which the test user(s) inputs various data (e.g. the patient ID, test time/date/location and results), such as a computer or portable tablet.
Language	Information on the language in which the device/test is programmed to operate, if relevant, or the language of the instructions.
DATA-RELATED	
Data capture	Information on what data to capture (which may be results-related but potentially includes the test time/date and patient information) and where/how to capture it.
Data handling	Information on gathering, recording and presenting information relevant to the test (test results or data referring to patients).
Data input	Information on how the user should input data into a product.
Data export (connectivity and interoperability, electronics and software)	Information on how the user should export the product's data, including information about connectivity, interoperability, electronics and software.
Result format and readout*,**	Information on how the test result is conveyed and displayed, i.e. the data types, outputs, display methods, overall system and how the system records the result. Note: This information is repeated in the 'Analytical performance' category and may not need to be included here. **A combination of four features listed in Cocco et al. (2020): 'Result', 'Readout/reading system', 'Result documentation-data display' and 'Test outcome (nature).'

7. Infrastructural Requirements

Commonly used features in this category across TPPs

This category covers features about any infrastructure-related requirements (such as facilities and equipment) or other operating conditions that must be established and maintained for the effective transport, storage, operation/use and/or disposal of the test. Cocco et al. (2020) listed 26 features in this category (see Figure 6), of which five featured in 50% or more of the TPPs they reviewed:

- Storage conditions and shelf life (70%)
- Temperature and humidity (61%)
- Power requirements (52%)
- Stability during transport (52%)
- Waste disposal (50%).

Scope for streamlining

Some features that currently appear separately in the Cocco et al. (2020) database could potentially be combined into one feature because of their conceptual proximity and similarities. A specific TPP's bespoke guidance would provide information on a feature's aspects relevant to a particular test. The feature groups that could be combined into one are:

- 'Instrument-infrastructural requirement', 'infrastructural requirement', 'supplies needed', 'need for additional equipment/ test/spare parts', 'ancillary supplies, and 'additional third-party consumables' and 'materials used'.
- 'Maintenance', 'service and support' and 'external maintenance'.
- 'Storage conditions and shelf life' and 'storage conditions prior utilisation'.
- 'Shipping conditions' and 'stability during transport'.

Figure A6. Features listed under the 'Infrastructural Requirements' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs that mentioned each feature

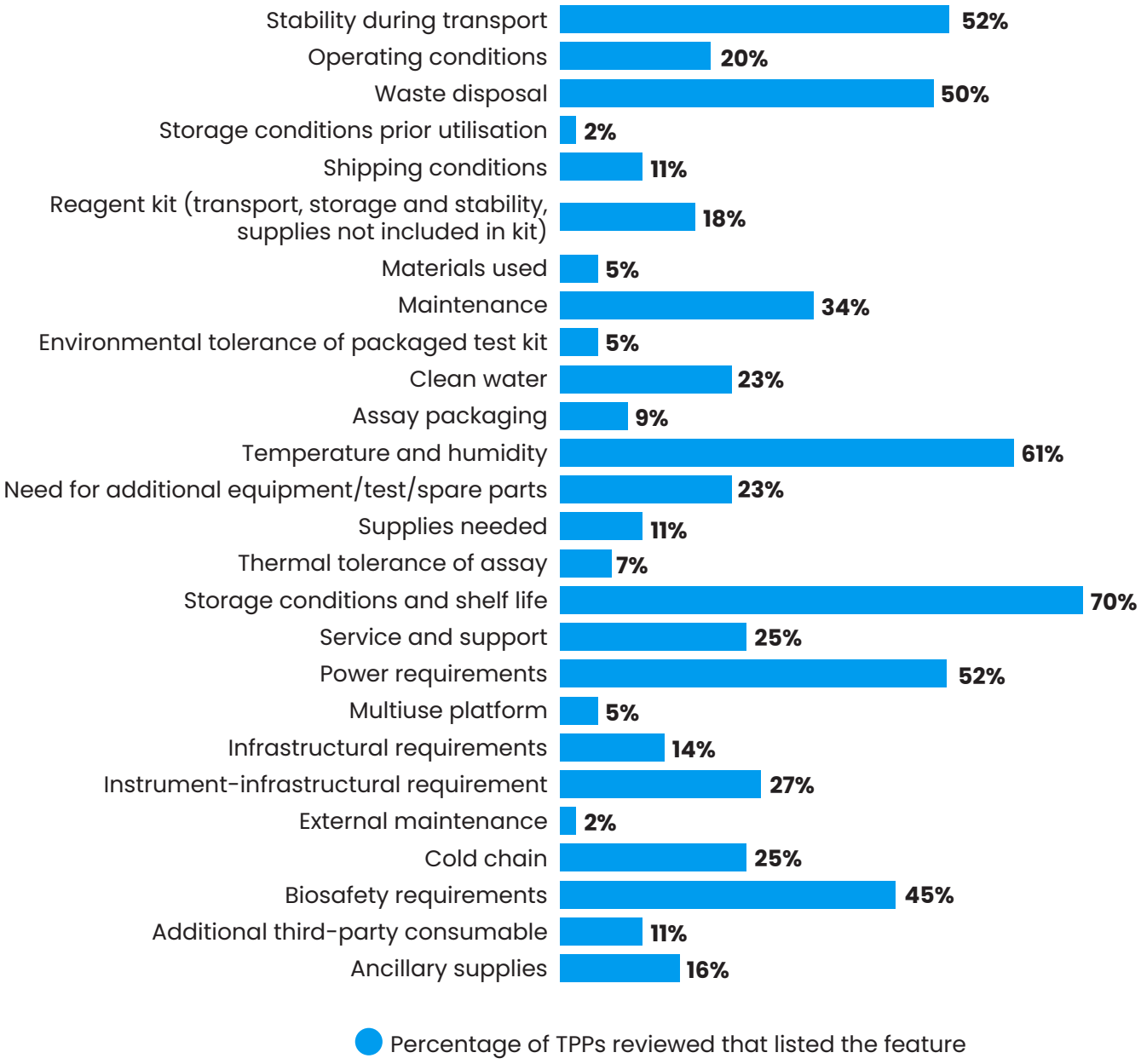


Table 6. Features listed under the 'Infrastructural Requirements' category in the Cocco et al. (2020) review, along with working explanations

Feature (*denotes common features)	Explanations
BASIC CONDITIONS RELATED TO INFRASTRUCTURE	
Operating conditions	Information on the external conditions necessary to support the medical test's operation (including storage, stability and supplies not included in the kit).
Biosafety requirements	Information about the biosafety requirements that must be in place for the test's safe operation.

Feature (*denotes common features)	Explanations
Cold chain	Information on the chain of events in temperature-controlled environments needed to store, manage and transport medical tests, including cold-chain equipment and facilities requirements.
Thermal tolerance of assay/test	Information on the optimal temperature for the medical test, including its capacity to tolerate temperature change.
Temperature and humidity*	Information on the optimal temperature for the test, including its capacity to tolerate temperature and humidity changes.
Environmental tolerance of packaged test kit	Information on the packaged test kit's tolerance to different environmental factors, such as temperature and humidity.
Clean water	Information on whether clean water is necessary for the test's operation.
Power requirements*	Information on the power supply necessary to support the test's operation.
Stability during transport*,**	Information on the test's stability and required conditions during transport. **A combination of two features listed in Cocco et al. (2020): 'Shipping conditions' and 'stability during transport'.
Waste disposal*	Information on the waste disposal practices of the medical test, including any potential special arrangements necessary, e.g. in the case of toxic or hazardous materials.
SPECIFIC INFRASTRUCTURE CONDITIONS FOR OPERATIONS	
Storage conditions and shelf life*,**	Information on the storage conditions necessary for the medical test, including information on its shelf life and the conditions necessary before use. **A combination of two features listed in Cocco et al. (2020): 'storage conditions and shelf life' and 'storage conditions prior utilisation.'
Equipment and supplies needed**	Information about the equipment and supplies necessary to support the operation of the medical test (that are not part of the test kit itself). This also includes Information about medical supplies and/or durable medical equipment necessary to operate and administer a medical test without being an integral part. Note: Individual TPP efforts may want to separate equipment from supplies and different types of supplies from each other. **A combination of seven features listed in Cocco et al. (2020): 'Instrument-infrastructural requirement', 'infrastructural requirement', 'supplies needed', 'need for additional equipment/test/spare parts', 'ancillary supplies' and 'additional third-party consumables' and 'materials used.'

Feature (*denotes common features)	Explanations
Multiuse platform	We assume this is about the information on a test platform applicable to multiple markers, but we are unsure. One advisor commented that this relates to the acceptability of a platform to test for a single disease.
Reagent kit (nature, transport, storage and stability, supplies not included in kit)	Information on the nature of the reagent kit and how it should be transported, stored and kept stable, as well as information on which supplies are not included in the kit and must be acquired externally. Note: Individual TPP efforts may want to distinguish between information about the reagent kit itself and information on how it is stored, transported and kept stable. If the reagent kit is part of the overall test kit, then information on the reagents could also be provided in the 'Analytical Performance Indicator' category. In contrast, storage, transport and shelf-life information could be covered under the 'Infrastructure requirements' category. Note: This also appears under 'Analytical Performance' category. It is likely unnecessary in both and may be better in one or the other.
Assay packaging	Information on assay packaging and safety seal used to confirm the contents are authentic and have not been tampered with and that they are authentic.
Maintenance (including servicing and support)**	Information on the external or internal maintenance required for the medical test's operation, including servicing and support. Some tests will require both external and internal maintenance of different types. **A combination of three features listed in Cocco et al. (2020): 'maintenance', 'service and support' and 'external maintenance.'

8. Costs/Economic Considerations

Commonly used features in this category across TPPs

This category covers features related to economic costs and other commercial considerations. Cocco et al. (2020) listed 11 features in this category (see Figure 7), of which only one featured in 50% or more of the TPPs they reviewed:

- Price/cost of individual test (61%).

Scope for streamlining

We suggest a scope for combining the following feature groups from Cocco et al. (2020):

- 'Potential market', 'Market segmentation/ channels to the market' and 'Region(s) of commercialisation'
- 'Capital cost per instrument' and 'Costs of platform to end user' (see our reasoning in the table below).

Figure A7. Features listed under the 'Costs' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs that mentioned each feature

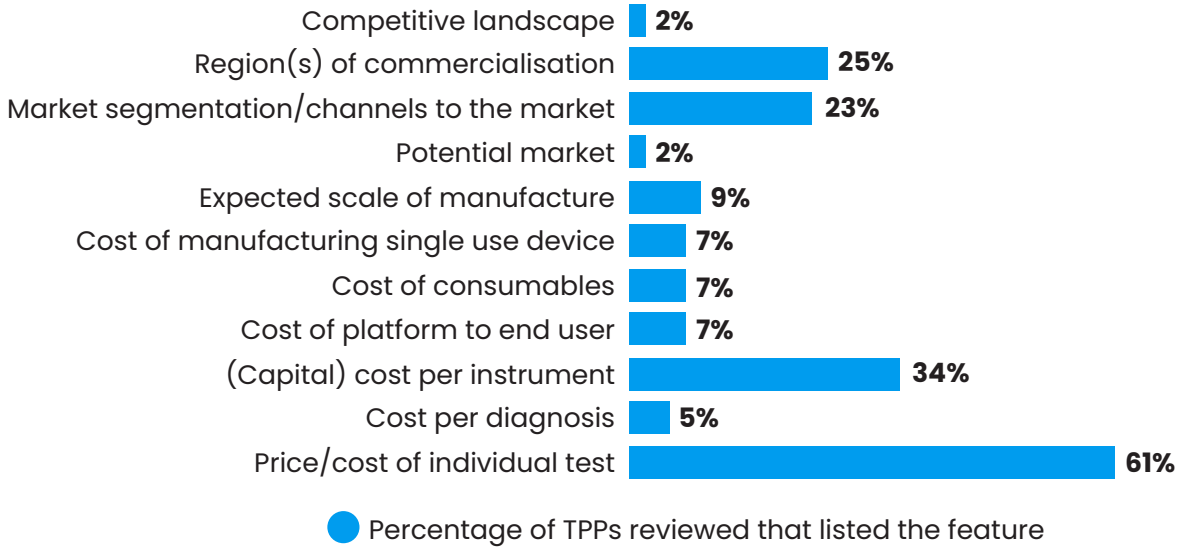


Table 7. Features listed under the 'Costs' category in the Cocco et al. (2020) review, along with working explanations

Feature (*denotes common features)	Explanations
Price/cost of individual test*	Information on an individual test's overall price or cost to the payer (NOT the production cost). One advisor felt a target should be estimated based on current tests for the same condition or similar approved tests.

Feature (*denotes common features)	Explanations
Cost per diagnosis	The overall cost of a test per diagnosis made. Note: This could include broader information extending beyond the base price/cost per individual test to consider cost factors related to its use in practice, e.g. how training costs and the number of individuals needing screening to find one case impact the costs per test. One advisor highlighted the need for an economic evaluation to choose a costing perspective. An estimate may need to be made in the early stages due to uncertainty.
(Capital) cost per instrument**	Information on the fixed one-time costs associated, for example, with purchasing instruments, equipment or infrastructure needed to run the tests. Note: we combined this with 'costs of platform to end user', which seemed vague and duplicative in the Cocco et al. (2020) list. *A combination of 'capital cost per instrument' with 'costs of platform to end user.'
Cost of consumables	Information about the costs of ongoing material supplies needed for the testing to happen, including reagents.
Cost of manufacturing single-use device	Information on how much it costs to manufacture a specific single-use test or testing device, specifically the manufacturing costs.
Expected scale of manufacture	Information on how much / how many units will be made.
Potential market	Information on a test's potential market size, number of users and market segments (different user segments, which may be geographical markets or different user types, e.g. hospital use or at-home use) or, in the case of multiplex platforms, uses for different diseases.
Market size, nature and segmentation**	Information on a test's potential market size, number of users and different market segments (different user segments, which may be geographical markets or different user types, e.g. hospital or at-home use) or, in the case of multiplex platforms, uses for different diseases. It also includes information on routes to market (e.g. who the payers are). **A combination of three features from Cocco et al. (2020): 'Potential market', 'Market segmentation/channels to the market' and 'Region(s) of commercialisation'. Note: Individual TPP efforts may want to divide this into more than one feature.
Competitive landscape	Information on costs of other tests and other available tests on the market.

9. Regulatory Requirements

Commonly used features in this category across TPPs

This category covers features related to the test's regulatory requirements and pathways information. Cocco et al. (2020) listed two features in this category (see Figure 8), of which none featured in 50% or more of the TPPs they reviewed.

Scope for streamlining

None.

Figure A8. Features listed under the 'Regulatory Requirements' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs that mentioned each feature

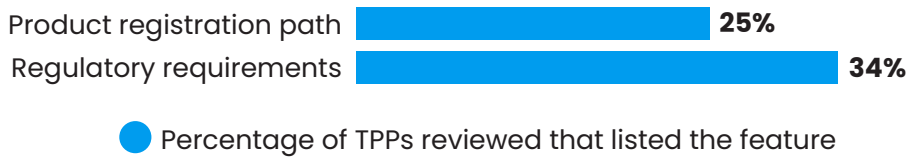


Table 8. Features listed under the 'Regulatory Requirements' category in the Cocco et al. (2020) review, along with working explanations

Feature	Explanations
Regulatory requirements	Information on what specific regulations the test must meet given its intended market. Note: Individual TPP efforts may develop specific subcategories to cover extensive regulatory requirements in oncology.
Product registration path	Information on whether the products are relevant for a particular jurisdiction. Product registration is the initiation of any regulatory process.

10. Environmental Impact

Commonly used features in this category across TPPs

This category covers features related to a test's impact on the environment. Cocco et al. (2020) listed one feature in this category (see Figure 9); this was not mentioned in 50% or more of the TPPs they reviewed.

Scope for streamlining

None, N/A.

Figure A9. Features listed under the 'Environmental Impact' category in the Cocco et al. (2020) review, including the percentage of reviewed TPPs that mentioned each feature

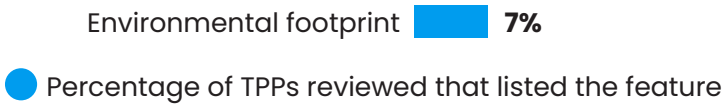


Table 9. Features listed under the 'Environmental Impact' category in the Cocco et al. (2020) review, along with working explanations

Feature	Explanations
Environmental footprint	Information about the environmental impact of test production and use, including the impact of the test's manufacture and use on the environment. Note: Individual TPP efforts may want to distinguish between the environmental impact associated with a test's manufacture and the environmental impact associated with its use.

11. Features and categories missing in the current evidence base

Features and categories which are missing in the current evidence base include the:

- Non-economic impacts on patients, e.g. patient acceptability and experience.
- Ability to address inequalities.
- Downstream effects on care pathways and processes, e.g. the implications of test findings on further care needs and processes.
- Downstream impacts on health systems and population-level outcomes (potentially as part of/through economic modelling).
- Non-economic downstream impacts on patients.
- Importance of meeting real-world performance needs, not just in lab conditions.
- Importance of cost-effectiveness in health economic modelling as well as cost considerations.

References

1 Clinical and Laboratory Standards Institute. 2023. *The Harmonized Terminology Database*. As of 03 April 2024: https://htd.clsi.org/

2 Medicines and Healthcare products Regulatory Agency. 2020. Vol. MHRA Guidance Note 8 (ed Medicines & Healthcare products Regulatory Agency). MHRA, London, UK.

3 Richards, M., Hiom, S. & Hamilton, W. 2023. 'Diagnosing cancer earlier: what progress is being made?' *British Journal of Cancer*, 1-2.

4 Armbruster, D. A. & Pry, T. 2008. 'Limit of blank, limit of detection and limit of quantitation'. *Clin Biochem Rev* **29 Suppl 1**, S49-52.

5 Saah, A. J. & Hoover, D. R. 1998. 'Sensitivity and specificity revisited: significance of the terms in analytic and diagnostic language'. *Ann Dermatol Venereol* **125**, 291-294.

6 Lister, A. 2005. Validation of HPLC methods in pharmaceutical analysis In *Separation Science and Technology*, edited by Michael W. Dong Satinder Ahuja. Academic Press Vol. 6, 191-217.

7 Patijn, J. 2019. 'Reproducibility protocol for diagnostic procedures in Manual/ Musculoskeletal Medicine'. *Manuelle Medizin* **57**, 451–479.

8 Aubry, A.-F. & Weng, N. 2015. 'So you think your assay is robust?'. *Bioanalysis* **7**, 2969-2971. As of 22 March 2024: https://doi.org/10.4155/bio.15.198

9 Dimeski, G. 2008. 'Interference testing'. *Clin Biochem Rev* **29 Suppl 1**, S43-48.

10 Misra, P., Kant, S., Guleria, R. et al. 2022. 'Test concordance and diagnostic accuracy of three serological assays for detection of anti-SARS-CoV-2 antibody: result from a population-based sero-epidemiological study in Delhi'. *BMC Infect Dis* **22**, 915.

11 Kim, W., Woo, Y. S., Chae, J. H. & Bahk, W. M. 2011. 'The Diagnostic Stability of DSM-IV Diagnoses: An Examination of Major Depressive Disorder, Bipolar I Disorder, and Schizophrenia in Korean Patients'. Clin Psychopharmacol *Neurosci* **9**, 117-121. As of 22 March 2024: https://doi.org/10.9758/cpn.2011.9.3.117

Annex C: Stakeholder Workshops Cross-Analysis Document

Authors: Mark L Cabling, Jessica Dawney, Matthew Napier, Zuzanna Marciniak-Nuqui, Fifi Olumogba, Larry Kessler, Amanda Cole, Lotte Steuten, Sonja Marjanovic

1.1. Introduction

Context

Annex C is the third of eight annexes complementing the main Cancer Research UK-funded project's final report: 'Advancing the development and use of diagnostic target product profiles for cancer.' The not-for-profit research institute RAND Europe led the project in collaboration with the Office of Health Economics. The project has benefited from ongoing support and advice from Professor Larry Kessler (University of Washington), a key consultant on the work. This document provides detailed findings and analysis of the six workshops conducted with different stakeholders; thus, like all the other annexes, Annex C is primarily meant to accompany the final report and is not meant to be read as a standalone document.

The following abbreviations for each stakeholder workshop are used when referencing sources of information (i.e. individual workshops) in the contents that follow:

- Advisory Group (AG)
- Academics, clinical academics and Research funders (ARF)
- Healthcare professionals and pathology laboratory managers (HP)
- Industry (IND)
- Policy, Regulators, HTA and Health Economists (PRHE)
- Patient, Carer and Public voice and Charities perspectives (PCPC).

Stakeholder workshops

As part of the project, we conducted a series of stakeholder workshops to engage the experiences and perspectives of individuals from academia and research, healthcare, industry, policymaking, regulation, health technology assessment (HTA), health economics and patient, carer and public perspectives. Table 1 below provides further information.[i] The workshops explored various issues to inform an approach to developing an oncology diagnostic TPP for use in the United Kingdom (UK) and identify the types of features that a TPP would need to consider. Discussions focused on understanding what stakeholders saw as key challenges to cancer diagnosis in the National Health Service (NHS) and associated areas in need of improvement, challenges to the development and adoption of improved diagnostic tests, the types of issues needing consideration in a demand-signalling TPP for cancer in the context of important types of features to specify, and what an approach to developing a diagnostic TPPs for oncology in the UK might look like, including how different types of stakeholders could contribute to future efforts.

We held six workshops facilitated by the project team between 22 May and 13 July 2023. Each workshop lasted 2.5 hours and was conducted online via MS Teams. In addition to participants from the project team and the client team (the latter as observers), the workshops gathered insights from 92 individuals across the advisory group and external participants. The wider project and workshops did not require ethical approval, according to the Health Research Authority (HRA). However, we complied with all principles of informed consent and legal data privacy and security requirements.

i All external participants for workshops ARF, HP, PRHE, IND and PCPC received a project information sheet and privacy statement explaining participation to assume consent to participate, be recorded and be named as a contributor in potential project outputs. At the onset of the workshops with external participants, participants were reminded that the sessions would be recorded and that we would like to name them as contributors in potential outputs and asked to let us know if they prefer not to be named. We did not receive any requests not to be named. Since the project's Advisory Group is an internal project structure, they received the agenda and were guided by an overarching document outlining their role in the project. However, we have also followed up with participants in the AG workshop to ensure they were comfortable potentially being named in project outputs as workshop participants, and they have all confirmed.

The contents below summarise key insights from across the workshops and draw on individual workshop write-up documents produced by the research team.

We fed the insights into the overall project's cross-analysis, synthesis, guide development and reporting in later project stages.

Throughout the document, we occasionally offer research team reflections on the insights offered and the conversations that took place during the workshops. When doing so, we clarify that these are our reflections using the term 'research team note'.

Table 1. Stakeholder workshops

Stakeholder group	Workshop date and time	Number of participants (external and from the advisory group)
Advisory group (multisector)	22 May 2023, 13:00 – 15:30	10 advisory group members[ii]
Academic/clinical academics and research funders	1 June 2023, 12:30 – 15:00	18 (2 advisory group members, 16 external participants)
Healthcare professionals	15 June 2023, 10:00 – 12:30	12 (1 advisory group member, 11 external participants)
Industry	23 June 2023, 11:30 – 14:00	21 (1 advisory group member, 20 external participants)
Policymakers/ payers, regulators, HTA and health economists	6 July 2023, 10:00 – 12:30	13 (1 advisory group member,[iii] 12 external participants)
Patients, carers, public voices and charities	13 July 2023, 11:00 – 13:00	18 (1 advisory group member, 17 external participants)
Total		92 external participants

ii We followed up with an AG member who could not attend the AG workshop (n=1) to gain their insights. An additional AG member joined after the AG workshop was held, but that member also attended another workshop.

iii This AG member joined after the AG workshop was held.

1.2. Key insights from the workshops

1.2.1. Should a TPP be developed? Key considerations related to the relevance and importance of developing a TPP

An overview of key considerations:

- **Several factors underpin the relevance of any effort to develop a TPP**. Broadly speaking, these include **(a) appropriately identifying an unmet need** for a novel diagnostic and hence for a TPP as a demand-signalling document, **(b) ensuring a TPP remains relevant and up-to-date but with sufficient longevity as a stable demand signal** without changing specifications too often, **(c) ensuring an appropriate market for a resulting product** (which may involve considering international as well as domestic markets). Each factor plays a crucial role in the relevance of an effort to develop a TPP, as elaborated below:
  - **A TPP must focus on informing the development of a diagnostic that responds to a well-informed and established unmet diagnostic need and clearly articulated features needing improvement (ARF, AG, PRHE). There is tension between specifying features for UK-specific needs (which may be distinctive) and for relevance in major markets globally (which may maximise the viability of development)** (IND, ARF). Global trends in diagnostics are important factors in incentivising industry innovation, even within the UK (ARF). At the same time, there may be tests that exist internationally but are not affordable for the NHS, giving rise to a need for improved diagnostics for the UK market specifically. Whether a TPP could help incentivise innovation for UK market needs alone will depend on that market's value and should be scoped and assessed before embarking on a TPP development effort for a UK-only market.
  - **The relevance of a TPP can also be influenced by how dynamic and responsive a TPP document is. There must be a balance between keeping a TPP a relevant but stable demand signal, allowing sufficient time for innovators to respond with a diagnostic development and capture value from their investments and innovation efforts** (ARF, AG, PRHE, IND). TPPs must be kept up to date (ARF). However, changing specifications too often can disincentivise innovators (whether in industry or the wider research and development community) [Research team note: We note that it is not just industry that acts based on TPPs, but also other stakeholders that may or may not be part of industry, such as researchers and developers in the public sector and those collaborating with industry]. This is due to risks of the nature of demand and the market for a product changing by the time a diagnostic that responds to an initial version of a TPP is developed (AG, IND). It is important to consider short- and longer-term relevance (ARF) at the outset of the TPP development effort, and about which features may be more stable or prone to changing specifications, and how that might impact the relevance, appropriateness and uptake of the TPP as a demand signal [Research team note: One advisory group participant flagged that they felt a TPP needs to aim to be relevant for at least 5–10 years from its publication to continue attracting and incentivising innovators (AG). We

recognise that this timeline may still be quite long for diagnostics, though, as they are often treated as medical devices and would have a shorter term of relevance of around 2–5 years due to technological advances. This can be a challenge if the field, technology or demand areas change rapidly]. There can also be resource limitations and a lack of stakeholder capacity to frequently update TPPs, which needs to be borne in mind (ARF).

- **A TPP can be a useful tool for signalling demand to innovators and may help in efforts to align supply with demand. However, broader systemic challenges in the NHS (e.g. funding, workforce, and challenges to implementing novel testing processes) have implications for the likely implementation of any novel test that might result from a TPP** (HP, IND). Although TPPs may help address information and evidence gaps on the test types needed, they cannot solve broader challenges to incentivising innovation and its adoption in the NHS alone without wider policy levers (IND).
- **TPPs create an opportunity to provide 'additional' supportive information and guidance that can help address some of the systemic challenges to novel test development (IND). However, it was acknowledged that TPPs sit within a broader landscape of innovation tools and resources, and a TPP's scope needs careful thought** (IND). Some examples of challenges the industry would appreciate guidance on concern clarity on regulatory requirements, HTA requirements, possible reimbursement pathways and ways of accessing clinical samples (IND) [Research team note: Whether an individual TPP needs to provide the information or can point to related information sources on these issues is worth considering].
- **TPPs can help industry and others who might be involved in developing novel diagnostics (such as academic and clinical researchers) navigate the market for diagnostics** (PRHE): the market for diagnostics is characterised by Small and Medium-sized Enterprises (SMEs) in addition to some larger players, and smaller players may have minimal experience with the process and requirements of health system stakeholders. TPPs can support SMEs in developing the right innovations to meet unmet needs by clarifying the diverse features needed. The diversity of relevant features reinforces the need for diverse stakeholder contributions to TPP development processes.

1.2.2. General insights on TPP development processes

TPP development tends to occur in two key stages. The first is the **inception/preparation stage**, establishing a core working group, governance and coordination arrangements, and an action plan. The second is the **implementation stage**, which has three phases (building on insights from a review by Cocco et al. (2020)).[1] These include **scoping** the unmet need and key novel test requirements, **drafting** TPPs, and **consensus-building** to explore and establish consensus on a final draft. Drafting and consensus building often happen iteratively. The sections below summarise workshop insights on these stages of TPP development.

1) Inception/preparation stage: Establishing a core working group, ensuring that the right expertise will input into TPP development processes, and deciding on the overall approach/plan of action

Establishing the core working group and plan of action:

- **A core working group is a helpful body to lead and oversee the TPP development effort and establish a plan of action** (AG, ARF, HP, PRHE, IND, PCPC):
  - **A plan of action will cover key aspects of TPP development** (AG, ARF, HP, PRHE, IND, PCPC), such as which stakeholders (organisations and individuals) should contribute to the effort, on what issues, how (which methods to use), how long developing a TPP is likely to take, how it should be phased, the resources available for the effort/funding, and what the desired outputs will be [Research team note: in reflecting on broader conversations during the

workshops, an underlying theme seems to be that rigour in TPP development is important but the methodological trade-offs those driving the effort need to consider will be partly influenced by financial resources available for TPP development and the envisaged urgency/acceptable timeframe for the task. This suggests a need for flexibility in the plan to accommodate potential variety in feasible and/or preferred ways of contributing from diverse stakeholders over time].

  - An advisor noted that it can take approximately six months to develop a TPP (AG) [Research team note: although insights from earlier research conducted by the project team suggest it may take a year or even longer]. When considering whether it is worth developing a TPP for a specific use case, it can be helpful to consider whether any diagnostic tests developed in response to it will reach the market before the technology becomes obsolete, particularly in areas where the pace and rate of scientific development is rapid (HP).

- **The core working group must reflect diverse expertise** (AG, ARF, PCPC, HP, PRHE, IND), as elaborated in the following sections. Terms of reference for the core working group must be made clear from the outset, including for different member types (PCPC).
- **Expertise across a whole research, innovation and care pathway is considered another important aspect of TPP development to ensure a range of perspectives** (e.g. from R&D to approvals and access) (HP).
- In a UK context, participants also highlighted the importance of having representation of each of the **devolved nations** (HP).
- **Academic/research expertise**: While scientific and technical expertise in a disease and diagnosis area from academics and clinical academics matters are important for a core working group (ARF, AG), other research expertise may also be important, e.g. health economics expertise (ARF) and implementation science expertise and social science expertise specialising in health inequalities (ARF). The clinical, academic, disease and diagnostic expertise needed will depend on the test type, cancer site, use setting and use case for which a TPP is being developed (AG). Pathology lab expertise is also often relevant in the research sense and regarding pathology lab practitioners (AG). Expertise in methods as well as topics is important. For example, health economics methods (ARF, AG) and horizon-scanning methods input into TPP development (ARF). Horizon-scanning expertise is important for assessing unmet needs vis-à-vis available tests (ARF). In addition, modellers, statisticians, and health economists can help other core working group members interpret the modelling necessary for health technology assessments by NICE (AG) to help the core group understand whether modelling validates a minimal and optimal specification range for features (AG) and identify which features and specifications might have the highest impact in a value proposition (PRHE).
- **Healthcare professional expertise** is also essential in a core working group (AG, ARF, HP, IND). Health professionals can provide insights on areas of unmet need regarding existing tests' 'technical' performance (e.g. accuracy) and unmet needs to ensure a diagnostic test fits with the health service's clinical and care pathways and usability (ARF). Therefore, healthcare professionals are key in defining and specifying the value proposition given the technical, human and adoption-context considerations (ARF). The healthcare professionals to engage will depend on TPP use cases (e.g. test types, cancer sites, and primary or acute care use settings) (ARF, AG, HP). Pathology and genomics lab managers' involvement was mentioned as necessary for many test types because they have everyday access to varying diagnostic sample types and expertise relevant to test validation and implementation into workflow streams (ARF, HP). Professional networks and organisations such as Cancer Alliances and potentially Royal Colleges also have a role to play (HP). Some participants highlighted the need for health professionals with clout and credibility to be on the core working group (IND).
- **Patient and carer representation** (lay and expert, via charities) is also vital in the core

working group and must be in place from the outset of developing a TPP to ensure it considers features related to the end-user experience, including test useability and accessibility (ARF, AG, HP, PCPC, PRHE). According to diverse participants in the workshops, the patient and public voice has been a lower priority than it could and should be in diagnostics development (and in developing TPPs in other disease areas), and this is an area ripe for improvement because patient experiences are central to understanding needs and inequalities (AG, PCPC, HP). However, there was also a recognition of challenges related to gathering patient and public representative views, such as time restrictions and creating effective ways for patients to communicate and articulate their needs (AG). Patient and public opinions can also be gathered indirectly, e.g. via a patient's partner or close relative (AG) and charities (PCPC). However, direct lay patient experiences also matter (PCPC).

- » **It is important to seek diversity in who contributes to the patient-and-carer voice and on what issues to ensure that diverse voices are heard during the TPP development processes** (PCPC). The most appropriate type of patient, carer and public-voice representation in the core working group will be context-dependent (e.g. cancer type and test use case). Also, the core working group's size must be manageable. However, workshop insights suggest that diverse patient, carer and public voice input should be sought in TPP development, not confined to the core working group representation (PCPC). Some cancer areas have established **panels and patient/carer/public voice groups/communities** for specific cancer types, which can help keep discussions focused and individuals engaged while still reflecting diverse voices. One participant advocated for consultation with **peer support groups** as a good source of information about diagnostic tests needing development and unmet needs (PCPC). Smaller groups can also be assembled to input into specific TPP development efforts (PCPC). It is essential to try to include the **underrepresented and under-involved groups in TPP development** (e.g. specific ethnic groups and those with learning disabilities). Patient, carer and public-voice groups can be self-selecting and often insufficiently representative of general patients. Therefore, it is important to solicit input from people not immediately interested in participating, not just those who are easiest to access. Some people will not want to engage, but there is a need to speak to as many people as possible (PCPC). **Charities** (research or other) can also provide helpful contributions (HP, PCPC).
- » **The terms of reference and role of patient, carer and public voice contributors must be clear to participants from the start, e.g. their role, commitment and compensation, and how they can provide the most beneficial input** (PCPC). [Research team note: However, we think caution is also needed to allow sufficient scope for bottom-up views on what is required from patient, carer and public-voice contributions to emerge, i.e. to be clear about the overarching themes but not be overly prescriptive from the onset].

- **Regulatory and health technology expertise and health economics are** important so that regulatory and HTA requirements can be considered early in TPP development (AG, ARF, HP, PRHE, IND) and made clear to innovators. Innovators responding to a TPP can then consider regulatory and HTA needs early in product design rather than risk their products failing on regulatory grounds and thus failing (AG, HP). Regulator/regulatory expertise can inform health economists on requirements that can feed into early economic modelling to inform TPP specifications (ARF).
- **Policymaking and arm's length bodies representatives (e.g. NHS England and All-Party Parliamentary Groups**, as some workshop participants mentioned) also matter. They considered these types of organisations/bodies to hold insights on what the policy priorities are in a given area, such as oncology, which may have implications for understanding whether a TPP is 'worth developing', i.e. whether any products resulting from it are likely to gain

traction in the NHS and to ensure that the criteria for what is paid for are clear from the beginning of the TPP development process (ARF, AG, HP).

- **Some specific expertise types are not confined to any one stakeholder group but could be important for a core working group**. For example:
  - **Product development expertise** (ARF) can come from industry, consultancy and academic research groups.
  - **Product implementation expertise** (ARF) can include innovators from academia/research and industry with experience implementing diagnostic tools in the NHS (including outside of cancer). They can provide useful expertise (if there is no conflict of interest) because clinical usability, cost-effectiveness, and implementation are vital considerations in a TPP. Implementation science/ implementation research expertise from academia can also add insightful contributions.
  - **Potential payers/commissioners** can provide perspectives from those who decide on (ARF, AG and PRHE) and pay for the adoption and uptake of diagnostic tests, helping shed light on existing tests' improvement needs for which there is a viable value proposition and broader cost considerations. Many tests can fail later in development because commissioner-related considerations were not considered early (AG). In some cases, centralised bodies with a policy role (like NHS England) may be the funding provider under national decisions to implement specific tests widely. However, other payer types may hold relevant budgets, e.g. hospital trusts.
- **TPP development also requires engagement from other stakeholders (such as industry)** (AG, ARF, HF). However, **views differ on whether industry should be (a) represented in a core working group or (b) consulted as part of the broader TPP development process, even if not represented in the core working group**.
  - Some felt that the risk of bias is too high with industry representation in a core working group (ARF). However, they noted that trade associations could potentially help mitigate the risks of bias from representing a small number of companies (IND). One attendee argued that ensuring any engagement with industry remains transparent could address real and perceived bias from industry during the TPP development effort.
  - **Diversity in terms of industry engagement in TPP development efforts matters, and a TPP needs to be usable for different types of industry innovators** (IND), including different company types and sizes (SMEs and large companies). There were mixed views as to when it is best to consult industry. For example, some participants in the advisory group felt developers should only consult industry in later TPP development stages once drafting is complete (essentially to elicit views on the appropriateness of a draft and explore consensus) rather than formally engage industry in consensus (AG). Others felt industry should be engaged from the beginning because they know the market and understand what is commercially viable (AG, IND).
- **Diagnostic equipment suppliers** were also considered important to involve in TPP development consultation to understand what is required to deliver the test and ensure readiness for deployment in clinical practice when required (HF).
- **Charities funding research and innovation activity and public sector funders** are also relevant to consult and will have perspectives helpful for informing unmet needs (ARF, HP) [Research team note: reflecting on workshop discussions, whether to involve research funders on a core group or to consult them more widely as part of the TPP development process may be a context-specific decision dependent on funder landscape (e.g. how diverse or concentrated in a few funding organisations it is, how possible it is to have independent representation/avoid bias) and the role of the funder relative to the TPP development process (e.g. are they funding the effort/ sponsoring it)].

- **A core working group must be manageably sized**, but no evidence suggests an appropriate number, which may be context-specific depending on the expertise types and diversity needed. According to one workshop member, the number ranged from 15–40 members in their experience (AG). Following up with this individual after the workshop, he also clarified that he experienced fewer than 20 people when working in a TPP development group but has heard of other groups as large as 40 working on TPPs.
- **Even if represented in the core working group, members of stakeholder communities identified above should also be consulted during TPP development on a broader scale** (ARF). Whereas some representatives may share views through the core working group, a wider range of individuals from most, if not all of, the stakeholder groups would likely need to feed into TPP development process consultations (e.g. a broader range of academics/researchers, healthcare professionals, patient voices, industry, experts in product development and implementation and possibly also policy, regulatory and HTA stakeholders depending on specific use case need and the nature of core working group representation) (ARF).
- [Team reflective note: While diverse perspectives must be included in TPP development, the nature of contributions made by different stakeholder groups must be proportional to the value distinct perspectives can add and feasible for all involved. While different stakeholders' views matter and diverse stakeholders need to be consulted, not every stakeholder group must be represented on a core group leading the TPP development effort. Core group members can include boundary-spanners whose roles allow them good insights into diverse stakeholders' developments and priorities. Some specific expertise types are not necessarily confined to any single stakeholder group.]

Governance of TPP development:

- TPP development is an inherently collaborative process. We understand TPP development governance to mean the organisation/institution formally hosting the TPP development process and being best placed to take ownership of steering and coordinating the effort (supported by the core working group). Hosting TPP efforts should be more of a stewardship than a strict 'ownership', given the collaborative nature of TPP development and the need for buy-in from diverse organisations (PRHE).
- **The host institution's reputation matters** (PRHE). They must be trusted and seen as unbiased.
- We asked workshop participants for their thoughts on who might oversee/govern, coordinate, and fund TPP development in a UK oncology context. No particularly strong view or clear message on this issue emerged from the workshops, but the following points were made:
  - The importance of organisations close to decision-makers/ decision-making points and procurement was raised. Organisations whose buy-in for TPP development is critical and who have a key role in any resulting test's adoption and uptake were suggested as one option for governance and steering/ hosting roles (ARF, AG, PCPC). Some examples given included NICE and NHS England. However, there were some reservations about sufficient independence and unbiased oversight from a body as close to policies as NHSE, even though involvement was seen as necessary in TPP development (AG). Another suggestion was to consider an all-party parliamentary group for diagnostics (HP); any organisation involved would need convening power and the capacity and coordination ability to oversee the delivery of the effort effectively.
  - Third-sector charities or public sector research funders were also considered potential 'hosts' who could bring good networks to the process and be trusted as independent (ARF, AG, PCPC, HP, IND). They were also considered capable of pursuing a longer-term and more strategic view than other organisations that may be more impacted by other agendas (AG). Depending on the use context, non-cancer-specific charities or cancer-specific ones like CRUK were

also mentioned. Charities were noted as a potential group (HP, PRHE, IND) that could support governance or funding or both (HP) and whose wide research and academic networks (for research active charities), convening powers, public good remit and potential funding could help (IND).

» However, some attendees mentioned that whichever organisation or body 'owns' or overseas a TPP development effort should depend on its use case. For example, some workshop participants flagged that if regulatory considerations drove a use case, a regulatory agency may be considered for an oversight role (IND) [Research team note: It would be critical to ensure no conflict of interest could influence subsequent licensing decisions].

» Some of those consulted in workshops (ARF) felt that while it is important to consult industry, its representatives should not steer/coordinate/govern the effort as this would risk biased commercial interests having too strong an influence (ARF).

» Cancer alliances and diagnostic networks were also mentioned as groups that could potentially oversee TPP development. Cancer alliances and the wider diagnostics community interact (HP). According to some workshop participants, diagnostic networks – including imaging, pathology and endoscopy – that were previously siloed are now increasingly consolidating and collaborating. Cancer alliances were identified as a group that could be involved in funding (resource permitting) and oversight or consulted as a community. Some suggested a role for **specialised professional bodies**, such as the Royal Colleges and Societies, who can support access to wider groups, members of Parliament, and Departments for Health (HP).

2) Implementation stage

The implementation stage of TPP development has three phases: (a) scoping the unmet need and key novel test requirements, (b) drafting TPPs, and (c) exploring and establishing consensus on a final draft. Drafting and consensus building, in particular, often happen iteratively. We summarise key workshop insights about these phases in the contents below.

Scoping phase

Before the workshop discussions, the research team provided an overview of insights on the scoping phase of TPP development based on desk research and interviews conducted earlier in the project. Annex A summarises the key points from the overview.

Key insights on the scoping phase from the workshop discussions:

- **The scoping process must be rigorous to ensure that an unmet need exists that accessible tests do not already respond to, and key requirements for a novel test must be identified and specified. This process requires diverse expertise and methods** (AG, HP, PRHE), e.g. evidence syntheses through systematic or rapid reviews of academic and grey literature depending on need and gaps in knowledge (ARF). The WHO's standards call for a literature review and technology landscape analysis (AG).[2]
- **However, there is a need to 'right-size' rigour and ensure a feasible TPP development process; the appropriate methods to support sufficient rigour will depend on various factors, such as the amount of existing research, key issues of interest, how consolidated the required evidence is or is not** (e.g. whether there are already recent systematic reviews addressing the issue), **how urgent the need for a novel test is and how much resource and capacity there is to invest in TPP development** (ARF). **While balancing optimal methods with pragmatic considerations is critical, there must be sufficient rigour to have confidence in the resulting TPP specifications**:
  - » For example, a systematic review may be warranted if a particular cancer area has already received considerable research activity relevant to diagnostic insights, but the evidence is not consolidated (ARF). However, if there are recent systematic reviews of good quality already, an update with a rapid review might be sufficient.

  - » A phased approach with more rapid assessments of unmet needs is also possible, followed by a call for more systematic means if needed (AG), i.e. if the rapid review reveals important evidence gaps.
  - » However, participants highlighted that TPPs must consider the evidence hierarchy, particularly when clinical expert opinion informs much of the specification work. The highest quality evidence possible is always preferable, and over-reliance on expert opinion alone should be avoided (where possible) when informing TPP specifications (PRHE).

- **Combinations of methods matter for the scoping phase to identify unmet needs and novel test requirements** (ARF, HF). For example:
  - » **If there are gaps in the research evidence (literature and expert opinion) on unmet needs and key test requirements, modelling may help shed light on the gaps and inform TPP specifications** (ARF, PRHE). Understanding which features matter most and the trade-offs associated with meeting different feature requirements is crucial early on. Early economic modelling can help complement insights from the literature by exploring feature combinations and associated trade-offs from different scenarios/combinations of specifications (ARF, PRHE) [Research team note: Such modelling can happen during the scoping phases but also in the drafting phase once some specifications have been commented on via wider consultation and are more firmed up]. Early economic modelling has a role in helping to address risks of overly aspirational rather than realistic TPPs by creating a more robust scientific basis for prioritising test features. This is because such modelling can help identify the features (and combinations of features) likely to yield the greatest beneficial impact and where effort may best be directed (PRHE).
  - » **Concerning mapping available diagnostic tests and/or horizon scanning for likely emerging tests, a decision must be made about whether to focus on tests available in the UK market only or internationally (ARF). This decision will influence which diagnostic databases warrant examination and which experts should be consulted for horizon scanning the existing landscape** [Research team note: For example, it may be that some policy efforts or national programmes to improve diagnosis have already engaged in some form of horizon scanning in an area of interest]. Some existing databases could be relevant for horizon scanning, e.g. methods and data from the NIHR Innovation Observatory[3] or technology landscape analysis from MHRA's PARD[4] and the EU's Medical Devices EUDAMED[5] [Research team note: However, further research is needed to assess how far these databases cover for oncology diagnostics and how up to date they are].

- **Different methodological options may be necessary to accommodate different stakeholders' inputs in expert consultation** (HP) (e.g. face to face meetings and/or hybrid, online questionnaire surveys or interviews).
- **As mentioned earlier, it is crucial to consider which features may be relevant for a TPP early on so that different features (and combinations) can be weighted**. Some features may be more important to specify in a TPP than others, with possible trade-offs between different features' requirements, e.g. technical performance versus accessibility (AG). Understanding which features must be specified can support more focused scoping activity further down the line and help justify feature specifications in later TPP drafting (ARF).
- [Research team note: Some TPP efforts can seek formal consensus on which features to include, not just on their specifications. Whether to formally seek consensus from a wider group of participants on which features to include in a TPP may partly be influenced by how aligned the core working group's views are and how much the certainty/uncertainty and agreement/disagreement stem from the number of external consultations, resources and timeline considerations in the scoping phase].

Drafting phase

Introductory overview of the drafting phase

The drafting phase aims to provide an initial draft TPP based on insights from the scoping phase. This phase lists features relevant to the TPP alongside specifications for each (where possible from scoping phase insights), ideally including the rationale for the specification to clarify the underlying reasoning for others consulted as TPP development evolves.

Typically, the core working group driving the development of TPPs leads the drafting phase but consults with broader stakeholders to evolve each one. The appropriate stakeholders to engage can vary depending on the use case/TPP effort. However, external consultation often involves technical or subject matter experts from the research community, product development and implementation experts, industry, healthcare professionals, and patient and public voice representatives.

Two to four drafts are typical, with the fourth being the final iteration. There does not seem to be any gold standard or standard method for evolving the drafts via consultation, which can include sharing drafts for free text comments on specifications for desired features and asking for further comment on areas of uncertainty where specifying a feature has proven challenging in the scoping and early drafting phases. It can also involve feedback via interviews, workshops, posted online drafts for broader public consultation and Delphi-inspired consensus surveys. During the drafting phase, economic modelling can also help arrive at specifications for tricky evidence-poor features (whether related to desired scientific performance or using modelling to identify downstream impacts on health outcomes and clinical utility). [Research team note: Such modelling can happen during scoping or drafting phases]. It is worth mentioning that early economic modelling for TPPs has rarely been used in past practice (based on the literature on TPPs), so its applications may be more aspirational than embedded.

Key insights on the drafting phase from the workshop discussions:

- **Methods of securing external input into TPP drafts must be practical for stakeholders to engage with** (ARF, HP, IND).
- **Sharing a detailed TPP draft with stakeholders may be helpful** (covering details beyond key features), **potentially supporting targeted follow-up engagement and less pushback in later TPP finalisation stages** (HP, ARF). **Information on the underlying evidence quality is vital when sharing feature specifications to clarify the justification and reasoning to those consulted**. Understanding the available evidence is as important as understanding where evidence is limited when developing TPPs (AG).
- The drafting phase could also explore acceptable evidence types/levels to demonstrate that a novel test meets TPP requirements. However, views differ significantly about whether TPPs should or should not engage with this aspect (discussed later in this Annex).

Consensus phase

Introductory overview of the consensus phase

Based on insights from earlier project stages, the consensus phase aims to seek and achieve consensus on a final TPP draft, often concurrently with the drafting phase. Consensus is generally sought internally amongst the core working group (via workshop or meeting) before consulting a broader set of stakeholders (e.g. healthcare, academia, research, industry and patient/public voices) to establish later drafts. There are different ways of exploring consensus, and not all are suited to each type of stakeholder. Methods include Delphi surveys, interviews, comments, workshops, etc. Public views informing consensus tend to be sourced via opportunities to comment on TPPs or answer questionnaires rather than through formal Delphis.

There may be one or more consensus rounds, as determined by the core working group. Any upfront decisions they made on this may evolve, particularly if there is more uncertainty than expected, necessitating multiple rounds. Pragmatic considerations can also play a role in initial decisions on the number of rounds and methods of consensus exploration (e.g. formal surveys versus more pragmatic workshops, interview-based or questionnaire-based consultation).

Consensus is sometimes sought for earlier drafts on the features to include in a TPP (rather than on their specifications). In other cases,

consensus is only sought on specifications for features initially decided on in the scoping phase. Although consensus thresholds can vary, 75% is typical in later drafts, with some efforts using 50% for earlier drafts.

Key insights on the consensus phase from the workshop discussions:

- **Views differ on who to consult to inform TPP development versus who to formally reflect in consensus calculations** (AG). For example, some workshop participants felt that regulators, HTA, industry, clinicians and patients were essential to consult and gain consensus from, but not necessarily on the same features (AG). For example, patients can comment on features they can relate to but probably not on highly technical specifications. There was also a suggestion that consensus should be explored separately with specific groups to understand within-group consensus (AG), not only between-group consensus. [Research team note: In such an approach, it would be important for the core working group to decide whether to give different stakeholder groups equal weighting in final consensus calculations]. Others felt that industry views should be consulted to understand and assimilate their perspective but not formally counted in consensus (AG).
- **It is important to tailor stakeholder engagement to the issues each stakeholder group can best contribute to (IND, ARF). However, caution is advised not to pre-judge who can best contribute to what too prescriptively**. For instance, it was noted that opportunities for academics to be consulted on an entire TPP draft when seeking consensus are welcome, even if not all can comment on each feature/ specification. This has implications for designing a consensus phase so that invited experts can (but do not have to) comment on/ rate every feature or only those they have knowledge or experience about (ARF). Being consulted on the entire draft could help academics see how their expertise has informed earlier phases and understand the bigger picture (ARF).
- **Public consultation for consensus building is not always robust in TPP development efforts**. One workshop participant mentioned their experience that even when a TPP is posted on a website for approximately a month for public consultation (collecting very few actual comments), it seems more an exercise in due diligence than a genuine attempt to garner public voice. This observation links to a need to consider how best to engage patient and public voices in various TPP development stages (AG).
- **Different stakeholders require different options and engagement methods** (HP, PCPC). For example, asking patients to contribute to detailed Delphi surveys may not be the most appropriate method. However, they might be able to contribute to specific sections and issues (PCPC). **Surveys** offer one way to engage patients (e.g. possibly via NHS, such as texting patients after they visit an NHS setting or giving them a paper survey if they are willing to contribute). **Straightforward information styles** were also highlighted as important for different needs and understandings. For example, one attendee with dyslexia highlighted that everyone has a personal strategy and way of adapting (such as particular colour filters, paper types or fonts). Participants also stressed the importance of **letting patients talk to each other** to share their experiences and opinions. [Research team note: This is better suited to interactive input formats than surveys, e.g. workshops, focus groups, face-to-face or online engagement]. Overall, the discussion highlighted that using a range of methods and a mix of qualitative and quantitative/quantifiable inputs would be appropriate. **Religious, cultural, age-based and other differences will also impact** how to engage groups (and on what issues). For example, one participant reflected on the importance of a proper structure with culturally sensitive facilitation to bring the best out of diverse people in a meeting. Another suggested going directly to smaller groups, e.g. through surveys, while remaining sensitive and careful around entering 'safe' spaces for specific patient groups who are often not seen or heard (PCPC).
- **Transparent reporting is critical, especially on the methods by which consensus was sought, from whom, and how successfully/ unsuccessfully for each feature, and whether there were higher levels of**

uncertainty regarding some feature's specifications than others (ARF, AG, HP, PHRE). Where consensus was not reached, or the core working group considered it less important, this must be explicitly clarified and justified (ARF, HF) [Research team note: consensus may matter for some core features more than for others, where flexible specifications may be preferable, and a lack of consensus could even inform flexible final specifications].

- **The core working group will likely be tasked with final decisions on specifications and how to present and explain them.** For example, the consensus phase might not achieve the consensus threshold for all desired features (IND). This outcome must be considered when designing a robust but pragmatic approach to specifying minimal and/or optimal/preferred features (given diverse stakeholder views):
  - » For example, views may differ between different stakeholder groups. One attendee used their experience working with TPP development efforts where clinical stakeholders focused more on the optimal specification for a feature responding to unmet needs. In contrast, industry stakeholders focused on what they felt was possible/impossible. Those managing and leading the TPP development effort undertook individual interviews to determine the rationale behind views on each feature's specifications. Based on this, the core working group decided to provide information on optimal specifications from a clinical perspective and what industry considered an acceptable specification (IND). [Research team note: On reflection, while there may be cases where the core working group decides on specifications with low consensus, transparency and underlying reasons for those decisions should be explicitly clarified in all cases].
- **There were no strong views or clear consensus on what constitutes consensus**. Based on the scoping-phase insights, the most common thresholds used are 75% or 50% of members agreeing, with consensus sometimes lower in earlier drafting phases than in the latter. However, advisory group members could not comment on this in detail. Some participants noted that a 75% or higher threshold for consensus would be helpful for the final stages. The idea of a simple majority versus complete consensus was also considered (HP). One workshop participant (AG) flagged a systematic review of Delphi methods that reported that the median threshold (when specified in Delphi studies more generally) was 75% but ranged from 50–97%.[6] The core group often deals with the most controversial features/specifications by consulting with relevant field experts before making final decisions.
- **The amount of consensus exploration can vary.** Based on personal experience, one participant suggested it can be anywhere from 100–200 people via an active Delphi process (AG).

1.3. Feature-related considerations in TPPs

Based on insights from our earlier project phases, TPPs typically provide lists of features and their specifications (minimal and sometimes optimal). Diagnostic test developers must then try to meet each feature's requirements. There is considerable diversity in whether TPPs list both optimal and minimal feature specifications and whether they provide accompanying reasoning. Individual features tend to be grouped into conceptual categories in TPPs, but the method varies, partly related to differing terminologies and organisational practices. However, many of the concepts covered are similar, and a 'typology of categories' may be possible, building on prior work.[1] This could be useful in helping to standardise efforts to develop TPPs for oncology, given the lack of incumbent practice. That said, any typology should be used as a guide rather than a mandate. While specifications for individual features will always vary across TPPs, yielding unique features, some features will also likely apply across diverse test types and use contexts.

The feature categories prominent in the existing landscape of diagnostic TPPs include unmet need (features describing the unmet needs a test responds to, such as medical use or target population), analytical performance (a test's ability to correctly measure the target measurand, e.g. the least amount needed in a sample for accurate detection of disease, or features related to test robustness under different conditions), clinical utility (how the test impacts on downstream outcomes such as quality of life or mortality, etc.), clinical validity (whether the test measures an appropriate disease marker, including features related to diagnostic sensitivity and specificity), human factors (how a test and its user interact, including training needs, test handling and administration to patient), infrastructure (e.g. facilities, equipment and consumable requirements), costs and economic considerations (e.g. price per test and routes to market, market segments – health economics rarely goes beyond simple cost considerations based on our analysis), regulatory requirements and environmental impact.

Several crucial categories are rarely covered in the existing TPP landscape. Examples include features related to patient acceptability and experience, inequalities, health equity (although this is sometimes implicit in TPPs developed for tests targeting neglected diseases and low resource settings), downstream effects on care pathways and care processes, including interactions with other tests or downstream patient care requirements and referrals, broader health system and population-level outcomes (though these links are difficult to make and not consistently credible), cost-effectiveness as opposed to price considerations and the importance of meeting real-world performance needs, not just in laboratory ones.

We discuss key insights on feature-related considerations from the workshops below.

1.3.1. Key insights on feature-related considerations

The importance of evidence-based specifications for features but also for a balanced approach that does not stifle innovation:

- **There was a general acknowledgement of the importance of robustly evidence-based underlying feature specifications** to avoid clinical scepticism and present a good case for replacing an existing technology (ARF, AG). In this light, it is also essential to be clear where evidence behind a specification for a feature is weak (AG).
- **However, it is also vital that specifications are achievable/realistic and do not stifle innovation** (ARF, AG, PRHE); **the same holds for expectations about the evidence types that must be provided for specifications** (ARF, AG). Some participants described

previous experiences where TPPs have been too optimistic or aspirational; it is essential that the TPP signals the demand for realistic innovations that would fit into real-world practice (PHRE).

- **Views differed on how flexible specifications need to be. There was a general recognition of the imperative not to impede innovation but also a need for sufficient steer and clarity** (ARF, IND, AG, PRHE). There was agreement that too strict parameters (e.g. 100% sensitivity and specificity) can stifle innovation by disincentivising innovators, meaning patients may not benefit from the potential technologies they could have developed. Thus, it is essential to consider what is achievable and realistic yet also clinically important (AG, PHRE). [Research team note: Specifications for some features may end up more prescriptive than others that require more range and flexibility. Decisions on what is needed are likely to be TPP and context-specific].
- **It may be necessary to prioritise which features need detailed specifications and which can be more flexible at the innovator's discretion, depending on a test's use case and the available evidence base** (ARF, AG). Many attendees acknowledged the need to keep TPP features open enough to spur innovation while specific enough to address a fundamental value proposition (ARF).
- **Views on whether TPPs should specify minimal and optimal feature requirements also varied. Participants highlighted that specifying 'minimal' requirements is necessary for key desired features and that this may be easier to do than for optimal requirements** (PHRE, IND). **Specifying 'optimal' requirements may not be possible for some features. Still, some workshop participants suggested introducing 'preferred' requirements (where possible) as a more appropriate way forward** (PHRE, IND). Some participants in the advisory group workshop felt that minimal and optimal specifications would be essential for some features, but others could have more room for flexibility (AG).
- Opinions on whether TPPs should specify acceptable evidence levels/evidence types and sources for demonstrating that a novel test meets TPP requirements differed within and across stakeholder groups (ARF, AG, IND). Some workshop participants emphasised the importance of the robustness of the supporting evidence for feature specifications included in a TPP. However, there was also discussion about the role of other functions and resources in the system (such as regulatory and HTA agencies to clarify what acceptable evidence levels are for diagnostic tests). Some felt that evidence requirements within a TPP would be redundant (provided this is clear in the evidence sources/guidelines from regulatory agencies, for example, and HTA). Some participants suggested that it should be left to manufacturers to determine appropriate evidence levels for their innovations, given regulation and HTA expectations. They commented that this is more conducive to innovation because it allows enough freedom, especially when it is difficult to know what an acceptable threshold should be. Some felt that additional evidence requirements within a TPP could be too restrictive for innovators because they would have to examine regulatory guidance *and* TPPs. However, there were mixed views on this. Some participants noted that TPPs could provide an opportunity to set appropriate evidence levels clarifying the reference standard and how to meet it for a specific subset of products. Such standards would need to consider the diversity of potential innovators and be feasible and not overly onerous for SMEs to respond to, ensuring a level playing field between larger and smaller companies (IND).
- However, evidence reporting standards should be clear to innovators responding to a TPP, including transparent sharing of study protocols, data sources and results (PRHE) (regardless of whether a TPP or documents from HTA agencies clarify the diagnostic evidence requirements).

General insights on the feature types specified by attendees:

- **All features included in a TPP must relate to the diagnostic's main intended benefit or value proposition** (AG), which must be clearly articulated in a TPP relative to the features covered and their specifications (ARF, AG, PRHE):

  - » **It is also helpful to specify what is not wanted.** TPPs are often framed as documents that show 'what is wanted', but it could also be useful for innovators to understand what is undesirable and worth avoiding (ARF, IND). A few attendees mentioned the need to identify a diagnostic's potential toxicity, dangers, harms and invasiveness and link that to information on safety requirements in a TPP (ARF, IND).

- **A TPP must look beyond scientific and technical performance features only to consider features related to the adoption context, clinical care pathways and diverse target population needs** (ARF, AG, HP, IND). A TPP must ensure good technical performance and cover related features (HP), but it is not sufficient to focus on technical performance only:
  - » For example, some workshop representatives flagged the importance of features related to **accessibility** (HP, PCPC, ARF), not just technical performance, and some use cases may involve trade-offs between accessibility and technical performance (ARF).
  - » Both **patient acceptability** (PCPC, HP) and **clinician acceptability** (HP) matter. Some tests are unacceptable to some patients due to discomfort or other reasons (HP).
  - » Participants also considered features related to **clinical utility** (e.g. impacts on care pathway and patient prognosis) important in articulating a value proposition (ARF).
  - » **Time-to-result feature specifications** are necessary for health professionals and patients to know when to expect a result (HP). Another practicality-related consideration raised concerns about the push/pull of timely access to results, i.e. whether a test result will be 'pushed' through to clinicians or whether they must actively seek it (HP).
  - » **Information on how results are displayed and can be interpreted and communicated** can help support clinicians using tests (e.g. features around result display or format, interpretation and communication) (HP).
  - » Participants also considered features that can help support products that address **challenges to test integration with NHS care pathways, systems, data and IT infrastructure** as important (HP). This consideration relates to information on other tests a patient may need to undergo beyond the test for which the TPP is being developed, how the test can inform further patient care decisions, and how the results will integrate into NHS IT systems and patient records. The NHS can struggle to connect current systems and test results, link different test results and ensure effective care. A TPP should specify how a specific test would integrate and work with data and IT infrastructure and other existing tests (HP).
  - » Broader requirements associated with **infrastructure** (e.g. space, equipment, facilities and IT) must be in place for testing and **transport** (HP).
  - » **How tests relate to subsequent treatment decisions** [Research team note: we are unsure whether this is within a TPP's scope. However, it could link to features on the medical decision a test informs] (HP).
  - » Clarifying **healthcare training needs for features related to human interaction is critical** in a TPP for novel tests, especially given the fast pace of technology development (e.g. genomic tests, AI) (HP).
  - » **Clinical validity** measures matter (e.g. positive predictive value and result validity in different populations) for scientific and technical performance. This issue also relates to measures of **analytical performance** because a test will have no buy-in if there is low confidence in its results, which can be a challenge for some tests (HP).
  - » Participants also raised **inequalities and equity** (HP, PRHE), i.e. whether performance is equally valid across a population. Some attendees noted that TPPs must be explicit about inequalities, differences or unequal performance, and ensure (where possible) that novel tests do not exacerbate disparities.

  - » **Broader social and ethical considerations must be reflected in feature specifications** (PHRE) related to data protection, mitigating inequalities, etc., and ensuring tests are not biased towards accuracy in particular populations only.
  - » Participants also highlighted that specifications that consider **real-world performance (not just lab conditions)** are essential as these features can influence test uptake by clinicians (HP).
  - » **Numbers related to the expected testing volume** (HP) [Research team note: this may link to information on the target population or anticipated market size, which matters for incentivising industry innovators and indicating supply volume expectations).
- **Views differed regarding the need for and feasibility of specifying requirements related to cost-effectiveness and the value of early economic modelling, even within the same stakeholder groups** (IND, AG, ARG).
  - » **Some participants felt that cost-effectiveness is important for a TPP and that early economic modelling is a valuable tool** (IND), **seeing early economic modelling as helping to identify the test features (or feature combinations) that most impact its cost-effectiveness** (PHRE). **Hence, early economic modelling could help articulate a test's downstream benefits** (e.g. generating savings in cancer treatments, hospital beds and treatments (IND, HP, PHRE) and rule out tests unlikely to be cost-effective (PHRE). Any new diagnostic test would have some cost burden and thus need a clear benefit to justify its implementation within the clinical pathway. Indications of cost and cost-effectiveness could also allow affordability considerations to be factored into a TPP early on (AG, ARF).
  - » **Early economic modelling was also seen as one way to help inform the evidence requirements for cost-effectiveness in a TPP relative to HTA** (AG, ARF, IND) **and perhaps bridge the disconnect between test development and reimbursement** (PHRE).
  - » **However, there were some concerns that the nature of current evidence requirements for diagnostics (being too drug-focused in the eyes of some participants) compromises the ability to demonstrate value effectively through cost-effectiveness analyses**. Thus, a TPP would help specify cost-effectiveness requirements only if there were appropriate requirements and clarity on them in the first place, which some participants felt was lacking (IND). Some participants advised caution in specifying costs/pricing in a TPP (AG) because the required economic modelling would be full of uncertainties initially (AG). The uncertainty in early economic modelling also makes confidence in borderline cost-effective tests challenging (PHRE).
  - » **Participants flagged a need for a broader focus in early economic modelling beyond sensitivity and specificity considerations alone** (PHRE). In addition to sensitivity and specificity, greater emphasis on accessibility, speed, and time would be helpful as these impact the patient's overall experience and the chance of success. The impact of these features on people – and on cost-effectiveness – could potentially be assessed.
- **Some participants suggested TPPs should consider features related to market considerations regarding clarity on routes to market** (IND). **Related to this, it would be helpful for innovators if a TPP could clarify what types of reimbursement pathways might apply to a potential product, perhaps in the context of information on routes to market that a test should consider** (IND) – even though participants acknowledged that a TPP is a specification document and cannot guarantee a product's successful uptake.
- A participant suggested that industry would benefit from guidance on features related to bringing an innovation to market. [Research team note: Despite a desire for information on guaranteed routes to market, decisions

about adoption extend beyond TPPs' role in providing information. At best, TPPs could provide information on potential routes to market unless there was a policy incentive such as a guaranteed purchase fund for a resulting competitive innovation].

- **As mentioned earlier, identifying trade-offs between feature specifications is vital** as not all features can simultaneously be met to ideal standards (AG, PRHE).
- **Since many features are interdependent, the combination of features matters as much as individual characteristics**. For example, clinician acceptability can be influenced by perceived clinical utility, time to result, ease of test displays and interpretation and fit with broader NHS IT infrastructure (HP). Patient acceptability is influenced by diverse factors such as accessibility, time to result, invasiveness and others (HP, PCPC).
- **Regulatory requirements must be clarified** (AG, PRHE, IND); **a lack of regulatory clarity is a key challenge for innovators, especially since Brexit** (IND). As discussed earlier, TPPs may need to clarify UK regulatory criteria (MHRA) and check international regulatory requirements for key jurisdictions, too, given that industry often develops products for multiple markets (AG, IND). Negotiating and meeting the necessary requirements for eventual market access in the UK can be challenging. A TPP could help outline the key requirements across the development cycle, increasing the likelihood of success (PRHE). **Considering regulatory clarity in light of differences between the devolved nations in the UK also matters for industry** (IND).
- **HTA assessment-related requirements that impact cost-effectiveness assessments also need to be clarified in a TPP (or at least by HTA agencies) and mitigate against requirements for assessing drugs when/if they are inappropriate for diagnostics** (IND). **Some participants mentioned challenges with inadequate clarity on HTA diagnostic requirements, given that NICE HTA in this area is less mature than for drugs** (IND). Some felt that NICE evidence requirements for diagnostics are not distinctive enough vis-à-vis drug-focused requirements and are still geared toward a pharmaceutical product. Examples concerned the requested outcomes data (e.g. mortality and morbidity-related outcomes), and there was a view that more bespoke indicators should be considered, e.g. changing diagnostic rates or decreasing late-stage cancers (impacted by earlier diagnosis and thus earlier/more timely treatment) (IND).
- **While many features may 'universally' apply to different oncology test types (even if actual feature specifications will differ), unique considerations must be borne in mind for bespoke future TPP development efforts, especially for novel test types** (like those using artificial intelligence [AI]) (AG). Many currently considered features derive from in-vitro diagnostics (though not all).
    - » Some areas, like AI, will need specific feature types for screening, format and display (AG). [Research team note: accuracy specifications will be unique in ensuring the algorithms underpinning the AI were developed for appropriate participants].
    - » Some feature requirements may be unique to a UK context, such as the NHS Net Zero criteria for procurement requirements and other environmental impacts (AG).
- **Demonstrating the clinical utility and economic value of early diagnosis in oncology is challenging** (PRHE). A barrier to translating innovation in oncology diagnostics to real-world practice is the difficulty in demonstrating clinical and economic value. Earlier diagnosis can often lead to a stage shift in cancer identification, which has often been established. However, this does not always necessarily lead to improved survival and health outcomes if no appropriate treatment or behaviour change can alter the course of the disease (i.e. does not alter the clinical care pathway). Thus, it can be difficult for some oncology diagnostics to demonstrate sufficient value to be considered cost-effective (PRHE).

The often-underexplored patient perspective on features and their input into TPPs:

- **A series of important patient-experience considerations should inform the features considered in a TPP development effort** (AG, PCPC). These include clarifying:
  - **Specifications relating to patient accessibility** (PCPC): These are test features related to access needs for diverse populations, inequalities and access routes.
  - **Specifications relating to human aspects such as required patient-clinician communications and patient care-pathway interactions** (PCPC), e.g. how a patient undergoing the test can effectively interact and communicate with healthcare professionals around gaining access, understanding what to expect from the testing process, any potential side effects, and receipt and interpretation of test results.
  - **How a test fits within patient care pathway** (PCPC): This concerns information on how the test must fit within a patient's care pathway, i.e. what type of health service pathway and organisation it must align with to support coordinated, timely and accurate diagnosis and an appropriate patient experience.
  - **Specifications explaining requirements for mitigating or managing inequalities:** (PCPC) These may relate to access, affordability, accuracy or other inequalities related to, for example, cultural sensitivities (PCPC).
  - **Specifications related to the speed and efficiency of testing processes and test-result turnaround times**: (PCPC, AG) These can impact patient anxiety and experience and influence the ability to access treatment as quickly as possible.
  - **Specifications related to test accuracy**: These are important for confidence in a result and the ability to secure appropriate treatment (PCPC).
  - **Specifications related to test invasiveness** can also impact patient experience.
  - **Eligibility criteria specifications that do not exclude patients needing screening and testing:** This can be challenging with rare cancers or for some patients due to age and other screening eligibility criteria.

Terminology:

- **Terminological clarity is essential for an unambiguous understanding of what certain features and their specifications actually mean in TPPs.** Terminology is sometimes mis- or inconsistently used in TPP development (see Annex A), e.g. 'analytical sensitivity' or 'screening.' There may be scope for considering an International Organisation for Standardisation (ISO) or otherwise standardised definitions (AG). [Research team note: in prior scoping documents, we have suggested the need for future work to develop a standardised set of terminology and feature definitions, at least for those likely to be relevant across multiple TPP contexts].

1.3.2. Prioritisation considerations

Summary of key insights:

- The workshops also examined potential criteria for future prioritisation of which TPPs should be developed. The identified criteria related to the nature and degree of unmet need:
  - **Cancers for which current diagnosis is poor, i.e. low screening rates** (AG, HP), **poor early diagnosis** (PCPC, HP, PRHE) or **inaccurate diagnosis** (PCPC), and **where current tests are inadequate** (HP). Participants suggested that it would be important to consider:
    - **Tests for earlier detection and diagnosis** (PCPC, PRHE, IND) to help patients secure earlier treatment, including better screening tests/ programmes.
    - **More accurate diagnosis** (PCPC), including **tests for better triage** (HP). **Tests that repeat the need for multiple other tests** (HP) **are better in rule-in/rule-out decisions** (HP).
    - **More accessible diagnosis** (AG, PCPC, HP), addressing access

inequalities and relating to an earlier point about improving access for hard-to-reach populations (HP).

  - » **Cancers with low survival rates** (ARF, IND) **with a clear association with diagnosis**, e.g. earlier or more accurate diagnosis leads to earlier treatment and increased survival chances.
  - » **Incidence and prevalence considerations** (ARF, HP) – although participants highlighted the importance of **not biasing against rare cancers** if there is an unmet diagnostic need (ARF, AG, IND).

- **Inequality-related considerations** (ARF, AG, PCPC, HP, IND) were discussed, namely deprived populations and those experiencing inequities, which need consideration when prioritising the focus of the first cancer-related TPP in the UK. This ensures that any innovation helps address unmet needs without introducing or reinforcing inequalities (ARF, AG), whether in access, affordability, or effectiveness. Related to this, **participants also flagged social, ethical and equity considerations** (PRHE).
- **Where there is a need for improved patient experience:**
  - » **Tests that can provide quicker results** for use cases where this is an issue. There is a challenge in developing improved novel tests or solutions to quickly rule cancer out (or in), especially for those in primary care, thus supporting triage and reducing capacity pressures associated with the more detailed diagnostic testing stage (PCPC, HP).
  - » **Tests that are less invasive for patients** (PCPC, HP).
- **Early diagnosis of first occurrence and recurrence where this is an issue** (PCPC).
- **The ability of improved diagnosis to reduce over-intervention**, e.g. where unnecessary biopsies are conducted or there is over-diagnosis/treatment (ARF).
- **Confirming the need for a novel test (and hence a TPP for it) should involve an understanding of how a novel test may help with workforce capacity constraints in the NHS or add to them** (HP), as this can affect the uptake of any resulting diagnostic and hence a TPP's ultimate impact.
- **Prioritising TPPs offering the greatest economic and clinical value (population *and* individual-level, including workforce impact)** (IND, PHRE).
- **Alignment with policy** is essential, especially for taking TPP and diagnostic development forward (IND). However, a point was made that alignment with policy may risk prioritising diseases that affect a larger portion of the population instead of rare diseases, which must be mitigated.
- Health technologies with the potential **for more medium-term impacts** (IND). [Research team note: The participant who mentioned this point did not clarify what they meant by medium-term as opposed to short or long-term].
- Cancer diagnostic programmes **that lack established benchmarks or standards** or a clearly defined patient pathway (with no existing NICE guidance) (IND).

1.3.3. Other points to note

Nascent diagnostic areas and evidence gaps in the AI field:

- **Many participants mentioned AI as a critical diagnostic that is quickly gaining traction, bringing specific TPP-relevant considerations into play**. Methodological and ethical research on AI is scarce, limiting its potential integration into diagnostics that are capable of making it to market within NICE guidelines and NHS infrastructure (e.g., **implications of the biases** from closed AI systems). Such considerations would need attention in a related set of TPP **requirements on the AI tool's development method and the data on which it was based**. AI data is often trained on research participants, meaning the data is frequently biased and may not apply to broader populations (ARF).
- **Some areas (like AI) will need specific feature types for screening, format, and display features (AG). [Research team note: This will require accurate**

specifications ensuring the AI's underpinning algorithms were developed for appropriate participants]

- **AI must be considered as a diagnostic aid or triage tool, and both must be considered in any TPP that involves it.** This is especially the case in pathology digitisation efforts, i.e. the medical decision that is determined would need to be explicitly clarified in any TPP using AI (ARF).
- **A couple of attendees also mentioned the challenge of regulating** something as self-directed as AI, with implications for regulators' capacity for constant amendments as AI changes. This would make it challenging to clarify regulatory requirements in a TPP in such a fast-paced and evolving regulatory landscape (ARF).

Diagnostic platforms versus individual tests:

- Some attendees highlighted that **diagnostic platforms raise unique considerations for a TPP compared to individual diagnostic** tests (IND). One participant used the example of genomics testing, suggesting that clarity on the need for tests to provide evidence about which biomarkers have evidence-based clinical utility and which do not would be helpful (as some tests span multiple biomarkers when one considers platforms). As demand signals, TPPs (as opposed to supply-side documents) may want to consider treating some diagnostic platform/multianalyte tests like assays in terms of the evidence requirements behind them (IND).

References

1 Cocco, P., Ayaz-Shah, A., Messenger, M. P., West, R. M. & Shinkins, B. 2020. 'Target Product Profiles for medical tests: a systematic review of current methods'. *BMC medicine* **18**, 1-12.

2 World Health Organization. 2023. *Performing a landscape analysis: understanding health product research and development*. As of 03 April 2024: https://iris.who.int/bitstream/handle/10665/372696/9789240073319-eng.pdf?sequence=1

3 NIHR Innovation Observatory (Homepage). 2023. *NIHR Innovation Observatory*. As of 22 March 2024: https://www.io.nihr.ac.uk/

4 Medicines and Healthcare Products Regulations Agency. 2023. *Public Access Registration Database (PARD)*. As of 22 March 2024: https://pard.mhra.gov.uk/

5 European Commission. *Medical Devices - EUDAMED*. As of 22 March 2024: https://health.ec.europa.eu/medical-devices-eudamed_en

6 Estrela, M. et al. 2021. 'Validation of the eHealthResp online course for pharmacists and physicians: A Delphi method approach'. *Biomed Pharmacother* **140**, 111739. As of 22 March 2024: https://doi.org/10.1016/j.biopha.2021.111739

Annex D: Cross-Analysis of Frontline Interviews with General Practitioners and Pathology/ Genomics Laboratory Experts

Authors: Mark L Cabling, Jessica Dawney, Matthew Napier, Zuzanna Marciniak-Nuqui, Fifi Olumogba, Larry Kessler, Amanda Cole, Lotte Steuten, Sonja Marjanovic

1.1. Introduction

Annex D is the fourth of eight annexes complementing the main Cancer Research UK-funded project's final report: 'Advancing the development and use of diagnostic target product profiles for cancer.' The not-for-profit research institute RAND Europe led the project in collaboration with the Office of Health Economics. The project has benefited from ongoing support and advice from Professor Larry Kessler (University of Washington), a key consultant on the work. This document provides detailed findings and analysis of the frontline clinician interviews to which the final report refers; thus, Annex D, like all the other annexes, is primarily meant to accompany the final report and is not meant to be read as a standalone document.

Context and aims of interviews with GPs and pathology/genomics laboratory experts

Supported by Cancer Research United Kingdom (CRUK), RAND Europe and the Office of Health Economics are conducting a project to advance practical knowledge on possible approaches in developing diagnostic Target Product Profiles (TPPs) for cancer. As part of the project, we conducted a series of interviews to better understand how end-users of any potential test resulting from a demand signalling TPP experience challenges with cancer diagnosis in the National Health Service (NHS) and associated areas needing improvement.

The interviews also aimed to gain insight into issues that need consideration in a demand-signalling TPP for cancer regarding the key features to specify. Finally, the interviews also aimed to understand how General Practitioners (GPs) and diagnostic laboratory experts (in both pathology and genomics) could best contribute to any process of developing a diagnostic TPP for oncology in the United Kingdom (UK). We conducted the interviews to complement insights gained through the stakeholder workshops based on CRUK's particular interest in better understanding gaps from these stakeholders' perspectives as potential end-users of possible innovative tests in future primary care. This aim was related to CRUK's identification of a gap in diagnostic testing for cancer in primary care and interest in complementing insights from the workshop conducted with healthcare professionals.

We interviewed nine GPs and four laboratory experts (three pathology laboratory experts and one genomics laboratory expert) between 14 June and 14 July 2023. The interviews were semi-structured and conducted online via MS Teams. All individuals participated with informed consent. We recruited GPs via convenience sampling through CRUK primary care panels and laboratory experts through CRUK, research team and advisory group networks. One caveat of the interviews is that all interviewees were based in England, and convenience sampling meant that diversity considerations were not factored into recruitment. However, it is worth noting that these insights complement workshop-based stakeholder consultation, in which we invited clinical perspectives from individuals with diverse roles and from diverse parts of the UK.

This annex provides an overview of key insights from the interviews. We aimed to gather a qualitative account of how those on the NHS frontline could best contribute to developing a diagnostic TPP for oncology. The aim was not to capture diverse views and experiences nor to quantify the strength of sentiment. Given their different roles in the diagnostic pathway, GPs and diagnostic laboratory experts emphasised different points, as expected. All interviewees focused on ensuring that any test to be developed due to a demand signalling TPP could fit within existing health system infrastructure and pathways. Given their patient-facing role, GPs focused on needing tests to effectively support accurate and early diagnosis while helping reduce workforce pressures, time constraints and patient backlogs. Diagnostic laboratory experts flagged the backlog of specimen samples and the need for novel tests to fit with the existing diagnostic laboratory infrastructure, workforce, physical infrastructure constraints and training needs. They also highlighted the need for, and potential in, automated diagnostic laboratory processes. All interviewees also felt it is important for GPs and pathology laboratory practitioners to participate in TPP development processes.

When citing interviewees to attribute findings to, we use 'GPINT' to indicate that the interviewee was a GP and 'PLINT' to indicate when the interviewee was a pathology or genomics laboratory expert. We occasionally offer additional research team reflections on the interview insights throughout the document. When doing so, we clarify that these are our reflections using the term 'Research team note.'

1.2. Areas needing innovation and improvement in diagnostic testing for cancer

1.2.1. Informing insights on the test types and features needed

Interview insights shed light on the types of cancers and use cases where NHS diagnosis-related challenges are particularly acute and where GP and diagnostic laboratory experts perceive a high need for improved diagnostic tests.

1) Developing tests that can be used for early detection and diagnosis in pre-symptomatic individuals

GPs considered cancers that present vague, non-specific, undifferentiated or non-deterministic symptoms that can lead to late presentation, worse prognoses and high mortality rates as a key challenge to timely diagnosis (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT7, GPINT14). Such cancers include any tumours or lumps that cannot be seen or felt. Some examples mentioned include lung (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT14), pancreatic (GPINT1, GPINT4, GPINT5, GPINT6), abdominal (GPINT1, GPINT6, GPINT14), brain (GPINT1, GPINT6), colorectal (GPINT4), ovarian (GPINT5, GPINT6) and haematological (GPINT6) cancers.

Some interviewees highlighted a need for innovative diagnostic tests that can focus on pre-symptomatic and pre-hospital parts of the patient care pathway and facilitate early diagnosis as particularly important (GPINT1, GPINT4) **to help address the challenges associated with cancers where symptoms only show at later disease stages.** Such diagnostics could include tests to screen asymptomatic and at-risk individuals (GPINT1). According to one interviewee, this could include self-administered at-home tests or those they can access via primary care (GPINT1). However, this would be more challenging for patients with cognitive impairment or physical difficulties (GPINT1, GPINT5). [Research team note: We advise caution with any considerations of at-home tests for cancer for any population group because a cancer diagnosis can be a challenging emotional experience requiring professional healthcare presence rather than a patient receiving the news when alone. Additionally, healthcare professional support is vital for interpreting the results and their implications]. Two GPs felt that a blood test that could detect cancer would be the 'holy grail' of early cancer diagnosis, especially if it could identify cancers in asymptomatic people (GPINT4, GPINT7). Two GPs also flagged that the pathway towards alternative diagnostics is unclear and inconsistent when a patient has been deemed ineligible for a particular diagnostic (GPINT12, GPINT14).

2) Developing improved tests for ruling cancers out as well as in

Some GPs suggested that tests able to rule cancer out as well as in would be helpful in primary care settings (GPINT4, GPINT6, GPINT7, GPINT14). Some emphasised that a rule-out test is just as valuable as a rule-in test because patients often present with imprecise, vague symptoms rather than according to guideline symptoms (GPINT6, GPINT7, GPINT12, GPINT14). One example given was the use of the Cytosponge test for Barrett's oesophagus (a condition which can lead to oesophageal cancer) in secondary care, which could potentially be used in primary care as a rule-out test rather than referring the patient to hospital for something more invasive like endoscopy (GPINT4). A simple rule-out test would enable GPs to sift through patients quicker, helping avoid unnecessary invasive biopsies and reduce the pressure on secondary care. Related to this, some interviewees flagged the need for tests with improved sensitivity, i.e. the ability to reduce false positives (particularly

for supporting early diagnosis of cancers that only present when already advanced) (GPINT1, GPINT3).

3) Developing tests that could address broader adoption issues regarding accessibility and acceptability

Many GPs flagged the need for any innovation to be as easily accessible to the patient as possible (GPINT2, GPINT4, GPINT5, GPINT6, GPINT12, PLINT10). One GP framed accessibility in terms of opening up the pathway to diagnostic testing so that patients could refer themselves if they meet specific criteria (GPINT2).

Interviewees also flagged examples of tests that currently have issues with patient acceptability (GPINT1, GPINT2, GPINT4, GPINT6). **Although interviewees considered patient acceptability important, they noted that achieving a test's desired patient acceptability must be balanced against achieving the GP's desired confidence in a test** (GPINT1, GPINT2, GPINT4, GPINT6)**. There can be trade-offs between achieving accuracy and acceptability.** The test's accuracy is vital because it is linked to trust in a test and healthcare professionals' confidence in the results (GPINT2, GPINT4, GPINT6). However, patient acceptability is essential to ensure a test's uptake (GPINT1, GPINT4). One GP felt that accuracy might trump acceptability for certain aggressive cancers. However, patient acceptability might trump accuracy for less aggressive cancers, especially if this would mean an easier pathway for patients (GPINT2).

Clinician acceptability partly depends on their trust in a test (GPINT2, GPINT4, GPINT6) **and its ease of use** (GPINT6). A typical GP only has a ten-minute consultation and no specialisation in cancer, so any test must be easy to use without specific oncology knowledge (GPINT6). A related issue is the ease of interpreting the test results and the ability to communicate them effectively. A test's interpretability and how to communicate the risks of an incorrect diagnosis matter (GPINT7, GPINT12, GPINT14), especially regarding what to do with a negative result if there are still suspicious symptoms (GPINT2, GPINT7, GPINT14).

4) Developing tests for a primary care setting

The pressures primary care faces in terms of demand for services, workforce capacity and the nature of existing diagnostic tests and pathways imply a need for improved tests that could be used more widely in primary care, are accurate and are easy for primary care professionals to administer/use and understand (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT12, GPINT14). GPs flagged challenges related to a lack of workforce capacity, resources and support to train primary care staff to perform tests, interpret, and/or act upon test results (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT12, GPINT14). One interviewee flagged the need to help GPs deal with current risks associated with diagnosing patients presenting to primary care, including current cognitive overload and capacity challenges (GPINT3). Two GPs emphasised that general practice is the 'first port of call' in the NHS but does not have easy, quick and direct access to either radiology or pathology services (GPINT6, GPINT12). As radiology is separate from primary care facilities, radiology referrals are time-consuming and often necessitate the patient facing logistical barriers to physically accessing radiology services (GPINT6, GPINT12) [Research team note: These two interviewees' implication was that some radiology imaging services should be included in primary care facilities]. Similarly, two GPs suggested that pathology services are linked to primary care too far down the care pathway (GPINT6, GPINT12) [Research team note: This implies that barriers to pathology access need addressing].

5) Multianalyte tests

Some interviewees highlighted the potential value of developing multianalyte tests capable of simultaneously testing for multiple cancer types for use in early diagnosis and complementing the types of screening assays compatible with what genomic laboratories already do (GPINT1, GPINT4, PLINT11). One interviewee noted that, from a histopathology perspective, it could also be helpful to have two separate 'pan-multianalyte tests' for men and women, with both tests still having cancer biomarkers common in either sex (PLINT8). However, using the example of circulating tumour DNA testing, one GP suggested that some novel tests designed to detect multiple cancer types still tend to pick up later-stage rather than early-stage cancers (GPINT4). [Research team note: There are ongoing trials of novel blood tests for detecting multiple cancers to see if they can aid with early detection, e.g. in asymptomatic individuals and when combined with standard cancer testing].

Multianalyte blood tests that could test for multiple cancers from a blood sample may also help respond to the need for tests that can be administered in primary care screening and are user-friendly for GPs (GPINT4, GPINT7). **However, interviewees voiced some concerns about aspects needing awareness and consideration** (though not all unique to multianalyte tests alone). Examples included:

- **The potential risk of multianalyte tests under-diagnosing cancers, given concerns over their accuracy** (GPINT7, GPINT14).
- **A need for expectation management for patients and physicians regarding multianalyte tests** (GPINT5, GPINT12). One GP clarified that their concern about managing expectations centred on the complexity of the multianalyte test output, which revolves around risks and probabilities (GPINT12). These are 'notoriously' difficult to communicate to patients, especially within a ten-minute appointment, although GPs can arrange longer appointments if they feel it appropriate. It is thus essential to factor in the time needed to do the tests and then deliver the results/diagnosis, particularly if there is a need for specialist support and difficulties with delivering news during time-limited appointments (GPINT12).
- **Downstream ethical implications for the patient, especially as multianalyte tests could pick up conditions other than cancer.** There could be unexpected findings from a multianalyte test, potentially disadvantaging the patient (e.g. when reporting on health or medical insurance in the system) (GPINT5, GPINT14). An example is lung screening, which picked up people with coronary calcification; this was not the original intention of the lung scan, and there was no clear treatment pathway for it (GPINT14). It is critical that a patient knows exactly what they consent to (PLINT9).
- **Ensuring tests are used appropriately, e.g. regarding collecting the required sample size.** One interviewee mentioned that the more tests are undertaken on a sample, the more attention must be paid to ensuring sufficient sample collection. (PLINT9).
- **Ensuring sufficient laboratory capacity and skilled staff to deliver the test meaningfully** (GPINT5, GPINT12, PLINT10). Related to this is ensuring appropriate sample storage capacity (and appropriate durations), special handling requirements, pre-analytics and fitting in with further downstream processes (PLINT10).

1.2.2. Examples of existing cancer diagnostics considered suboptimal

Interviewees highlighted various cancer types for which they saw a need to improve existing tests and/or diagnostic pathways (GPINT2, GPINT3, GPINT7, GPINT12, GPINT14). These include:

- **Ovarian cancer:** More specifically, the CA-125 blood test for ovarian cancer was considered sub-optimally sensitive or specific (GPINT3, GPINT7), and it is unclear whether the test is used appropriately in conjunction with ultrasound (GPINT14) in existing care pathways.
- **Prostate cancer:** Limitations with Prostate Serum Antigen (PSA) blood tests were mentioned, e.g. that they do not provide information about how deadly the cancer is (i.e. the likelihood of associated mortality) (GPINT3, GPINT7, GPINT12). [Research team note: Other limitations identified in the literature,[1] such as low accuracy (e.g. three in four men with raised PSA levels will not have cancer), the PSA test cannot distinguish between fast and slow-growing cancers and thus can cause unnecessary worry].
- **Tests for upper versus lower gastrointestinal cancers**, since tests such as colonoscopies have low rates of patient acceptability (due to the invasiveness of the test), linking to the earlier points about accessibility and acceptability (GPINT3, GPINT14).

Interviewees also highlighted improvements needed for different diagnostic test types, i.e. different technological approaches, highlighting imaging techniques (GPINT2, GPINT7). They mentioned that Artificial Intelligence (AI) could fast-track chest X-ray processing (GPINT2); this is critical for lung cancer diagnosis and CT scans (applicable to diverse cancers), for which the NHS has limited availability (GPINT2).

1.2.3. Innovation-related infrastructure implications in cancer diagnostics

Our interviews flagged some insights about the broader healthcare capacity and infrastructure that any efforts to develop new tests must consider achieving uptake and integration. We elaborate on these insights below, specifically in informing the types of diagnostic tests needed and the types of infrastructure development that the health system may need to integrate them. We flag those aspects raised by the interviewees specifically but acknowledge that there may be other relevant aspects that did not arise in the interviews. These interview insights are also relevant in informing how TPP development efforts approach considerations related to infrastructure requirements and specifications.

1) The need for tests that can fit with wider histopathology and genomic laboratory infrastructure

Any efforts to develop a new test must consider how it will be implemented in laboratories and how it would work within, be integrated into and/or support existing pathology and genomic laboratory processes and infrastructure (PLINT10, PLINT11). According to one pathology laboratory expert, histopathology faces a 'crisis' with a considerable country-wide backlog (PLINT10). Other experts flagged general workforce capacity/staffing, training and infrastructure challenges (such as space) in histopathology and genomic labs (PLINT8, PLINT9, PLINT10, PLINT 11). Both pathology and genomics laboratories often want to innovate and look at new tests/diagnostics, but it is difficult to do so under current time and monetary constraints (PLINT9, PLINT11). A laboratory's design will also impact what tests they can or cannot implement (PLINT10). Considering the types of samples needed is another important factor (PLINT10).

2) Needs for improved automation of diagnostic testing

Pathology laboratory experts flagged improved automation of diagnostic testing (i.e. routine sample processing) as important, as it would support pathology laboratories with current work capacity constraints and help address current backlogs (PLINT8, PLINT9, PLINT9, PLINT10). An overall quicker process within histopathology would also be beneficial (PLINT8). Although new technologies have emerged, they tend to be very costly (PLINT9).

Increasing automation (of tools, processes, routine work and result delivery) is something pathology laboratories are considering (PLINT8, PLINT9, PLINT10). A large proportion of histopathology work is done through human intervention and is thus a manual process for staff (PLINT8, PLINT9).

However, automation does not necessarily fix the problem in and of itself because it only plays one part in addressing the challenge of how the whole histopathology pathway can be redesigned to accommodate the growing workload. As one interviewee explained, adding equipment will not by itself solve the problem, as there are upstream and downstream processes that also need to be adapted and changed (PLINT10).

1.3. Insights on key features to consider specifying in a TPP

Interview insights point to the importance of specifying both technical performance requirements and a broader set of features that can help support the fit of any resulting test with clinical and care pathways and health systems infrastructure, and different interviewees highlighted diverse features **(**GPINT1, GPINT2, GPINT3, GPINT5, GPINT6, GPINT7, GPINT14, PLINT8, PLINT9, PLINT 10, PLINT11).

Many flagged that features related to technical performance, such as a test's desired accuracy and validity, are as essential to consider as features supporting its availability and accessibility (GPINT1, GPINT2, GPINT4, GPINT5, GPINT7). While some felt that technical features are necessary for a TPP (which we elaborate on in the following paragraphs), one interviewee noted that other bodies sometimes already clarify those specifications, implying they may already be stated and known, e.g. the Royal College of Radiologists (GPINT12). [Research team note: Such evidence could be a source to inform specifications on technical features in some TPPs].

Interviewees highlighted the following feature types and feature-related considerations as essential to consider specifying in a demand-signalling TPP informing the development of novel diagnostic tests for cancer:

- **The test's target population** (GPINT1, GPINT5, GPINT6, GPINT14, PLINT11). Certain populations will have specific challenges (GPINT1, GPINT5, GPINT6, GPINT14, PLINT11). It is as important to identify who the test is *not* for as who it *is* for (GPINT6). Interviewees noted that diverse considerations such as socioeconomic deprivation (GPINT1, PLINT11), gender (GPINT1), physical and cognitive disabilities (GPINT1, GPINT6, GPINT14) and cultural sensitivities (GPINT5) should be accounted for. For instance, with the Faecal Immunochemical Test (FIT), it is unreasonable to give a kit to a blind patient (GPINT6). People with disabilities must be included in any diagnostic recommendations (GPINT1).
- **Eligibility criteria and clarity concerning symptoms:** Two GPs conveyed that it would be helpful if TPPs could clarify what types of symptoms would imply a need/eligibility for using a diagnostic test (GPINT3, GPINT12) because:
  - » Patients with the same cancer can present with different symptoms (GPINT3, GPINT12).
  - » Patients can communicate about the same symptoms differently (GPINT3).
  - » Practitioners can have different interpretations of the same symptoms (GPINT3).
- A related consideration is how patients with different risk profiles (high, medium and low) would fit into the clinical pathway in which the test would operate (PLINT8), i.e. clarity on populations who might be tested relative to their risk for the specific cancer the diagnostic tests for.
- **A test's specificity-related features** to reduce false positives (GPINT1, GPINT3, GPINT6).
- **A test's sensitivity** (i.e. accuracy in avoiding false negatives). Miss rates are important because they affect GPs' confidence in ruling out cancers. According to one interviewee, the lower the miss rate, the higher the GP's confidence in the test and the likelier they are to use a test (GPINT6).
- One pathology laboratory expert flagged features related to general technical performance, including analytical performance requirements (PLINT10).
- Features **clarifying the interpretation of test results** for different patient characteristics, e.g. ethnicity and sex (GPINT3, GPINT4).
- **TPPs should also consider flagging features related to potential ethical issues test developers must consider** (GPINT1,

GPINT2, GPINT4, GPINT5, GPINT7, GPINT12, GPINT14), such as those associated with any incidental findings a test result might detect (GPINT14). [Research team note: However, dealing with incidental findings is a strategic issue for healthcare providers and decision-makers, extending beyond a TPP's scope].

- One pathology laboratory expert highlighted **sample handling requirements,** such as **storage** and **processing** requirements (PLINT9).
- **Infrastructure-related requirements** (PLINT9, PLINT10, PLINT11), including **laboratory buildings, capacity (i.e. staff time and availability) and space** to provide the necessary **equipment** that a test would require (PLINT10). Specifying the infrastructure, including **physical, data and IT compatibility**, is crucial for laboratories (PLINT11). The TPP must reflect consideration and understanding of laboratory processes, including staff, equipment, staffing, and any training required (PLINT9, PLINT10).
- **Specifications for features related to the medical decision a test will affect and how this will inform downstream-care-pathway needs.** Therefore, the treatment pathway is a critical consideration, underlining the importance of actionable results and a seamless path to interpreting results and, if needed, follow-up and treatment (GPINT2, GPINT3, GPINT4, GPINT7, GPINT12, GPINT14, PLINT8).
- **Broader features related to the test's acceptability and usability among healthcare professionals** (GPINT2, GPINT4, GPINT7, GPINT12), such as **capacity, training and ease of use**. To elaborate:
  - » **Human factors such as workforce capacity and training requirements**: For example, interviewees mentioned that due to already-small workforces in pathology laboratories, it is essential to consider who can deliver a test and ensure its sustainability for the workforce and system (PLINT8).
  - » **Factors related to ease of use** and reducing demands on GPs as gatekeepers (GPINT2, GPINT4).
  - » **Related specifications for features reflecting implementation within the current care pathway** (GPINT3, GPINT5, GPINT14, PLINT9, PLINT10, PLINT11), e.g. **integration into existing workflows** (PLINT10). **Understanding how a test would interact with others in the care pathway is important** (GPINT3, GPINT5). Assessing how cancer tests might be added to routine tests for other conditions (GPINT3) or whether a test replaces, complements or enhances current testing may be part of this so that primary care professionals know the next steps in informing the patient (GPINT5).
- **Features related to patient accessibility, acceptability and affordability** (GPINT1, GPINT3, GPINT4, GPINT6, GPINT7, GPINT12, GPINT14, PLINT8, PLINT9, PLINT10), which influence **patient experience**. However, one interviewee flagged that although it is important to consider patient experience, patients are also quite tolerant if they perceive a test's benefit (GPINT4). More specifically:
  - » **Patient accessibility requirements need specification in a TPP** (GPINT1, GPINT5, GPINT7, GPINT12, GPINT14, PLINT9, PLINT10): One interviewee commented on the inaccessibility of genomic testing for many (GPINT4). Others flagged the importance of easier access and fewer visits to health facilities (e.g. **more accessible patient testing**, affordable parking or free transport to facilities) (GPINT1, GPINT3, GPINT7, GPINT14, PLINT8). Some noted the importance of access via public transport, a significant factor in rural areas and counties where money, travel time, distance and availability of transport links are issues (GPINT7, GPINT12, GPINT14, PLINT9). For instance, while a local community practice can conduct a blood or FIT test, radiology may require substantial travel for patients (GPINT7). Patients should not be disadvantaged by where they live and their local health facilities (PLINT9). [Research team note: Although interviewees raised the theme of making facilities as accessible as possible to patients, it is important to caveat that distance from built-up areas, including those that have medical facilities and transport infrastructure, can be a

patient's purposeful choice and thus out of the NHS's control].

- **Specifications related to patient acceptability of a test:** As well as a test being as non-invasive as possible and easy to receive administration-wise, there are issues related to understanding test results that need consideration (GPINT7, GPINT12). For instance, one interviewee flagged that if patients were informed about a chest X-ray's poor sensitivity for detecting lung cancer, they would want CT scans instead (GPINT12).
- **Affordability-related specifications:** Affordability for the patient matters (GPINT1, GPINT3, GPINT7, GPINT14, PLINT8). [Research team note: Even if testing is free on the NHS, they face important time costs and out-of-pocket costs associated with access and travel].
- **Other features affecting a patient's experience**, including the **speed** of testing and receiving results (GPINT6) and convenience, i.e. how easily testing is administered (GPINT6).

- **Specifications related to how the test relates to inequalities [**Research team note: These considerations are also interdependent with accessibility and acceptability considerations]:
  - **TPPs could specify whether a test addresses an inequality or could potentially exacerbate one and how/ to what extent it can balance a test's acceptability in the broader population with ensuring marginalised groups have access** (GPINT1, GPINT5, PLINT9, PLINT10, PLINT11).
  - Overall, **specifications must be transparent** about how the test relates to inequalities, with open information available to explain the reasoning when something is not accommodated or clarify what the complete support process will be (PLINT9, PLINT10, PLINT11). Any decision related to inequalities needs an accompanying explanation of what a test can and cannot do, with concomitant reasoning (PLINT10).
  - The above point relates to the earlier one about clarity on the groups the test may be unsuitable for and why (GPINT1, GPINT5, GPINT6, GPINT14, PLINT11).
- **Specifications for features related to economic considerations (costs and cost-effectiveness):**
  - **TPPs should consider product costs and broader costs associated with a test's use and implementation in the health system** (GPINT1, GPINT2, GPINT4, GPINT5, GPINT7, GPINT12, GPINT14, PLINT10), e.g. related to test processing and result delivery (PLINT10), including the health workforce and their time (GPINT2, PLINT10), and equipment maintenance costs (GPINT4, GPINT5, PLINT10).
  - How often a test will likely be used – i.e. its likely **use volume** – is also important for a TPP to specify so laboratories can forecast budgets and impact on other work (PLINT10). One expert flagged that piloting test implementation would be a good way to assess these broader aspects of a test's cost (GPINT5), but this cannot happen at TPP stages. However, **early economic modelling** can help identify broader system costs (GPINT5).
  - **Cost-effectiveness is essential** (GPINT2, GPINT12, GPINT14, PLINT 9, PLINT10): Early economic modelling could be undertaken for specific scenarios (GPINT7) and in consideration of the National Institute for Health and Care Excellence (NICE) Health Technology Assessment (HTA) requirements (GPINT2, GPINT12). Economic cost-effectiveness modelling should also consider broader downstream system costs (GPINT5, GPINT12, GPINT14). [Research team note: Some costs may be one-off, and others may be recurrent]. Many interviewed experts emphasised that a technically difficult or expensive test to deliver must demonstrate other benefits justifying its costs to the NHS (GPINT1, GPINT4, GPINT5, GPINT6, GPINT7, PLINT8, PLINT9, PLINT10, PLINT11).
  - Where possible, cost-related specifications in a TPP could distinguish **between short-term versus long-term cost-effectiveness, including potential NHS savings** (GPINT4, GPINT6, PLINT8). One expert flagged that although the

upfront costs for tests like a multianalyte test may be high, the individual test's cost plus the potential cost savings of early diagnosis diagnosing could imply longer-term savings for the NHS (PLINT8) [Research team note: This was a tentative comment that would need supporting evidence to justify its inclusion in economic assessment].

- **Features related to environmental considerations and requirements:** Now the NHS is aiming for net-zero emissions, a test's environmental impacts must also be considered (GPINT1, GPINT14). Environmental aspects are also crucial regarding waste disposal and management in laboratories, e.g. how to appropriately dispose of the toxic chemicals needed for some tests (PLINT8).
- **Specifications for effective performance in 'real-world' clinical practice and pathology and genomic laboratories, not just in clinical-trial settings** (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT7): Some interviewees pointed out that clinical practice poses very different challenges for a diagnostic test than a clinical trial (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT7). The potential list of what could 'go wrong' in real-world primary care is difficult to gauge in a study setting (GPINT7). Differences between clinical practice and trial environments relate to diverse issues, with implications for features needing specification. For example:
  - » Study-based selection bias due to **trial participants' characteristics versus patients in real-world practice** (GPINT7).
  - » Differences in **how frequently a test is used (i.e. the volume needed) can link to shelf-life specifications** and the potential for waste (GPINT1, GPINT5).
  - » **The test's use environment can also vary from trial settings (e.g. temperature-change sensitivity), and logistical hurdles such as transport infrastructure to safely transport and store samples must also** be considered (GPINT1, GPINT4, GPINT12, GPINT14).
  - » **The real-world requirements for specifications related to histopathology and laboratory infrastructure** requirements are also essential to consider in a TPP (PLINT9, PLINT10, PLINT11), including how the test must fit within existing space constraints and the necessary conditions for storing samples and ensuring appropriate cross-contamination controls (PLINT9).
  - » **Specifications related to ongoing maintenance requirements and costs** (GPINT5).
  - » **Quality control** (GPINT5).
  - » Real-world workforce-related requirements, such as available staff and NHS capacity to implement and interpret a test and any related training requirements (GPINT1, GPINT2, GPINT3, GPINT12, GPINT14, PLINT9, PLINT10).

1.4. The process of developing diagnostic TPPs for cancer

1.4.1. Stakeholders to involve in TPP development

Interviews with primary-care GPs and pathology and genomic laboratory experts shed light on the types of stakeholders who could be important to involve in TPP development:

- **Interviewees from primary care felt it would be important for GPs to be part of a core working group leading TPP development** (GPINT1, GPINT3, GPINT4, GPINT5, GPINT6, GPINT7, GPINT12). **Diverse types of GP expertise were considered necessary** (GPINT2, GPINT3, GPINT5, GPINT6, GPINT12, GPINT14), both research active (GPINT2, GPINT6) and those fully engaged in clinical practice (GPINT3, GPINT6, GPINT14). Primary care professionals involved with cancer networks (e.g. Cancer Alliances) (GPINT5, GPINT12) and charities (such as CRUK) (GPINT6) and those with wider public-health and oncology-relevant expertise and interests could add important subject matter expertise (GPINT5, GPINT6).
- **Other primary care professionals' early involvement was also considered helpful by many interviewees, bringing a perspective on the 'practical' elements of a test and its usability** (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT7, GPIN12, GPINT14) in a way that other stakeholders perhaps cannot (e.g. industry may have scientifically sound ideas but may not be as familiar with NHS practicalities). **Overall, those consulted as part of the working group or during the process will depend on the use case and specific TPP** (PLINT8, PLINT11). Mentioned examples **include nurses, healthcare assistants, practice managers and pharmacists** (GPINT1, GPINT2, GPINT5, GPINT6, GPINT7, GPIN12, GPINT14); **physician associates** (GPINT3); **data specialists, paramedic practitioners and cancer care coordinators** (GPINT5); **and representation from Royal Colleges** (GPINT3).
- **Some interviewees saw other diagnostic end-users in the healthcare system beyond primary care as important**, e.g. secondary care doctors, radiographers, counsellors in cancer trials and oncologists (GPINT7, GPINT12, PLINT10).
- **Interviewees from the pathology community also emphasised their expertise relevance in a core working group** (PLINT8, PLINT9, PLINT10), suggesting it would be important to ensure their engagement throughout the process to sense check and ensure tests are feasible and realistic under current laboratory capacities and capabilities (PLINT8, PLINT9, PLINT10). [Research team note: This would only apply for tests requiring pathology laboratory inputs].
- **Many interviewees discussed the role of other stakeholders beyond primary care and pathology/genomic laboratory experts in TPP development efforts** (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, PLINT8, PLINT9, PLINT11). Mentioned examples included:
  - **Experts in the diagnostics a proposed innovation would replace would be vital during the scoping phase of TPP development,** connected to the importance of understanding and clarifying any proposed innovation's value proposition (GPINT2). This group can include academic expertise but is not confined to academics only (GPINT4).
  - **Research expertise is important to provide methodological inputs in TPP development efforts, e.g. conducting literature reviews** (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5) and evidence that a test offers a measurable improvement on an existing diagnostic (GPINT2, GPINT4).

- **Patient representation and input matters in a core working group to assist with co-producing a TPP** (GPINT1, GPINT4), **providing insights and feedback on specifications related to usability and accessibility** (GPINT2, GPINT7, GPINT12, PLINT8, PLINT9), as well as **acceptability and potential unintended consequences** (PLINT8, PLINT9). Involvement could be enabled via the system's existing Patient Public Involvement and Engagement (PPIE) infrastructure, such as Cancer Alliance Patient Advisory Groups (PAGs) (GPINT12). One interviewee emphasised that any TPP development process must involve PPIE and seek diverse inputs, e.g. from various ethnic groups and religions and considering various inequalities (GPINT7).
- **Third-sector organisations, e.g. charities** (GPINT1, GPINT6).
- **Policymakers and regulators** (GPINT2, GPINT6, GPINT7, GPINT12, PLINT8) are important to involve in TPP development. For example, NICE guidance could signal a diagnostic tool's credibility to GPs (GPINT6). GPs often use NICE criteria as guidance to choose whether or not to refer a patient based on symptoms. Clarifying NICE requirements regarding a test's eligibility and referral criteria is important for a TPP. However, one interviewee highlighted the importance of keeping HTA involvement at arm's length to minimise the risk of stifling innovation if introduced too early in TPP development (GPINT2). The reason the interviewee (GPINT2) gave for this was their perception that NICE might quash innovation if involved too early because TPP development processes likely involve many features that initially seem costly; however, costs may go down or up or new developments mean certain features are more or less valid than they were initially. Therefore, involving NICE too early may quash development before a case could be made (in later stages) for a diagnostic's benefits (GPINT2).
- A related point is that some interviewees considered **health economist** expertise relevant for understanding economic considerations and cost-effectiveness (GPINT7, GPINT12).
- **Existing health and innovation networks in the system, such as Academic Health Science Networks (AHSNs), Cancer Alliances and the Royal College of GPs,** were brought up as possible stakeholders to provide approval or review (GPINT2, GPINT3, PLINT11) – although only once primary care professionals have helped develop a draft TPP (GPINT2, GPINT3) to avoid 'onerous' engagement with an AHSN (GPINT3).
- Industry could also be consulted during the scoping phase, as they know what is new and cutting-edge (PLINT8). **One interviewee noted that industry should be part of the core working group alongside academics and GPs** (GPINT4). [Research team note: It is important to consider ways of mitigating bias]. Suppliers of testing platforms were also flagged (PLINT9).
- **The UK Accreditation Service (UKAS), with which laboratories must be accredited before operating,** is another group worth engaging in later TPP development stages (PLINT8).

1.4.2. Key insights for engaging primary-care professionals in TPP development

Any approaches to engaging GPs in TPP development must be mindful of time and capacity constraints (GPINT2, GPINT5, GPINT14). Cancer leads may be good to engage because they can use the time allocated for these activities instead of clinical time (GPINT2). **One GP flagged the importance of pragmatic engagement methods because of GPs' limited capacity to utilise more systematic means** (GPINT2)**. However, interviews identified diverse possible strategies for engaging GPs and laboratory experts in TPP development** (GPINT1, GPINT4, GPINT6, PLINT8) [Research team note: However, these must not compromise necessary rigour]:

- While there may be diverse preferences for engagement, one GP noted their preference for face-to-face involvement, no matter the method of eliciting engagement and information (e.g. interviews or surveys) (GPINT4).
- However, another interviewee noted that many different engagement methods can work if GPs are adequately prepared and notified in advance (GPINT1, GPINT2, GPINT6). Effective GP engagement could be facilitated by sharing appropriate preparatory material in advance (GPINT1, GPINT6), especially written documentation with a clear background (GPINT2). As long as GPs are first given background information and material (with clear objectives, desired activities, outputs and next steps) before any engagement, they will likely be open to engaging with any scoping, drafting, and consensus-building method (GPINT6).
- A pathology expert mentioned using Delphi surveys to engage with GPs for consensus-building alongside interviews during the scoping phase (PLINT8).

Initial early engagement with primary care professionals was considered important for informing TPP specifications in a timely way (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5). Such engagement would allow primary care professionals to help set the direction and have an early checkpoint on progress and timely feedback (GPINT1, GPINT2). **Consensus considerations with any stakeholder group involved in TPP development should ensure decisions align with primary efforts toward patient benefit and acceptability** (GPINT6). One expert flagged that this is key to finding consensus between industry, academics and other stakeholders, even on topics that could be controversial (GPINT6). **Another GP flagged the importance of ensuring GPs are properly renumerated for time** (GPINT12). While this may introduce some bias, GPs may not engage unless their time is remunerated.

1.5. Prioritisation

Based on interview insights, the following prioritisation considerations emerged as important, primarily centred on improving patient outcomes and care pathways:

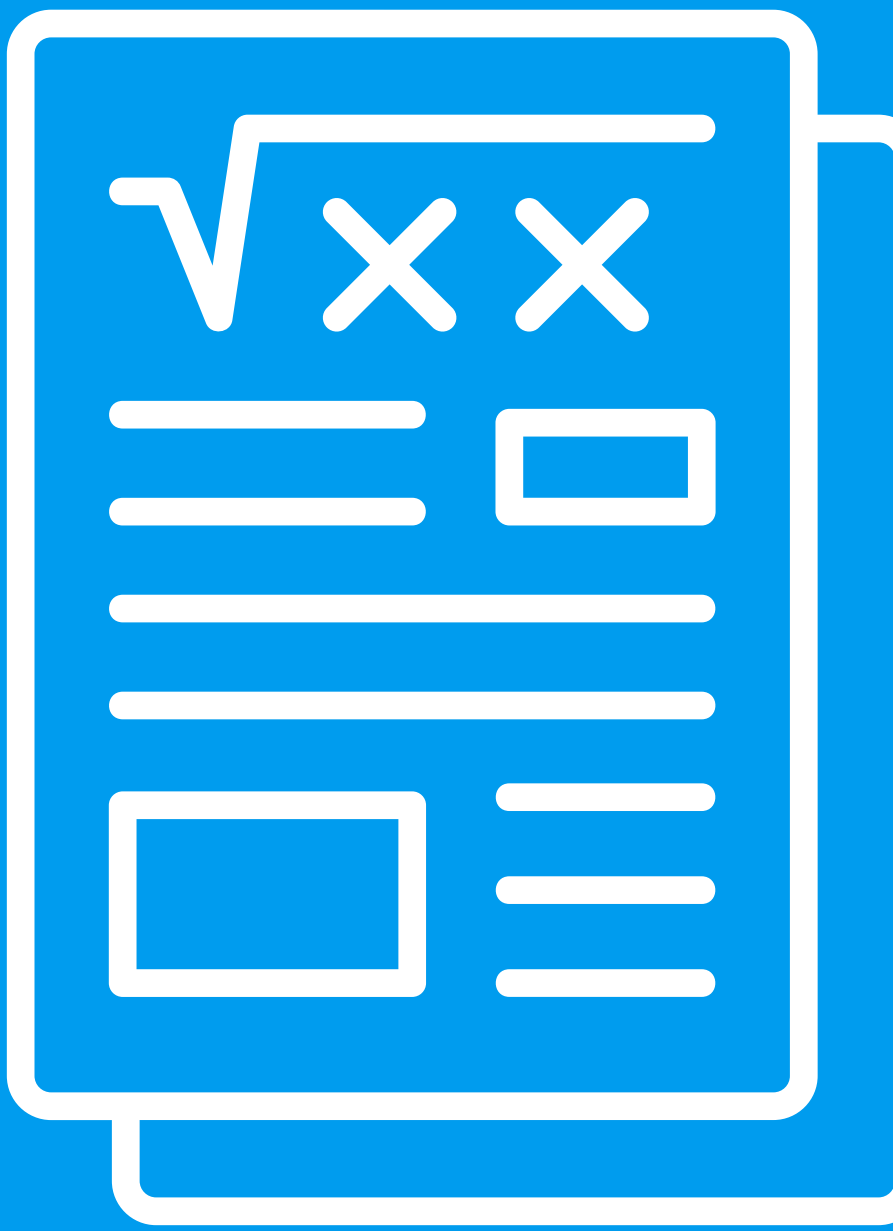
- **Target population size and incidence data (i.e. most common cancers) for a test's highest impact** (GPINT1, GPINT3, GPINT7, GPINT14, PLINT10) **while not neglecting tests for rare cancers** (GPINT1, PLINT8).
- **Opportunities to improve early diagnosis** (GPINT1, GPINT5, PLINT9) would improve treatment outcomes and quality of life (GPINT5). This includes a need for:
  - **Tests enabling quicker diagnosis or diagnostic pathways while still meeting high performance standards** (GPINT2, GPINT4, GPINT7, PLINT8, PLINT9, PLINT10), and where AI can help meet this demand (GPINT2, GPINT4, PLINT8, PLINT9).
  - **Tests capable of detecting and diagnosing early cancers that present vague, non-specific, undifferentiated or non-deterministic symptoms** (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT7, GPINT12, GPINT14).
  - **Innovative diagnostic tests focused on pre-symptomatic and pre-hospital parts of the patient care pathway** to facilitate early diagnosis (GPINT1, GPINT4, GPINT5, GPINT7).
  - **Tests that can rule cancers out as well as in and have improved specificity** (GPINT1, GPINT3, GPINT4, GPINT6, GPINT7, GPINT14).
- **Test that can improve patient accessibility** (GPINT5, GPINT12) **or acceptability** (GPINT1) or both (GPINT2, GPINT4, GPINT6, GPINT14, PLINT9, PLINT10).
  - **Related to patient accessibility, it is vital to identify the key inequalities across patient populations, which would be the first step in ensuring a test's accessibility** (GPINT14).
  - **Related to patient acceptability, less invasive tests that are easier on the patient administration-wise** (GPINT1, GPINT2, GPINT4, GPINT5, GPINT6, GPINT12) would help increase patient uptake and ease the NHS workforce pressures.
- **Tests that can improve a test's clinical acceptability**, e.g. by:
  - **Raising trust in a test, along with ease of use** (GPINT2, GPINT4, GPINT6) and **ease of result interpretation** (GPINT2, GPINT7, GPINT12, GPINT14),
  - **Helping address broader NHS workforce capacity and demand considerations** (GPINT3, GPINT5, GPINT12, GPINT14, PLINT8, PLINT9, PLINT10), including **tests that can help catch up on testing backlogs** (PLINT8, PLINT9, PLINT10). Pathology laboratories often receive gastrointestinal, breast, urology, bowel and skin samples. As such, support addressing backlogs in these areas would be helpful in current challenges.
  - **Improving tests that can help reduce pressures in primary care,** particularly tests that are easy to use, administer and understand in this setting while still accurate (GPINT1, GPINT2, GPINT3, GPINT4, GPINT5, GPINT6, GPINT12, GPINT14).
- **Multianalyte tests that can detect multiple cancer types simultaneously,** including for early diagnosis (GPINT1, GPINT4, GPINT7).
- **Tests that can help mitigate legal liabilities,** e.g. for cancers and testing areas subject to the most medical claims or those for which GPs and primary care are most sued (GPINT14).
- **Tests that are responsive to cost pressures** to ensure they are not too expensive for people to 'buy into' them. This also requires an effective business case outlining the system and patient benefits (PLINT9).
- **Tests that can be implemented effectively and add value to current care pathways and patient outcomes** (PLINT11).

References

1 NHS England. 2023. 'Should I have a PSA test?' As of 22 March 2024: https://www.nhs.uk/conditions/prostate-cancer/should-i-have-psa-test/

Annex E: Economic Modelling Tool

Authors: Matthew Napier, Amanda Cole, Lotte Steuten, Mark L Cabling, Jessica Dawney, Zuzanna Marciniak-Nuqui, Fifi Olumogba, Larry Kessler, Sonja Marjanovic



The economic modelling tool: A guide

The economic modelling tool is intended as a resource to identify a diagnostic test's main value proposition and the health economic impact of different Target Product Profile (TPP) specifications, focused on those expected to directly affect its health economic value. The tool comprises an Excel-based model and is accompanied by a Word document that acts as a 'guide'. Together, they can be used to perform a simplified early economic evaluation primarily aimed at identifying the TPP requirements expected to have the most significant impact on cost-effectiveness and indicating the test's potential cost-effectiveness, assuming 'perfect implementation'. The value provided is solely indicative and should be considered just one of the many pieces of information needed to develop the complete TPP.

Annex E is the fifth of eight annexes complementing the main Cancer Research UK-funded project final report 'Advancing the development and use of diagnostic target product profiles for cancer', led by RAND Europe and the Office of Health Economics. This document provides a detailed economic modelling tool guide to which the final report refers; thus, Annex E, like all the other annexes, is primarily meant to accompany the final report and is not meant to be read as a standalone document.

This annex contains three sections. Section 1 aims to provide an understanding of the key features of a proposed diagnostic that are expected to impact its potential cost-effectiveness. Such features include a test's analytical performance, whether/how it impacts the current care pathway, and whether it will likely save costs for the healthcare service. Resources are provided to help inform the answers to these questions: for example, when defining the comparator, the National Institute for Health and Care Excellence (NICE) guidance can provide useful information on the current care pathway.

Section 2 aims to support the parameterisation of the relevant variables for input into the economic modelling tool. We followed the NICE reference case since we focused on a United Kingdom (UK) context. Therefore, only health-related quality of life and costs from a National Health Service (NHS) and personal social services perspective are deemed relevant to the analysis.

Section 3 introduces the MS Excel modelling tool, outlining the modelling approach taken and discussing the tool's limitations and scope. We then present an illustrative example to demonstrate the tool and Early Economic Evaluation (EEE) modelling in use.

Section 1. The value proposition

Insights from our workshop findings and literature review suggested that UK-based TPPs should consider using NICE's approach to assessing cost-effectiveness. By considering a health technology's cost and health impact through an economic evaluation, NICE assesses whether it offers good value for money.

Even when a TPP is under development, the impact of specific diagnostic characteristics on its overall cost-effectiveness can be estimated using early economic evaluation.

The first step of this process is articulating the test's value proposition. We outline a series of questions/steps below designed to draw out responses informing diagnostic features that will impact its cost-effectiveness (see Table 1 for further details):

- Firstly, describe the diagnostic, including its specific use case, the unmet need it meets and a clear value proposition.
- Secondly, define the existing comparator test and current care pathway. This is important as NICE assessments involve an incremental evaluation, comparing current practice to the proposed 'new' pathway or technology.
- Outline the diagnostic target population, highlighting how many people will likely be eligible for the test and whether any potential sub-groups might benefit more or less than the average population.
- Next, explore the new diagnostic's impact on the health service, including whether it will lead to disinvestment of an old test or potential savings in the health system (e.g. in hospital bed days).
- Define the potential health impact of the new diagnostic, outlining how it might change the clinical care pathway and whether it might lead to a more accurate or timely diagnosis.
- Finally, outline whether there are any expected changes in the policy landscape or treatment options in the future. Any change to the broader context will likely impact the test's cost-effectiveness, particularly if a new treatment is on the horizon.

Table 1. TPP Features related to the diagnostic's cost-effectiveness

Categories:	Questions:	Link to TPP features	Responses:
Test description: provide a value proposition for the proposed test	What is the use case? What key unmet needs does it meet? What are the expectations for its analytical performance (sensitivity and specificity) at this stage?	'Unmet need', 'Analytical performance'	

Categories:	Questions:	Link to TPP features	Responses:
Comparator: usual clinical practice is the appropriate comparator (or the relevant test/technology recommended in current NICE guidance)	What is the current care pathway? Will the test replace or be used in addition to an existing test?	'Clinical utility', 'Human factors', 'Downstream impacts on care pathways and processes'	
Population	Who will be eligible to receive the test? How many people are expected to be tested? Are there any subgroups for sensitivity analysis?	'Inequalities and health equity', 'Clinical utility', 'Cost/economic considerations'	
Health service impact: changes in service delivery costs for the NHS and personal social services	Linked to the comparator question, will the test replace current practice (lead to disinvestment)? Could there be savings in staff time, hospital bed days or General Practitioner (GP) visits, etc.? Will the diagnostic require training or investments in infrastructure?	'Human factors', 'Cost/economic considerations', 'Downstream effects on care pathway and care processes', 'Infrastructure'	
Patient/Health Impact: how the test may lead to improved health-related quality of life	How might the clinical care pathway change by introducing the new test? Could the test lead to earlier/more timely diagnosis and potentially reduce late-stage cancer?	'Clinical utility', 'Downstream impacts on care pathways and care processes', 'Patient acceptability and experience'	

Categories:	Questions:	Link to TPP features	Responses:
Key uncertainties and research questions	Are there any new treatments on the horizon? Are there any significant upcoming policy changes you know of (e.g. the NHS's major conditions strategy)? If you expect there to be multiple use cases, outline these.	Depending on the uncertainties and research questions that the person enters, any or all of the following: 'Clinical utility', 'Downstream impacts on care pathways and care processes', 'Patient acceptability and experience', 'Human factors', 'Cost/ economic considerations', 'Infrastructure', 'Inequalities and health equity', 'Unmet need', 'Analytical performance'	

Useful background information sources when defining a value proposition:

- NICE methods guide: useful for information on the methods and processes NICE follows when undertaking health technology assessments.[1]
- Published NICE guidance on diagnostics: helpful when defining the current care pathway and eligible population.[2]
- All NICE guidance: useful for defining the population, current care pathway and comparator(s).[3]
- General information on the condition, cause and diagnosis.[4]
- The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) for health economic evaluations for background.[5]

Section 2. Parameterisation

Once the main drivers of cost-effectiveness have been outlined, their respective input parameter values need estimating. This step enables comparison of the new diagnostic's costs and health-related quality-of-life impacts against the current care pathway, providing an initial sense of whether a new diagnostic meeting the TPP could offer value for money.

This comparison is undertaken using an incremental cost-effectiveness analysis, with the main output being the Incremental Cost-Effectiveness Ratio (ICER) – the incremental cost per incremental quality-adjusted life year (QALY) of the new technology. In England, the health technology assessment body, NICE, sets a threshold ICER against which technologies are compared to determine whether they offer value for money, generally £20,000-30,000 per QALY.

The parameterisation comprises three parts: population-level data, QALY impact and costing.

Population-level data:

Population-level data refer to the size of the population that may be eligible for the diagnostic, their average age and the cancer's prevalence:

- Previous NICE guidance is a good resource when defining the size and eligibility of the relevant population.[2] Published NICE guidance for all types of health technologies is also available.[3]
- Cancer Research UK has published statistics on the incidence, mortality, survival and prevalence of multiple cancers, providing statistics by cancer type. All the statistics are available.[6]

QALY impact:

In terms of health outcomes, NICE considers QALYs, representing the estimated survival of patients combined with their health-related quality of life (between 1–0) in their life span.

One QALY equals one year in perfect health, equivalent to two years with a health-related quality of life of 0.5 or four years with a health-related quality of life of 0.25. Death is equal to a quality of life of zero.

Health-related quality of life:

Estimates of health-related quality of life can be found using age-specific population-level values and adjusting these by the cancer-related decrement. Those who do not have cancer (a true negative diagnosis) can be modelled according to the population-level values:

- Age and sex-adjusted health-related quality of life values.[7]
- Quality of life disease weighting/decrement (i.e. health-related quality of life relative to perfect health for someone with a particular cancer).
    - » A useful source of previous cost-effectiveness analyses may provide some input values.[8]
    - » The Cochrane Library may also provide useful input values.[9]

Survival:

Baseline survival estimates by sex and age can be adjusted to incorporate the increased likelihood of mortality due to cancer using a relative risk (of mortality) value. Those who do not have cancer can be modelled according to population-level survival expectations.

Alternatively, cancer-specific survival data could be used when estimating survival (see below):

- Baseline survival – age and sex-adjusted.[10]
    - » Data can be used to estimate life expectancy by sex and age in the UK.
    - » The relative risk of mortality due to cancer can be applied to calculate the probability of mortality.
- Alternative - cancer survival data.[11]

The health-related quality-of-life and survival data can be combined and modelled to

calculate the expected QALY gains for each diagnostic pathway.

Costing:

The relevant costs from the NICE reference case are from NHS and Personal Social Services (PSS) perspectives, and do not represent broader societal costs (e.g. lost patient productivity due to treatment, illness or death). Various sources can be used to monetise costs, whether related to drugs, staff time, commissioning NHS services, or inflation:

- Drug costs from the NICE British Nation Formulary (BNF).[12]
- Staff time costs and uprating costs with inflation indices[13]: The unit costs of health and social care for 2022 provide data on the annual cost estimates for delivering health and social care services. For example, the cost per working hour for different NHS staff (e.g. qualified nurses by seniority and salary band).
- NHS services costs:[14] This data is collected on the aggregated costs of defined services and activity levels (how many patients receive care), providing the average unit costs of providing defined services and a cost based on the patient's specific interaction. For example, the data set would give the number of patients suffering a stroke (by Casemix Companion score, which classifies patient care based on expected clinical resource use for providing that care), the total cost of this, and the cost per patient.

Section 3: MS Excel modelling tool

The tool and the modelling approach are intended to support the TPP development process in understanding where a test's primary economic value lies and where uncertainties may be most significant. The modelling tool applies a generic diagnostic modelling structure. It consists of a decision tree, shown in Figure 1, where all patients eligible for testing are tested and either have cancer or do not, receiving a positive or negative diagnosis. This results in four groups of results:

1. True-positives
2. False-positives
3. True-negatives
4. False-negatives.

The patient's outcomes are then modelled over an appropriate time horizon, depending on the test outcome and whether they have cancer. Those with true negative and false positive results are modelled according to the general population's survival and health-related quality of life.[7,10] Those with false negative and true positive results enter a three-state transition model commonly used in oncology: progression-free disease, progressed disease and death. Figure 2 shows the model's structure.

Figure 1. Decision tree

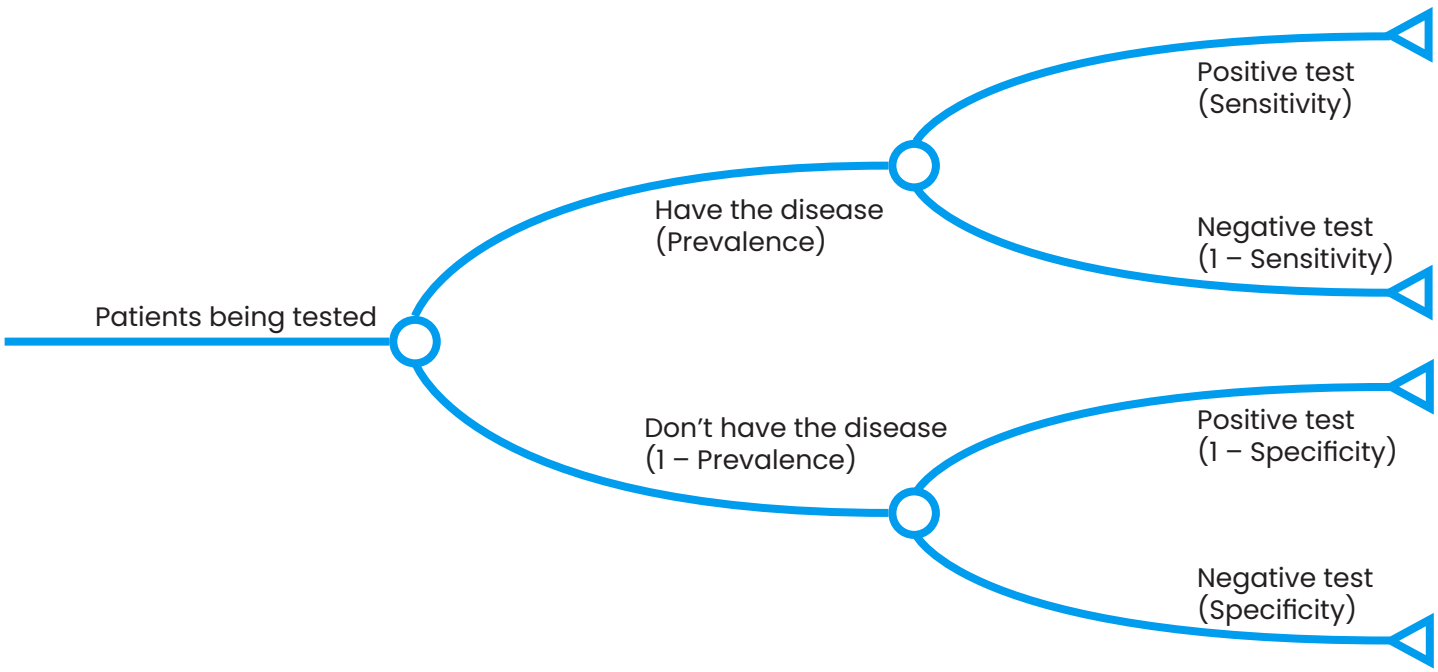
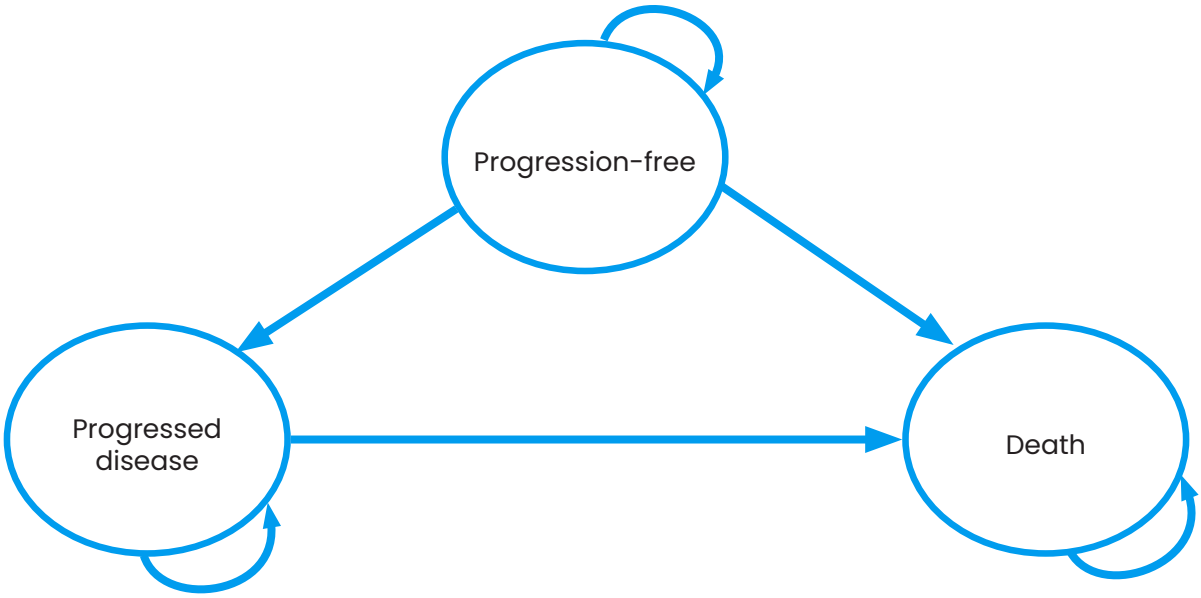


Figure 2. State transition model



Although a degree of flexibility is allowed in populating the model, the structure itself is fixed and may not be suitable for modelling all diagnostic technologies. For example, a discrete event simulation model may be more appropriate for a particularly time-critical diagnosis or if the new technology significantly changes the patient pathway. In addition, as with all diagnostic models, the cost-effectiveness is highly dependent on the availability of subsequent treatments, thus whether there are new treatments or broader changes in health service delivery on the horizon that may impact the diagnostic's cost-effectiveness when it reaches the market.

Bespoke modelling efforts will need to be undertaken to derive more precise cost-effectiveness estimates, particularly given the heterogeneity in how diagnostics derive health economic value and affect the care pathway. Therefore, this model should be considered as a starting point when developing a TPP for providing initial estimations of:

- The likelihood that the diagnostic will be cost-effective (under reasonable assumptions), ruling out those diagnostics that will be highly cost-ineffective.
- The parameters with the most significant impact on cost-effectiveness under the current assumptions.
- The impact of varying individual parameters on the cost-effectiveness of the diagnostic (e.g. the direct effect on cost-effectiveness of varying the test's sensitivity/specificity).

An illustrative example using the modelling tool

Table 2 details the illustrative example: a hypothetical new software developed in response to a given TPP to support the use of biopsy when diagnosing cancer. It is hoped that the technology will lead to fewer missed cancer cases and fewer repeat biopsies, but it comes at a higher cost compared with the current diagnostic approach.

Table 2. An illustrative example of a hypothetical diagnostic

Categories:	Questions:	Responses:
Test description: provide a value proposition for the proposed test	What is the use case? What key unmet need does it meet? What are the expectations for its analytical performance (sensitivity and specificity) at this stage?	A new technology has been proposed to diagnose a form of cancer. The current unmet need is for a more precise diagnostic to detect the cancer more quickly. The software would overlay an MRI image onto a live ultrasound image when performing a biopsy, which may lead to fewer missed cancer cases (false negatives) and fewer repeat biopsies (reduced costs). Thus, the sensitivity is expected to improve compared to usual care.
Comparator: usual clinical practice is the appropriate comparator (not best practice)	What is the current care pathway? Will the test replace or be used in addition to an existing test?	Previous NICE guidance describes the current care pathway well, and the new diagnostic doesn't affect the subsequent care pathway. The test would replace an older form of biopsy performed using previously taken MRI images and live ultrasound imaging to guide the biopsy needle. The newer form of biopsy is expected to replace the existing testing method.
Population	Who would be eligible to receive the test? How many people are expected to be tested? Are there any subgroups for sensitivity analysis?	People with suspected cancer who have had an MRI scan indicating potential cancer, which applies to 5,000 people per year.
Health service impact: changes in service delivery costs to the NHS (and personal social services)	Linked to the comparator question, will the test replace current practice (lead to disinvestment)? Could there be savings in staff time, hospital bed days, GP visits, etc.? Will the diagnostic require training or investments in infrastructure?	The overall cost to the health system associated with the new diagnostic is expected to be greater than the cost of usual care. However, there will be some savings associated with fewer repeat biopsies and fewer adverse events. The increased costs will be mainly due to staff training, system maintenance and installation of the required software.

Categories:	Questions:	Responses:
Patient/Health Impact: how the test may lead to improved health-related quality of life	How may the clinical care pathway change with the introduction of the new test? Could the test lead to earlier/more timely diagnosis and potentially reduce late-stage cancer?	It is expected that the new diagnostic will improve health outcomes and satisfy some of the current unmet needs. The improved health outcomes are expected due to improved (early/increased) cancer detection and fewer misdiagnoses (fewer false negative results).
Key uncertainties and research questions	Are there any new treatments on the horizon? Are there any significant upcoming policy changes you know of (e.g. the NHS's major conditions strategy)? If you expect there to be multiple use cases, outline these.	Uncertainty about the effectiveness of the technology due to limited evidence. No new post-diagnostic treatments have been identified on the horizon.

Population-level data

Table 3 (below) shows the population figures for this example and some modelling parameters. The key parameters highlighted are the proposed and current diagnostics' sensitivity scores: 0.85 and 0.65, respectively. This is one of the key values associated with the proposed diagnostic, which is expected to identify more true positive cases and fewer false negative results, as outlined in Table 2. The cycle length refers to the time interval for which the model calculates costs and outcomes. After the model has completed a cycle, patients move health states according to the transition probabilities, and this is repeated until the full time is reached. The discount rate is applied to bring costs and outcomes obtained in the future (i.e. beyond the current year) to their present value, with 3.5% being NICE's chosen annual discount rate.[1]

Table 3. Population-level data

Parameter	Value
Time horizon (years)	10
Cycle length (years)	1
Prevalence of the cancer	0.7
Mean age at diagnosis (years)	60
Proportion male	0.5
Proportion female	0.5
Discount rate	3.5%
ICER threshold (£/QALY)	20,000
Sensitivity (proposed diagnostic)	0.85
Specificity (proposed diagnostic)	0.95
Sensitivity (current diagnostic)	0.65
Specificity (current diagnostic)	0.95
Eligible population per year (n)	5,000

Costing data

Table 4 presents the costs associated with each diagnostic strategy. As outlined in Table 2, the overall cost associated with the proposed diagnostic is expected to be greater than that of the current diagnostic due to the additional procurement, installation, and training required. The diagnostic does not lead to different treatments, so the longer-term costs for each pathway are the same. However, we expect more patients to be correctly diagnosed with the new diagnostic, meaning fewer patients will be diagnosed with late-stage disease. This will help lower costs for the health system, as the cost of later-stage disease is more significant than that associated with progression-free survival.

Table 4. Costing data

Parameter – Costing	Value (£)
Procurement cost	10,000
Installation cost	100
Maintenance	200
Training	100
Procedure time (cost)	100
Total (proposed) diagnostic-related costs	10,500
Maintenance	200
Procedure time (cost)	100
Total (current) diagnostic-related costs	300

QALY impact

The diagnostic's impact on survival and health-related quality of life must be estimated to calculate the QALY impact. To model the QALY impact, the utility values for the UK population,[7] adjusted by age and sex, are decremented by the disutility associated with progression-free and progressed cancer. In addition, the disutility associated with misdiagnosis needs to be accounted for. Table 5 shows all disutility values.

Table 5. Health-related quality of life

Parameter – Utility Decrement	Value
Progression-free	0.1
Progressed	0.4
False positive (misdiagnosis)	0.05
False negative (misdiagnosis)	0.1

The baseline survival is modelled using ONS life tables data.[11] Then, relative risk values are applied to the data to account for the increased risk of death associated with having progression-free and progressed cancer. An increased risk of death is also applied to those who receive a false negative result, accounting for the increased risk of death due to a less timely diagnosis. Table 6 summarises the parameters.

Table 6. Relative risk of death compared to baseline

Parameter – Relative Risk	Value
Progression-free to Death	1.5
Progressed to Death	2
Progression-free (false negative) to Death	2
Progressed (false negative to Death)	2.5

Results and analysis

Applying the parameter values under the current assumptions, the results indicate that the proposed diagnostic would be deemed cost-ineffective, as shown in Table 7. For an additional £9,000 cost over the 10-year time horizon, a further 0.19 QALYs are expected. £9,000/0.19 => £47,368 per QALY, which exceeds the NICE threshold range of £20,000-30,000 per QALY.

Table 7. Results

Headline results	
Difference in costs	£9000
Difference in QALYs	0.19
ICER	£47,368/QALY

As discussed, another important finding from early economic modelling is the most sensitive variables affecting cost-effectiveness, i.e. the magnitude of the effect that a change in a variable's value will have on the ICER. This is explored through a one-way sensitivity analysis, with the Tornado diagram in Figure 1 showing the top ten most sensitive variables.

As Figure 3 demonstrates, the most sensitive variables are the sensitivity of the proposed and current diagnostic and the cost associated with the proposed diagnostic. These results make intuitive sense as these are two of the key elements of the proposed diagnostic: it comes at a higher cost than the current diagnostic but provides a more accurate diagnosis (and has a downstream impact that improves health outcomes).

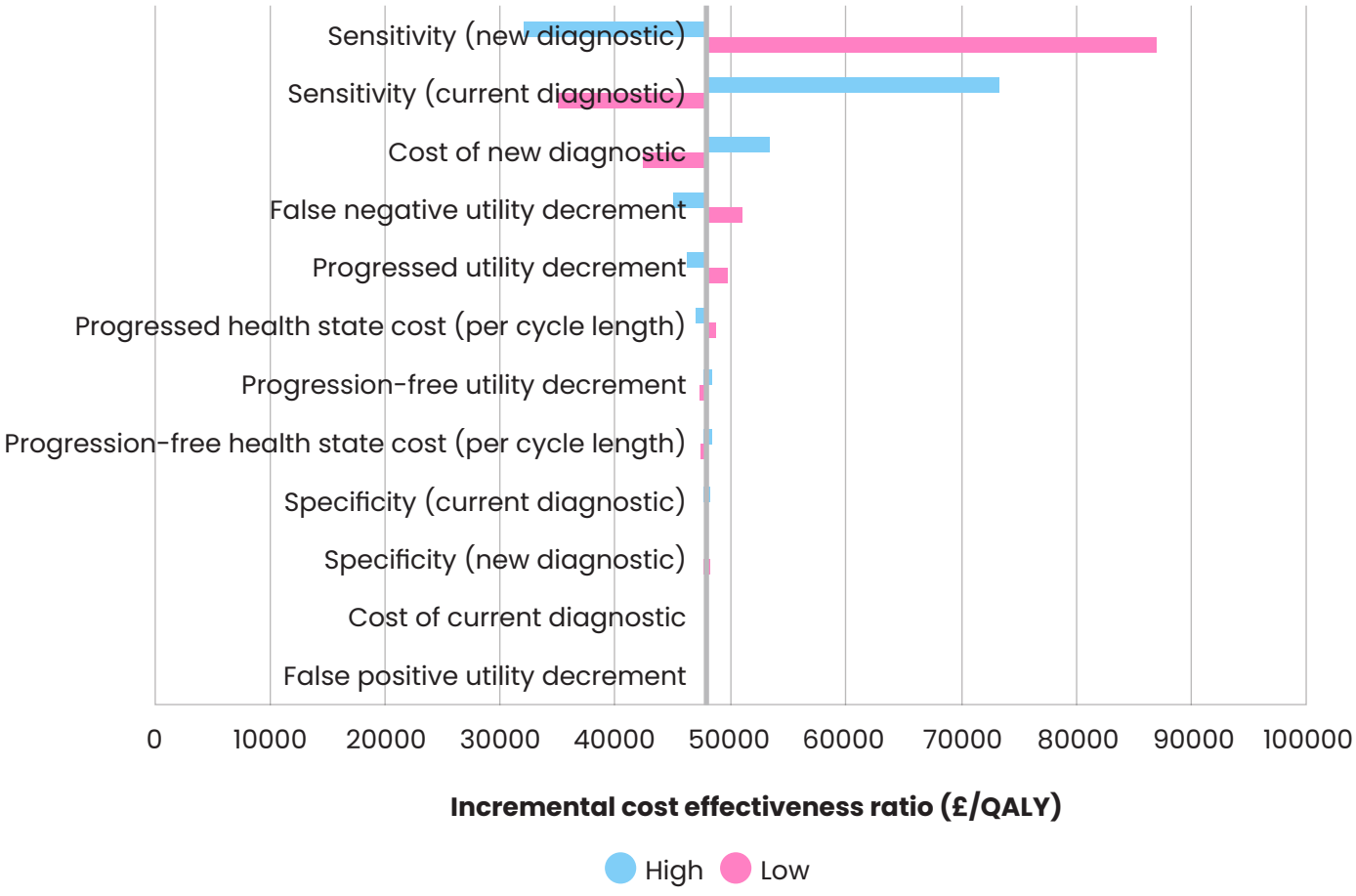
In contrast, a change in the diagnostic's specificity does not lead to much variation in the ICER. Again, this makes intuitive sense; since the specificity is already relatively high (only 0.015 receive a false positive in each pathway), there is little scope for improved specificity to lead to better outcomes and, thus, improved cost-effectiveness.

When developing a TPP, it is essential to consider the potential downstream impacts of introducing the proposed diagnostic, particularly when this impacts its potential cost-effectiveness. Such considerations include the impact on the health system, the potential health/patient benefits and whether there are potential new treatments on the horizon.

We hope that through this illustrative example, we have demonstrated the role early economic modelling can play in TPP development by:

- Indicating the potential cost-effectiveness of the proposed diagnostic.
- Showing the impact that varying individual parameters have on the cost-effectiveness of the diagnostic.

Figure 3. Tornado diagram showing the top ten most sensitive variables



References

1 National Institute for Health and Care Excellence. 2023. *NICE health technology evaluations: the manual*. As of 22 March 2024: https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation

2 National Institute for Health and Care Excellence. 2024. *Guidance programme: Diagnostic guidance*. As of 22 March 2024: https://www.nice.org.uk/guidance/published?ngt=Diagnostics%20guidance&ndt=Guidance

3 National Institute for Health and Care Excellence. 2024. *Guidance, NICE advice and quality standards*. As of 22 March 2024: https://www.nice.org.uk/guidance/published?sp=on

4 World Health Organization. *Cancer*. As of 22 March 2024: https://www.who.int/health-topics/cancer#tab=tab_1

5 Husereau, D. *et al.* 2022. 'Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations'. *BMC Med* **20**, 23. As of 22 March 2024: https://doi.org/10.1186/s12916-021-02204-0

6 Cancer Research UK. 2024. *Cancer incidence statistics*. As of 22 March 2024: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence

7 McNamara, S., Schneider, P. P., Love-Koh, J., Doran, T. & Gutacker, N. 2023. 'Quality-Adjusted Life Expectancy Norms for the English Population'. *Value Health* **26**, 163-169. As of 22 March 2024: https://doi.org/10.1016/j.jval.2022.07.005

8 Tufts Medical Center. 2024. *The Cost-Effectiveness Analysis (CEA) Registry*. As of 22 March 2024: https://cear.tuftsmedicalcenter.org/

9 Cochrane (Homepage). 2024. *Cochrane Library*. As of 22 March 2024: https://www.cochranelibrary.com/

10 Office for National Statistics. *National life tables – life expectancy in the UK: 2018 to 2020*. As of 22 March 2024: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetables unitedkingdom/2018to2020

11 Office for National Statistics. 2019. *Cancer statistics explained: different data sources and when they should be used*. As of 22 March 2024: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/cancerstatisticsexplaineddifferentdatasourcesandwhentheyshouldbeused

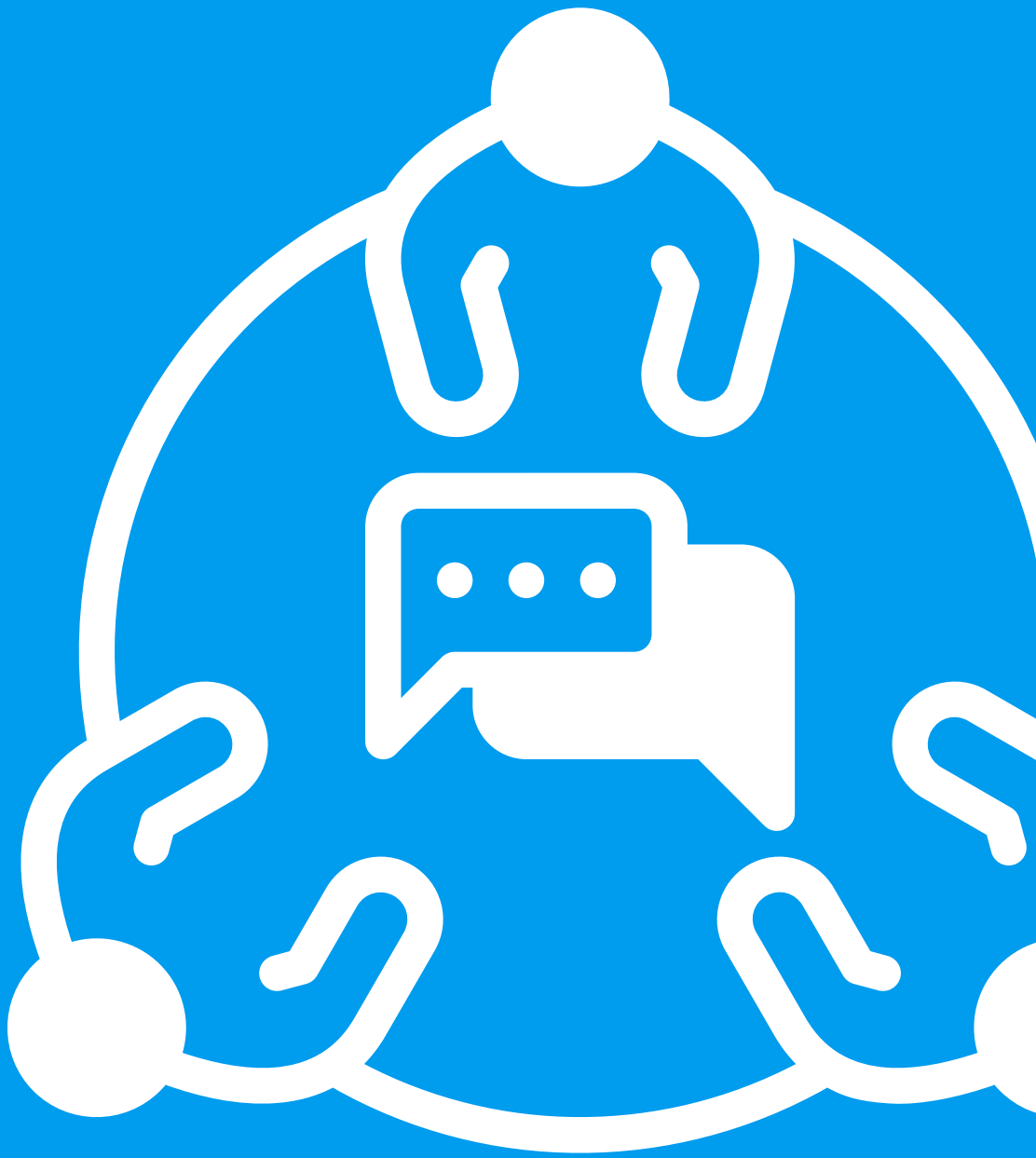
12 National Institute for Health and Care Excellence. 2023. *British National Formulary (BNF)*. As of 22 March 2024: https://bnf.nice.org.uk/

13 Jones, K. W. H., Birch, S., Castelli, A., Chalkley, M., Dargan, A., Forder, J., Gao, M., Hinde, S., Markham, S. Ogunleye, D. Premji, S., Roland, D. 2022. *Unit costs of health and social care 2022 manual*, Social Services Research Unit, Centre of Health Economics, University of York, Kent, United Kingdom. As of 03 April: https://eprints.whiterose.ac.uk/196043/1/Unit_Costs_of_Health_and_Social_Care_2022.pdf

14 NHS England. 2024. *National Cost Collection for the NHS*. As of 22 March 2024: https://www.england.nhs.uk/costing-in-the-nhs/national-cost-collection/

Annex F: Guide-Testing Workshop Document

Authors: Mark L Cabling, Jessica Dawney, Matthew Napier, Zuzanna Marciniak-Nuqui, Fifi Olumogba, Larry Kessler, Amanda Cole, Lotte Steuten, Sonja Marjanovic



1. Introduction

We facilitated an online workshop to engage the experiences and perspectives of the project's advisory group and individuals working in diverse stakeholder groups. The workshop explored a range of issues to help refine a draft of an oncology diagnostic draft Target Product Profile (TPP) guide developed based on previous work packages and provided to the participants before the workshop. Discussions focused on garnering suggestions and feedback on the format, content and framing of the TPP guide, including TPP features and priority areas. The workshop was held on 16 November 2023 and lasted 2.5 hours, conducted via MS Teams.

Annex F is the sixth of eight annexes complementing the main Cancer Research UK-funded project's final report: 'Advancing the development and use of diagnostic target product profiles for cancer.' The not-for-profit research institute RAND Europe led the project in collaboration with the Office of Health Economics. The project has benefited from ongoing support and advice from Professor Larry Kessler (University of Washington), a key consultant on the work. This document provides detailed findings and analysis of the stakeholder TPP guide testing workshop to which the final report refers; thus, Annex F, like all the other annexes, is primarily meant to accompany the final report and is not meant to be read as a standalone document.

In addition to participants from the project team and Cancer Research UK (the latter as observers), the workshops gathered insights from 16 individuals spanning the advisory group and external participants. The participants reflected diverse voices from academia and research, healthcare professionals, industry, policymakers, regulators, health technology assessment experts, health economics and patient, carer and public voice perspectives. Table 1 below provides further information. According to the Health Research Authority (HRA), the workshops did not require ethical approval. All principles of informed consent and legal requirements relating to data privacy and security were complied with.

Table 1. TPP development guide workshop participants

Overall, 24 participants attended the workshop:

- **Client commissioning staff:** CRUK client staff (n=4)
- **Research team** (n=4)
- **Workshop external participants** (n=16):
  - Project advisory board (n=8)
  - Academics (n=3)
  - Clinicians (n=2)
  - Industry (n=2)
  - Patient and Public Involvement (PPI) (n=1).

Participant	Affiliation
PROJECT CLIENT	
Dr Sarah Cook	Cancer Research UK
Samantha Harrison	Cancer Research UK
Dr Jessica Lloyd	Cancer Research UK
Sarah Brookes	Cancer Research UK

Participant	Affiliation
PROJECT RESEARCH TEAM	
Dr Sonja Marjanovic	RAND Europe
Fifi Olumogba	RAND Europe
Dr Mark L. Cabling	RAND Europe
Professor Larry Kessler	University of Washington
EXTERNAL RESEARCH PARTICIPANTS	
Advisory group	
Professor Michael Messenger	Leeds MedTech and In-Vitro Diagnostic Co-Operative
Dr Bethany Shinkins	University of Warwick
Professor Bernard Rachet	London School of Hygiene and Tropical Medicine
Dr Brian Nicholson	Nuffield Department of Primary Care Health Sciences, University of Oxford
Jacob Grant (deputising for Sarah Byron)	NICE
Dr Gillian Rosenberg	Innovation, Transformation Lead, National Cancer Programme, NHS England
Helen Dent	British In Vitro Diagnostic Association (BIVDA)
Dr Rebecca Riches-Duit	Medicines and Healthcare products Regulatory Agency (MHRA)
Academics	
Professor Phil Crosbie	Division of Infection, Immunity and Respiratory Medicine, University of Manchester
Professor Sue Mallett	Professor in diagnostic and prognostic medical statistics, University College London
Professor Jon Deeks	Professor of Biostatistics, University of Birmingham
Clinicians	
Dr Anthony Cunliffe	National Lead Medical Adviser, Macmillan Cancer Support
Dr Mark McCleery	Radiologist Consultant, NHS Greater Glasgow and Clyde
Industry	
Nishan Sunthares	Association of British Healthcare Industries
Ed Godber	Guardant
PPI	
Peter Clark	Lay PPI participant

We use the following abbreviations for each stakeholder group when referencing information sources in the following contents:

- Advisory Group (AG)
- Academics (AC)

Healthcare professionals (HP)

- Industry (IND)
- Patient, Carer and Public voice and Charities perspectives (PCPC).

The following section summarises key feedback regarding the format, content, and framing of the draft TPP guide. Throughout the document, we occasionally offer additional research team reflections on the insights and conversations during the workshops. When doing so, we clarify that these are our reflections using the term 'research team note' in brackets.

2. Feedback on features and prioritisation

2.1. General feedback on the first draft

2.1.1. Clarity of content

We considered and addressed all of the comments relating to the discussion of clarity of content in the final report.

Participants proposed that the draft guide should explicitly clarify its purpose and state who it is for (IND, AG, AC). Other suggestions included:

- Emphasising that implementation issues must be considered explicitly and upfront (IND).
- Clarifying who the document is for upfront and referring to relevant stakeholders early in the guide, not just in later sections.
- Highlighting the importance of accessibility in the report, e.g. noting that technical jargon should be explained in a TPP so that if lay people (e.g. PPI) access documents, they can understand, i.e. Patient, Carer and Public Voice and Charities Perspectives (PCPC).
- The global dimension of innovation must be overt in the guide, even if the project focus is UK-specific (IND, AC, AG). The following points were raised:
  - » Balancing international and UK-specific TPP considerations is important because clinical pathways and reimbursement differ between countries (AC, AG). International differences in reimbursement mechanisms would affect early economic modelling (AG).
  - » Equity of diagnostic access is vital to consider in TPPs. Both UK-specific populations and international perspectives must be considered, as the populations that suffer from inequities will be different in some cases in the UK compared to other countries and between countries/regions (AG, IND).
  - » There were some views that TPPs should have a global applicability when it comes to feature specifications and evidence for those specifications (AG, IND) [Research note: However, one participant used terminology that distinguished 'target product profiles' that, in their view, should have this global applicability, and 'target implementation profiles' that should have the more localised view. The discussion's context was issues with local adoption of an innovation, balancing a TPP's global vs. local focus, and the need for implementation considerations alongside technical criteria].

2.2. Content to add or make more prominent

We considered and addressed all comments relating to content additions and prominence in the final report:

- There were diverse views regarding whether or not payers (i.e. anyone who might pay for diagnostic tests in the NHS) should contribute their perspective to TPP development efforts. However, there was a general steer that efforts to develop TPPs need to be cognisant of and reflect procurement realities, e.g. how features related to economic considerations are addressed (AG).
- The life history of any TPP document must be clear. For example, how a TPP fits into the broader regulatory innovation and procurement landscape and process so that it is used (AG) It was noted that procurement and HTA organisations could have additional requirements for diagnostic procurement, not all of which might be considered at TPP specification stages (e.g. impact on patient outcomes) (AG).

2.3. Features

2.3.1. Human factors (separated into health professionals vs. expert users/patients/carers)

We considered and addressed all of the comments relating to the discussion of human factors of relevance in this section of the final report.

- Workshop participants flagged the importance of human factors as significant for both healthcare professional and patient-related features in a TPP (AG, AC).
- Workshop participants flagged the importance of ensuring balance in focus (AG, HP). They mentioned that clinical effectiveness, regulatory safety, risk management (e.g., adverse events and potential of misuse) and end-users' capability to use and interpret tests are also important (AG, HP).
- Patient acceptability, in particular, was flagged as key to identify within the TPP guide (AC).
- It was noted that qualitative exploration of the patient perspective in TPP development is limited (AG) [Research team note: This reflects our learning that patient-related features and engagement are areas needing improvement].
- Different types of modelling can help inform TPP features. For example, system challenges (e.g. waiting times) are easier to model than the HTA evidence requirements of long-term mortality and quality-of-life benefits because it is difficult to definitively link a diagnostic test with survival/QoL (AG). This can help in thinking about proxy measures of clinical utility.

3. Feedback on the TPP development process

3.1. General clarity of content

We considered and addressed all of the comments relating to the clarity of content in this section of the final report.

During a discussion about the challenges of consensus, it was noted that the TPP guide document should clarify that those developing TPPs should provide details on how each specification was derived, the evidence on which the specifications are based and the uncertainties surrounding them (HP, AG, AC).

3.2. Inception stage

Consulting various stakeholders is not just about developing the TPP but also about networking opportunities that may go beyond TPP development efforts (HP) [Researcher note: This participant was referring to how getting together and identifying a group of diverse stakeholders can also bring the benefits of networking and idea exchange to the fore, not just on working on the TPP. We clarified the communication aspect of TPP development value in the final report].

3.3. Implementation stage

Although the implementation stage of TPP development includes scoping, drafting and consensus phases, only the consensus phase was discussed by participants during the workshop.

3.3.1. Consensus phase

Key insights:

- What is important to find consensus on can depend on whether the TPP is international or UK specific, e.g. there may be some cases in which an internationally focused TPP would require consensus on features to do with global markets. In contrast, a UK-specific TPP would need consensus on features that are UK-specific (AG) [Research note: A participant continued the conversation previously initiated about the international vs. UK-specific scope of a TPP, showing that this theme appears to run throughout the workshop across many areas of TPP prioritisation, features, and development processes].
- Early economic modelling can help provide ranges for features. Some specifications might need to be presented as ranges rather than absolute values (AG) for consensus-building. Ranges could also provide the limits of acceptable trade-offs within a TPP for some specifications (e.g., accepting a sensitivity reduction of up to 'X' if the specificity increases) (AG).
- Participants brought up some challenges to consensus building in TPP development:
  - » Consensus must be evidence-based and, if possible, based on more than expert opinion (AG). [Research team note: This strengthens the need for multiple methods and types of evidence from desk research, expert consultation, and modelling to arrive at specifications before exploring consensus on them].
  - » It is critical to consider ways of handling dissenting expert opinions in consensus-building processes (AG).
  - » Consensus across stakeholder groups may be difficult for various reasons, such as differences in perspectives, expertise and interpretations of technical jargon (AG, PCPC, HP).
- Beyond identifying the challenges in consensus-seeking and building efforts, participants also identified ways to address challenges related to stakeholder diversity (HP, AG):
  - » One way to consider stakeholder diversity in the consensus exploration phase would be to engage separate

panels when exploring consensus (HP, AG) to capture views at the stakeholder level, giving voice to all, including the patient voice (AG).

  - Another way to help respond to diverse stakeholders' varying expertise and technical knowledge would be different versions of the same TPP, one with highly technical language and another with generalisable language and footnotes – or the same TPP with explanations in technical and lay terms (HP).

- Qualitative methods for exploring consensus may be valuable (e.g. through interviews and workshops), as opposed to only quantitative ones (e.g. rating-based surveys) (AG). It is important to capture information explaining the reasons behind stakeholders' views (AG).
- One participant mentioned the importance of unforced consensus, noting that there should be more consideration and elaboration on uncertainties and how to discuss uncertainties (as opposed to forcing consensus) (AG).
- Some key issues that must be clarified when presenting information on consensus in any TPP document include:
  - The minority opinion and reasons for that dissenting opinion (AG), as these nuances could prove helpful as TPPs evolve and new evidence or technology moves forward (AG).
  - The consensus thresholds were used in consensus-building efforts (HP).
  - Areas for which no consensus was reached (HP, AG), including a justification for why and how parameters were chosen without it (AG).
  - Evidence for consensus, which must be published (AG).

4. Feedback on the economic modelling tool

The workshop also discussed the economic modelling tool developed by the project team. Feedback and reflections were as follows, and all are considered and addressed as appropriate in the final report:

- There was a discussion about how and why the early economic tool is useful:
  - The tool can have significant value in helping to understand what characteristics drive cost-effectiveness (AG, AC, IND).
  - The tool can help to focus on what is plausible overall and what a range of likely values might be (AG, AC). A couple of participants opined that the tool also offers an opportunity to test different scenarios (AC).
  - The tool can efficiently utilise existing data, such as on diagnostic technologies evaluated by NICE, even if not recommended, because such cases could show what was required for a positive recommendation and feed into a TPP for a new test (AG).
- Participants also mentioned points regarding important considerations to keep in mind whilst using the tool:
  - The tool should not be used in isolation and must consider broader context (AG, HP, AC). Some necessary value judgements may override specific considerations (e.g. if it is deemed unacceptable for a new test's sensitivity to be lower than the test it replaces) (HP, AC). Other examples are where the test may address health equity issues (AG).
  - In tests targeting earlier diagnoses, it is imperative to account for overdiagnosis, where true-positive results lead to no patient benefits (AG).
- In addition to the tool's usefulness, participants also discussed its potential challenges. There were comments around the complexity of diagnostic introduction, where infrastructure and capital costs may be significant (IND, AG). It can be difficult for diagnostics to demonstrate value where medical value is fully captured in the price of a treatment pathway (AG) [Research team note: This is an issue for economic evaluation of diagnostics generally, rather than early economic evaluation specifically. Infrastructure modelling would be beyond the scope of this project].
- Points were raised about how to refine the tool:
  - There were suggestions to include more guidance on incorporating the early economic modelling elements into drafting a TPP and on drafting those statements specifically (AG).
  - There were suggestions around the possibility of making the model probabilistic, as often there can be lots of parameter uncertainty in the evidence-base (AGs) [Research group note: We would counter that probabilistic sensitivity analyses (PSA) would not be appropriate for this exercise because it would formalise the results in a way that would not match with the nature of the task, and potentially over-state the confidence in the results. PSA is more important to consider when the evidence base becomes more developed, and it is more important to characterise the uncertainty. The tool sits at the beginning of TPP development; hence, a probabilistic model is not helpful. Instead, a deterministic model (which is what our model incorporates) is more useful and relevant].

Annex G: TPP Feature-Related Considerations – Insights from Exploratory Desk Research and Stakeholder Consultation (Refinement of Annex B)

Authors: Mark L Cabling, Jessica Dawney, Matthew Napier, Zuzanna Marciniak-Nuqui, Fifi Olumogba, Larry Kessler, Amanda Cole, Lotte Steuten, Sonja Marjanovic

1. Background

Annex G is the seventh of eight annexes complementing the main Cancer Research UK-funded (CRUK-funded) project's final report 'Advancing the development and use of diagnostic target product profiles for cancer', led by RAND Europe and the Office of Health Economics. Like all the other annexes, Annex G is primarily meant to accompany the final report and is not meant to be read as a standalone document. Furthermore, this Annex is a refinement of the feature considerations in Annex B based on feedback and insights gained from further work packages after the initial scoping work underpinning it.

This document expands on information about features included in demand signalling diagnostic TPPs based on an exploratory analysis of prior efforts and stakeholder consultation. As discussed throughout the report, these efforts largely derive from the infectious diseases field (given the lack of publicly accessible diagnostic TPPs in the oncology space). However, our stakeholder consultation confirmed that many of the features considered in TPPs focused on infectious diseases would also be relevant for oncology use, even though the specifications for the features would be unique. That said, there will likely be some additional features that need considering in an oncology context, especially given that many (although not all) TPPs in other areas have referred to in-vitro diagnostics and that other test types, including multi-component platforms, imaging tests and Artificial Intelligence (AI) and digital-technology-informed diagnostics have a role to play.

This document seeks to inform those who might develop TPPs for specific cancer sites, test types, use cases and use settings about feature types which may be relevant to consider, but is by no means intended to be all-encompassing, nor does it suggest that all features are necessary for each effort. Prioritising a subset of relevant features is likely to be important in future TPP development efforts for diagnostic tests in oncology.

Section 2 of this document focuses on features identified in prior work and draws heavily on the in-vitro diagnostics space. It is based primarily on our analysis of supplementary material referred to in a recent systematic review by Cocco et al. (2020)[1] and additional research we undertook to test and validate their typology by analysing a sample of TPPs[2-9] (elaborated on in Annexes A and B). When considering the table of features, please note that we present the features used in prior efforts as an informative resource for those developing TPPs to consider. We are not judging whether they are suitable or whether some are more important than others, which depends on specific TPP use cases and test types. For example, according to one of our advisors, there is some controversy around whether analytical sensitivity is a robust measure, or whether limits of detection or quantification are more appropriate. Discussions around the suitability of terminology and features would need to be advanced in future research.

Section 3 outlines some additional considerations related to TPPs for digital imaging tests, multi-component tests and multianalyte platforms based on insights from stakeholder workshops (more specifically, the industry, advisory group and HTA workshop) and interviews conducted (more specifically, with INT8 and INT9) as part of the project. The scope of our project did not allow us to examine these test types in detail, but we have highlighted some initial considerations that matter for future research to build on and refine.

2. Features used in demand signalling TPPs: Insights from a systematic review and analysis of a sample of TPPs

Based on a recent systematic review by Cocco et al. (2020)[1] and our additional research to test and validate their typology, features used in diagnostic TPPs tend to fall into nine key conceptual categories. These conceptual categories and brief explanations of what they mean are summarised in Table 1 below. While they draw on areas other than oncology (where diagnostic TPPs as demand-signalling documents do not yet exist), our research suggests that the categories (and their respective features) will likely also be relevant in an oncology context. However, specifications for individual categories and features will be unique to specific tests and TPPs. Our research has also flagged historically neglected considerations within these categories, highlighted in the underlined text.

It is important to note that the evidence base on features in demand signalling diagnostic TPPs draws on efforts in the in-vitro diagnostics space. Future research will need to draw out unique features that might apply to other types of diagnostics, such as digital and imaging-based tests, including those with machine learning and AI capabilities and multi-component and multi-analyte testing platforms.

Table 1. Conceptual feature categories: An overview

Conceptual categories of diagnostic test features in TPPs	Explanations
1. Unmet need	Features related to the unmet need a diagnostic test must respond to and its scope of application, spanning information on intended use, the medical decision to be influenced, the medical need, target level of health system, fit with clinical workflows, target population, target user and description of the test concept. (Information on how a test might need to interact with or affect other testing needs and tests or downstream decisions about patient care is often missing).
2. Analytical performance	Features related to a test's ability to correctly detect and measure a disease analyte/marker, i.e. whether the test accurately measures what it needs to. This category includes 'output' features on analytical performance and features related to requirements to support appropriate analytical performance.

Conceptual categories of diagnostic test features in TPPs	Explanations
3. Clinical validity	Features demonstrating that what is being measured correlates appropriately with a physiological condition, pathological process or state, i.e. information to ensure that the test measures an appropriate marker of the disease. (Information on how the real-world conditions a test will be used in might differ from those in an experimental laboratory setting is often missing, i.e. a recognition that a test must consider real-world performance needs, not just those in lab settings).
4. Clinical utility	Features describing how a proposed test must influence downstream care outcomes, e.g. given features demonstrating links to survival, quality of life and patient experience (direct or indirect contributions to intended outcomes). (Proxy measures of clinical utility are often missing, an area needing further research in the context of cancer diagnostic TPPs)
5. Human factors	Features describing how individuals must interact with the test and how it needs to fit in with user skills and abilities (e.g. when administering the test to a patient, preparing it for use, interpreting and capturing results, and any training needs implied) can apply to how healthcare professionals and patients might interact with a test. (Considerations related to patient acceptability, accessibility and experience are often missing, including how the test needs to mitigate inequalities. Examples include requirements related to invasiveness, cultural acceptability for specific target groups and ease of access).
6. Infrastructure	Features describing requirements related to infrastructure (such as facilities, equipment, supplies and IT systems) or other operating conditions that must be established and maintained for the test's effective transport, storage, operation/use and/or disposal.
7. Costs/economic considerations	Features describing requirements related to economic and other commercial considerations (e.g. requirements related to what will likely be an acceptable price/affordability considerations and information on alternative tests and their costs/competitive landscape). (Consideration of cost-effectiveness as opposed to price is often missing, i.e. 'value for money', for which early economic modelling can help inform specifications).
8. Regulation	Features related to a test's safety, quality and efficacy/effectiveness requirements, considering regulations. (In addition, specifying features related to data governance, ethics and informed consent should not be neglected).
9. Environmental impact	Features related to requirements about the test's impact on the environment.

In the contents that follow, we reflect on the features and categories covered in the Cocco et al. (2020) review[1] (in supporting material) and highlight common features across TPPs (the * symbol denotes features that appear in half or more of the demand signalling TPPs included in the Cocco et al. review).

A range of features are covered within each category, as summarised in Tables 2–10. Some features appear in several categories; where possible, we have streamlined these to avoid duplication due to diverse terminology to describe conceptually similar/identical features in different TPPs. Some categories have a few features, while others have many.

We also provide 'working explanations' for features based on desk research, consultation within the project team and advisor comments. We checked the Clinical and Laboratory Standards Institute (CLSI) Harmonized Terminology Database for less clear features.[10] Not all terms featured in the database, which is not specifically TPP-focused. The explanations are not formal definitions, which would require a project in and of itself. Any project to define features in further detail would need to relate the definition to the nuances of each specific test type or use case to contextualise it.

Individual TPP development efforts will need to consider the level of granularity associated with any one feature. Thus, some features in the lists below may merit dividing into component features in specific TPP development efforts, e.g. separating infrastructure features associated with physical equipment from those associated with supplies/consumables.

We are not proposing that all demand signalling TPPs include all these features or that others may not be relevant. Instead, the tables below offer a reference and resource for what already exists in the landscape. As discussed in the guide draft, developing demand-signalling TPPs is time and labour-intensive, and a TPP cannot provide everything of interest to decision-makers and innovators. While the feature categories discussed below must be considered, developing demand-signalling TPPs must also be a feasible process, distinguishing between features central to a proposed test's value proposition for which specifications are essential (versus features where specifications can be more flexible and left to the innovator's discretion) and features which may not be relevant to the specific TPP in development.

Table 2. 'Unmet need' category

The 'Unmet need' category includes features related to the unmet need the diagnostic test responds to and its scope of application.

Features in the 'Unmet need' category	Explanation
Intended use*	Purpose of the test, i.e. **what it will be used for:** 'function and intended use of the product' and whether the product 'may be administered…with a view to… achieving a medicinal purpose.'[11]
Medical decision(s) to be influenced	**The medical decision the test can inform/ influence** and how it will achieve its purpose, e.g. deciding whether a person has a disease or not, deciding how to prioritise patients for further care (e.g. whether to refer to secondary care or to put on an urgent treatment list, etc.).
Medical need	Information on **why the test is needed** and what unmet need it addresses, potentially including the scale of need and the rationale for developing the test.
Target population*	Information on **who will be tested** (i.e. the eligible population based on risk factors, symptoms, demographics, etc).

Features in the 'Unmet need' category	Explanation
Target user*	Information on **who does/administers the test** (e.g. a lab technician, a nurse, a patient at home, or possibly more than one person; for example. one person might collect the sample, and another conduct the test).
Fit with clinical workflow	**How the test fits into existing patient care pathways** and processes and whether it disrupts them positively or negatively, e.g. whether patients would see different healthcare professionals, whether healthcare professionals would assume new responsibilities and whether some patients would exit the system).
Target level of health system*	Information about **whether the test is meant to be used in specific settings, e.g. primary care, acute care, community care or at home.**
Description of the test concept	Information on **the possible range of technologies for accomplishing the test's aims** (as far as can be specified, e.g. sample type and analysis). An example is genomic sequencing of lab blood tests for chemical analysis.
Proof of concept	Information on the test's proof of scientific concept, i.e., evidence proving its feasibility based on prior research and trials, **or the evidence type needed to ensure proof of concept.** (Though currently grouped under 'Unmet need', this feature could also fit under the 'Clinical validity' category).

*Denotes features that appear in half or more of the demand-signalling TPPs included in the Cocco et al. (2020) review.

Table 3. 'Analytical performance' category

The 'Analytical performance' category includes features related to a test's ability to correctly detect and measure a disease analyte/marker, providing information about whether the test accurately measures what needs to be measured. This category includes 'output' features on analytical performance and features related to requirements to support appropriate analytical performance.

Features in the 'Analytical performance' category	Explanation
Subcategory: Key test features and analytical performance information	
Assay (test) design/ format	Diagnostic test type, e.g. molecular, serologic, antigen-detection, etc. Individual test-kit components may also belong in this category.
The target molecule for detection	Information about the specific molecule the test aims to detect.
Analytical specificity	A test's ability to specifically detect the intended specimen/ substance, not others.
Analytical sensitivity/ limit of detection	Information about the smallest quantity of analyte necessary in a sample for accurate disease detection.

Features in the 'Analytical performance' category	Explanation
Strain specificity	A test's ability to accurately detect specific strains/distinguish between different disease strains.
Cross-reactivity	The extent to which different disease markers appear similar in the test (e.g. possible sources of false positives).
Reproducibility	Whether a test gives the same result across repeat samples from the same person. Other important aspects include the test's reproducibility across test settings and near the clinical threshold/ clinical value). (Some efforts classify this factor under precision, categorising it as a 'Clinical validity' feature).
Robustness and interferences	Information on how far a test is affected/unaffected by changes in test conditions as an indicator of its robustness, alongside information on when a test result may be falsely altered (interferences).
Control or comparative reference method	Information about the other test types / reference methods that a new test's performance must compare favourably against (e.g. existing tests/ diagnostic methods).
Indeterminate test results	Information about whether a test can/does give invalid results and, if so, how and why. Some tests give neither positive nor negative results but inconclusive ones.
Device failure/invalid rate	Information on the cases/conditions/rates for which the device fails to give a result. This differs from an indeterminate result. (e.g. only showing part of a result).
Duration of valid sample	Information on how long the sample can be used once collected (e.g. blood, saliva, etc).
Time to test result*	Information on the length of time between completing a test and seeing a result.
Duration of valid result	Information on how long the reading is valid for, e.g. whether the test is stable over time or whether the disease changes over time, such that the result may not apply after 'X' time. This may be similar to result stability (to discuss in meeting).
Result stability	This refers to 'the degree to which a diagnosis is confirmed at subsequent assessments.'[12]
Subcategory: Operational and analytic requirements to reach desired analytical performance	
Sample/specimen volume	The sample amount needed (e.g. volume of blood, saliva, etc) for testing.
Sample type*	Information on the sample type, e.g. blood, saliva, urine or other tissue, etc.

Features in the 'Analytical performance' category	Explanation
Sample/specimen preparation*	The steps, processes and conditions necessary to prepare the sample for testing, including a description of what (if any) specific volume is required.
Throughput	The number of tests that can be completed in a specific period. This may apply to various aspects of throughput (e.g. platform, human capacity, sample/specimen throughput).
In use stability	Information on whether the material the diagnostic test uses (e.g. reagents) remains stable during use and for how long it can produce an accurate result.
Type of analysis	Information on the type of analysis to be conducted (e.g. qualitative, quantitative, chemical, genomic, etc.).
Quantification/ quantitation	Whether a test provides a quantitative measure/result, e.g. on the severity/ spread/quantity of disease.
Result (format and readout)	Information on how the format of the test result is conveyed/ displayed.
Multiplexing	Information on the process of simultaneously detecting/identifying multiple biomarkers in a single test.
Kit-quality indicators	Elements of the test kit that indicate any degradation in the test components' ability to do their job.
Reagent integration/ preparation	How to prepare the test reagent for use and how to package it.
Reagent kit (nature, transport, storage and stability, supplies not included in the kit)	The nature of the reagent kit and how it should be transported, stored, and kept stable, as well as information on which supplies are not included in the kit and must be acquired externally.
Calibration	Information on correctly setting up and calibrating the test for accurate use.
Quality control (internal and/or external)	Information about the quality control procedures and requirements associated with test use, including internal and external quality control.

*Denotes features that appear in half or more of the demand signalling TPPs included in the Cocco et al. (2020) review.

Table 4. 'Clinical validity' category

The 'Clinical validity' category includes features related to whether a test's primary measure(s) appropriately correlate(s) with a physiological condition, pathological process or state, i.e. whether the test measures an appropriate disease marker.

Features in the clinical validity category	Explanation
Diagnostic/ testing sensitivity*	Diagnostic sensitivity is **a test's ability to correctly identify individuals with the disease (i.e. not yield false negatives).** Thus, it represents the probability of a positive diagnostic test in a person with the illness.
Diagnostic/ testing specificity*	Diagnostic specificity is the probability of a negative diagnostic test in an individual who does not have the disease (i.e. not yield false positives). Thus, it represents **a test's ability to correctly rule out those without a disease.**
Positive predictive value (PPV)	PPV is the **probability of a confirmed diagnosis among those testing positive**. (PPV and NPV depend on disease prevalence; in contrast, sensitivity and specificity are not prevalence-dependent).
Negative predictive value (NPV)	NPV is the **probability of a confirmed diagnosis among those testing negative**. (PPV and NPV depend on disease prevalence; in contrast, sensitivity and specificity are not prevalence-dependent).
Field performance	Information on **how well the test performs in the real world rather than a controlled laboratory environment.**
Precision/ concordance	**The agreement level between a test's results and other tests applied to the same sample/individual. This feature is closely related to validity and how a test's results compare against a recognised gold standard** (where this exists).
False recent ratio (%)	Information about **the proportion of diagnosed cases falsely classified/ misclassified as recent.**
Test performance with disease subgroups	**Information about a test's performance in distinct patient profile types** (i.e. patients with different demographics or different stages/severities of the disease or demographics) to include performance information across certain demographic groups.
What is the risk of an inaccurate test result?	Information on the types of conditions (e.g. human user, sample, operating environment conditions) that could lead to inaccurate results and, if possible, information on the risk level to patients.

*Denotes features that appear in half or more of the demand signalling TPPs included in the Cocco et al. (2020) review.

Table 5. 'Clinical utility' category

The 'Clinical utility' category includes information about whether a test will positively affect the intended outcomes, e.g. patient quality of life and longer lifespan, and how (direct or indirect contributions to the intended outcomes).

Features in the 'Clinical utility' category	Explanation
Intended outcome and linkage to care	Information about the type of patient outcome the test is meant to contribute to and how it will likely link to care pathways. For example: • Is it meant to enable faster/earlier diagnosis or better triage? • Will it support outcomes such as longer life, better quality of life, etc.?

Table 6. 'Human factors' category

The 'Human factors' category includes information about how individuals need to interact with the test (e.g. when administering it to a patient, preparing it for use and interpreting and capturing results, and any associated training needs).

Features in the 'Human factors' category	Explanation
Subcategory: General use related human factor considerations	
Test size and portability	Information on the physical size and weight of the end product. This could include information on the test's sensitivity to external factors such as temperature, movement, storage and dampness that can impact its portability and implications for human use.
Equipment-specific human factors	Information about how an individual must engage with the test to appropriately operate the equipment.
Patient identification capability	Information on the in-product capability for patient identification, i.e. the process of correctly matching a patient to an appropriate intervention/test and communicating information about their identity accurately and reliably throughout the care continuum (e.g. radio-frequency identification tags).
Safety precautions (biosafety requirements)	Information about the biosafety requirements that must be in place for the test's safe operation.
User/use-induced failure rate	Information on the rate of human error in operating the device/product.
Ease of test result interpretation	Information on human involvement in interpreting/analysing the test results (analysis type, complexity and simplicity).
Rate of errors in device interpretation	Information on the rate of human error in interpreting the test results.

Features in the 'Human factors' category	Explanation
Subcategory: Test operation specific (other than data-related)	
Training and education*	Information on any training/education the user must have to effectively engage with any aspect of the test, e.g. preparing the kit, collecting a patient sample, administering the test and interpreting the results.
Tool format and complexity	Information about the product complexity level, i.e. how specialised the user has to be to use the product correctly. This may include information about any associated skills needs.
Hands-on time	Information about the time needed to conduct the medical test.
Labelling	Information on designing, reviewing, producing and attaching labels for the tests.
Walkway operation	Information on whether the assays must be supervised or can be left for hours/days (as in the case of cultures, for example) before returning to complete them.
Instruction for use	Information on the necessary instructions for using the test/test kit.
User interface	Information on the platform through which the test user(s) inputs various data (patient ID, test time/date/location and results), e.g. a computer or portable tablet.
Language	Information on the language in which the device/test is programmed to operate (this could also refer to the language the instructions are in).
Subcategory: Data-related human factors	
Data capture	Information on data to capture (e.g. data related to the test result, time, date or patient) and where and how to capture it.
Data handling	Information on gathering, recording and presenting information of relevance to the test (data resulting from test results or referring to patients).
Data input	Information on any human involvement needed to input data into the product.
Data export (connectivity and interoperability, electronics and software)	Information on how data from the product should be exported by a human user, including connectivity, interoperability, electronics and software details.

*Denotes features that appear in half or more of the demand signalling TPPs included in the Cocco et al. (2020) review.

Table 7. 'Infrastructure' category

This category includes requirements related to infrastructure (such as facilities, equipment, supplies) or other operating conditions that must be established and maintained for the test's effective transport, storage, operation/use and/or disposal.

Features in the 'Infrastructure' category	Explanation
Subcategory: Basic conditions related to infrastructure	
Operating conditions	The external conditions necessary to support the test's operation (including storage, stability and any supplies not included in the test kit).
Biosafety requirements	The biosafety requirements that must be in place for the test's safe operation.
Cold chain	The chain of events in temperature-controlled environments needed to store, manage and transport tests, including cold-chain equipment and facilities requirements.
Thermal tolerance of assay/test	The optimal temperature for the test, including information on its capacity to tolerate temperature change.
Temperature and humidity*	The optimal temperature for the test, including information on its capacity to tolerate temperature and humidity changes.
Environmental tolerance of packaged test kit	The packaged test kit's tolerance to different environmental factors, such as temperature and humidity, etc.
Clean water	Information on whether clean water is necessary for the operation of the medical test.
Power requirements*	Information on the power supply necessary to support the operation of the medical test.
Shipping conditions	Information on the shipping conditions needed for the medical test during transport.
Stability during transport*	Information on the stability of the medical test during transport.
Waste disposal*	Information on the waste disposal practices of the medical test, including any potential special arrangements necessary, e.g. in the case of toxic or hazardous materials.
Subcategory: Specific infrastructure conditions needed for operations	
Storage conditions and shelf life*	Storage conditions necessary for the medical test, including information on its shelf life and conditions necessary prior to use.
Equipment and supplies	The equipment and supplies necessary to support the operation of the medical test (and not included in the test kit). This includes medical supplies and/or durable medical equipment required to operate and administer the test without being an integral part. Note: Individual TPP efforts may want to separate equipment from supplies and different supply types from each other.
Multiuse platform	Information about the test platform that can be applied to multiple markers.

Features in the 'Infrastructure' category	Explanation
Reagent kit (nature, transport, storage and stability, supplies not included in the kit)	Information on the nature of the reagent kit and how it should be transported, stored and kept stable; information on which supplies are not included in the kit and must be acquired externally. Note: We have left this separate for now, but individual TPP efforts may want to distinguish between information on the nature of a reagent kit and how it is stored, transported and kept stable. Note: This also appears under the 'Analytical Performance' category; it is unnecessary in both and may be better in one or the other.
Assay packaging	Information on the assay packaging and safety seal used to confirm the contents have not been tampered with and are authentic. This is also covered under 'Human factors' but may not be needed in both.
Maintenance (including servicing and support)	The external or internal maintenance required to operate the test. Some tests will require distinct types of external and internal maintenance. This can include information on the servicing and support needed to maintain the product.

*Denotes features that appear in half or more of the demand signalling TPPs included in the Cocco et al. (2020) review.

Table 8. 'Cost and economic considerations' category

This category covers features related to a test's economic costs and other commercial considerations.

Features in the 'Cost/economic considerations' category	Explanation
Price/cost of individual test*	Information on an individual test's overall price/cost to the payer (NOT the cost of production).
Cost per diagnosis	Information on a test's overall cost per diagnosis. Note: This could include broader information on costs beyond the individual price/cost per test to consider how the test is used in practice (e.g. the impact of training costs and the number of individuals that need screening to find one case on the costs per test).
(Capital) cost per instrument	Information on the fixed one-time costs associated with, for example, purchasing the instruments, equipment or infrastructure needed to run the tests.
Cost of consumables	Information on the costs of ongoing materials needed, including reagents.
Cost of manufacturing a single-use device	Information on how much it costs to manufacture a specific single-use testing device (specifically manufacturing costs).

Features in the 'Cost/economic considerations' category	Explanation
Expected scale of manufacture	Information on how much / how many units will be made.
Potential market: size, nature and segmentation	Information on the market size, the potential number of users, and different market segments (e.g. geographical markets or different user types, such as hospital versus home-based). Alternatively, in the case of multiplex platforms, this could also cover uses for different diseases.
Routes to market	Information on routes to market (e.g. who the payers are).
Competitive landscape	Information on the costs of other available tests on the market.

*Denotes features that appear in half or more of the demand signalling TPPs included in the Cocco et al. (2020) review.

Table 9. 'Regulation' category

This category covers features related to a test's regulatory requirements and pathways information.

Features in the 'Regulation' category	Explanation
Regulatory requirements	Information on the specific regulations a test must meet given its target market. Note: Individual TPP efforts may develop specific subcategories to cover the extensive regulatory requirements regarding specific oncological tests or parts of tests.
Product registration path	Product registration is the initiation of any regulatory process. This feature covers information on how to begin the regulatory process, who to register with/which agency and which areas the TPP applies to/focuses on/is relevant for.

Table 10. 'Environmental impact' category

This category covers features related to a test's impact on the environment.

Feature in the 'Environmental impact' category	Explanation
Environmental footprint	Information about the environmental impact of a test's production and use. Individual TPP efforts may want to distinguish between the environmental impacts associated with manufacture and those associated with use.

3. Considerations for TPPs for digital imaging tests, multi-component tests and multianalyte platforms

3.1. Considerations for digital imaging tests, including those with artificial intelligence and machine-learning capabilities

Recently, there has been growing interest in applying digital technologies to cancer diagnosis, particularly in digital imaging and screening combined with clinical decision support software and/or artificial intelligence (AI) and machine-learning capacities.

Many attendees in stakeholder workshops brought up the use of AI for cancer screening and diagnosis as an important domain in diagnostics that is quickly gaining traction, bringing into play specific considerations relevant to a TPP. Attendees also recognised that methodological and ethical research on AI is scarce and unsuitable for informing TPPs for diagnostics that can make it to market, including potential biases in AI from closed AI systems and how such biases compare to biases in clinical judgement. Participants also flagged a need for any TPP using AI to make clear requirements for tool development, especially concerning data informing algorithm development. Data conveyance from AI systems can also be challenging, and attendees highlighted specifications for screening (object recognition), format and display features as important. Participants also noted the importance of clarifying explicit links between AI and medical decision-making, e.g. whether AI and image data will be used to aid diagnosis or as a triage tool. The challenges of a TPP specifying regulatory requirements or signposting information on them were also identified, especially given the fast-changing and evolving regulatory landscape.

The research team also consulted experts from Team Consulting, a consultancy specialising in medical device design and development, and they highlighted some important issues to cover when considering features to specify in TPPs for diagnostic tests involving digital technology, AI and machine learning. We are particularly grateful to Ben Cox, Charlotte Harris, Thorbjorg Petursdottir and Thomas Watts from Team Consulting for sharing their insights (named with permission).

Examples of the information to consider specifying features for in digital diagnostics that they flagged as essential included:

- Information on the platform type the test would operate on (e.g. IOS/Android)
- The need for integration with other devices and systems
- Cybersecurity and data-privacy considerations
- Unique user-skill and data-interpretation needs.

Numerous considerations were also flagged as relevant for demand-signalling TPP development for machine learning and AI imaging-based tests, including:

- Data collection and processing (e.g. whether the AI/model will need to build on pre-existing or new data).
- Whether the data set needs to be representative of gender, ethnicity and other variables to help mitigate bias.

- Any potential needs for data augmentation (e.g. capacity to rotate images to allow the model to better generalise insights).
- Data labelling (to ensure consistency and accuracy).
- Establishing truths (is this based on the test itself or verified by a doctor).

With evolving AI environments and their applications for diagnosing disease, we are beginning to learn more about the data-related considerations relevant to a TPP that engages with these technology types. This includes considerations related to the underlying algorithms, such as content transparency, quality control, algorithm validation and machine learning element requirements. It also includes data-related features (e.g. data capture, validation, ownership, storage, recovery, flow, reporting, provenance, dictionary, security and privacy).[7]

Developing a machine-learning model also brings to mind additional considerations concerning:

- Whether innovators must use their own or established external neural networks for image processing (and if so, licence information, accuracy metrics and criteria, inference time and support specifications)
- How to select an appropriate neural network for a given task (e.g. object detection or image classification).

Finally, such a TPP must consider specifications related to human oversight and regulation:

- The human-machine interface is also important to ensure that specifications will enable users to understand the provided outputs and decide whether to use the AI system outputs or override them. The data/analytics format and display, as well as the analytics and criteria for user-friendly data conveyance, come into play here.

3.2. Additional considerations for multi-component diagnostic tests (using an example involving digital technologies)

We also considered an example of a multi-component test demand-signalling TPP we identified through our scoping work. We deliberately chose the example of a test involving digital technology, as these tests are more novel and have not featured prominently in prior literature on diagnostic TPP development and features.

Although the scope of the current project does not allow us to consider complex digital diagnostics in detail, we wanted to explore unique key learning that might stem from such a space alongside learning related to a multi-component test. To do this, we looked at an example TPP for an electronic-clinical-decision-support-algorithm-based (CDSA-based) test, which also incorporates the use of point-of-care diagnostic tests as part of a multi-component kit.[7]

The TPP defined a toolkit consisting of the clinical decision support algorithm and point of care tests to support evidence-based clinical decisions by capturing patient, clinical and contextual data and diagnostic test results to arrive at diagnosis and patient care needs recommendations. The algorithm integrates the diagnostic test results with the other relevant information, all embedded in an app.[7] The TPP reflected several unique considerations and a bespoke structure:

- The first category of features covered was **characteristics describing the general scope of the test**, including features typical of many TPPs [and what Cocco et al. (2020)[1] classified under the 'Intended use' category]. In this case, features included intended use, target population, setting and target end-user specifications.
- The second category sought to cover characteristics that describe the **core components of the test kit** (an algorithm,

associated point-of-care testing tool, compatible devices on which the app would function and related operating systems, and app-related features in this case). Features within this category primarily focused on the key component characteristics essential for an accurate and clinically useful test and most broadly correlating to general features in the Cocco et al. (2020) framework regarding analytical performance (accurately measuring/capturing what it intends to capture – patient, clinical and contextual data and diagnostic test wise), clinical validity (measuring/capturing and conveying the correct information) and clinical utility (informing the appropriate clinical decision-making for desired patient care outcomes), as well as regulatory requirements. Given the test's nature, this related to features such as algorithm access format/design (i.e. access via an app), content informing the algorithm (e.g. underlying data input requirements for the algorithm to be informed by credible, clinically valid evidence), information related to treatment recommendations (e.g. compatibility with national guidelines to provide appropriate clinical validity and utility), information on compatible/additional associated diagnostic tools to be used/prompted by the app (e.g. point-of-care tests to support clinical validity), regulatory considerations for diagnostic tools, information on compatible device requirements (e.g. tablets, phones and laptops) and compatible operating systems.

- Thereafter, the TPP took different test-kit components as an overarching category and discussed remarkably diverse features for each category. Thus, the TPP includes categories for the **clinical-decision-support algorithm element, the point-of-care tool element, the app component/device element, and the data component/element** separately. Each includes diverse features, many of which speak to categories broadly correlated to those in the Cocco et al. (2020) framework, even if not organised as such. For example, some features are compatible with indicators covered in the Cocco et al. (2020) framework under 'Analytical performance' and associated operational requirements, others under 'Clinical utility' and others under 'Human factors', 'Infrastructural requirements' and 'Regulation'. Procurement is also mentioned, but no further details are provided.

In reflection, while some overarching feature categories covered in the Cocco et al. (2020) framework may apply to multi-component and digital technology diagnostics (e.g. intended use, environmental impact, cost and health-economics-related information), it may be necessary to consider other feature categories at the individual test-component level (e.g. addressing the digital-imaging device, AI/machine-learning and decision-support software and app components separately). This separation might apply to features concerning unique analytical performance (whether the relevant component is accurately capturing and measuring what it aims to), clinical validity (whether the relevant component is measuring and capturing the correct information), clinical utility (whether the appropriate component is influencing the patient care pathway and outcome as intended), human factors (e.g. training, instructions for use, result format/visualisation and interpretation) and infrastructure and regulatory requirements.

What is common across test/TPP components and what may be unique to specific components may vary on a case-to-case basis. In addition, regulation of software as a medical device is evolving; thus, regulatory specifications may present additional layers of complexity for such TPP specifications, depending on regulatory areas.

3.3. Diagnostic platforms versus individual tests

Multianalyte tests that can check for diverse cancer types simultaneously have the potential to significantly improve early diagnosis prospects and complement the screening assays that genomic laboratories already undertake. Multianalyte blood tests that could detect multiple cancers from a blood sample may also help meet the need for tests that GPs can easily administer in primary care screening. Our research also highlighted that

such diagnostic platforms give rise to unique considerations compared to TPPs for individual diagnostic tests. Examples include:

- **Clarifying the required sensitivity and specificity requirements, especially given the potential risk of multianalyte tests under-diagnosing cancers:** e.g. there are concerns that reasonable specificity may come at the expense of sensitivity.
- **Clarifying the required sample-size collection:** Testing for multiple cancers may require different sample sizes than single-analyte tests.
- **Clarifying specifications about the necessary laboratory infrastructure capacity and staff skills:** Additional infrastructure requirements are necessary to accommodate these testing modalities.

References

1. Cocco, P., *et al.* 2020. 'Target Product Profiles for medical tests: a systematic review of current methods'. *BMC medicine*. **18**(1): p. 1-12.
2. Dailey, P.J., *et al.* 2019. 'Defining System Requirements for Simplified Blood Culture to Enable Widespread Use in Resource-Limited Settings'. *Diagnostics*, **9**(1): p. 10.
3. World Health Organization, FIND and Medecins sans frontieres. 2020. *A Multiplex multi-analyte diagnostic platform*.
4. Kadam, R., *et al.* 2020. 'Target Product Profile for a mobile app to read rapid diagnostic tests to strengthen infectious disease surveillance'. *PLoS One*, **15**(1): p. e0228311.
5. Mather, R.G., *et al.* 2019. 'Redefining typhoid diagnosis: what would an improved test need to look like?' *BMJ Global Health*. **4**(5): p. e001831.
6. Program for Appropriate Technology in Health. 2018. *Diagnostics Instrument–Target Product Profile. Diagnostic Instrument: Hemoglobinometer*. Seattle, WA, USA
7. Pellé, K.G., *et al.* 2020. 'Electronic clinical decision support algorithms incorporating point-of-care diagnostic tests in low-resource settings: a target product profile'. *BMJ Global Health*, **5**(2): p. e002067.
8. Vetter, B., *et al.* 2021. 'Development of a target product profile for a point-of-care cardiometabolic device'. *BMC Cardiovasc Disord*, **21**(1): p. 486.
9. World Health Organization. 2020. *Target product profiles for antibacterial resistance diagnostics*. 2020. World Health Organization.
10. The Clinical and Laboratory Standards Institute. 2024. *Harmonized Terminology Database*. As of 22 March 2024: https://htd.clsi.org/
11. Medicines and Healthcare Products Regulatory Agency. 2020. *A guide to what is a medicinal product.* MHRA: London, UK.
12. Kim, W., *et al.* 2011. 'The Diagnostic Stability of DSM-IV Diagnoses: An Examination of Major Depressive Disorder, Bipolar I Disorder, and Schizophrenia in Korean Patients'. *Clin Psychopharmacol Neurosci*, **9**(3): p. 117-21.

Annex H: Stakeholders to Involve in TPP Development Document

Authors: Mark L Cabling, Jessica Dawney, Matthew Napier, Zuzanna Marciniak-Nuqui, Fifi Olumogba, Larry Kessler, Amanda Cole, Lotte Steuten, Sonja Marjanovic

Introduction

Annex H is the eighth of eight annexes complementing the main Cancer Research UK-funded (CRUK-funded) project's final report: 'Advancing the development and use of diagnostic target product profiles for cancer.' The not-for-profit research institute RAND Europe led the project in collaboration with the Office of Health Economics. The project has benefited from ongoing support and advice from Professor Larry Kessler (University of Washington), a key consultant on the work. This document provides detailed descriptions of which stakeholders can be involved in Target Product Profile (TPP) development efforts that the final report refers to; thus, Annex H, like all the other annexes, is primarily meant to accompany the final report and is not meant to be read as a standalone document.

1. Academic/research expertise

Why it matters: Different types of clinical, natural and social science research expertise may have a role to play in specifying technical performance requirements for a novel diagnostic test and economic considerations and criteria related to ensuring a diagnostic's good fit within healthcare services.

Which types of expertise might be relevant: Clinical research, scientific and technical expertise related to a specific cancer site and diagnosis in that area is essential for ensuring a TPP defines evidence-based specifications for technical performance criteria and responds to a clearly articulated unmet need. Such expertise can be found in academic departments with active cancer research across the UK, as well as in specialised research institutes (such as those funded by CRUK, including CRUK research centres) and in research networks (e.g. Academic Health Science Networks [AHSN], or cancer-specific networks such as the CanTest Collaborative). Health economics, statistics and modelling expertise can help in early economic assessment of the potential health and economic value and trade-offs involved in meeting different combinations of specifications for features outlined in a TPP (e.g. diagnostic sensitivity and specificity, accessibility and cost criteria). This process can help identify features and specifications with the highest impact on a novel test's overall value proposition. Expertise in implementation science and social sciences can help shed light on specifications related to a potential diagnostic's fit in clinical and care pathways (e.g. by helping make related user-skills and training requirements clear) and in meeting different populations' use requirements (e.g. inequality-related considerations that may link with the design features of a potential novel test). Human factors and implementation science expertise can also be helpful.

2. Healthcare professional/ diagnostic laboratory expertise

Why it matters: Healthcare professionals can provide insights on areas of unmet need in terms of existing tests' 'technical' performance (e.g. diagnostic sensitivity and specificity), their fit with the healthcare service's clinical and care pathways and usability in a given context (e.g. the specifications a novel test needs regarding the healthcare service pathway and organisation it must align with to support coordinated, timely and accurate diagnosis and an appropriate patient experience, including considerations concerning required workforce capacity, skills and workflows). Therefore, healthcare professionals can be key in defining and specifying the value proposition regarding technical and human factors and adoption-context considerations. Healthcare professional representatives on regional bodies such as cancer alliances and Integrated Care Boards may also be well-informed about procurement channels and payer realities.

Which types of expertise might be relevant: The types of healthcare professionals to engage will depend on TPP use cases (e.g., test types, cancer sites and use settings in primary or acute care). Specific diagnostic laboratory expertise (such as pathology and/or genomics laboratory expertise) can be relevant to help identify requirements for ensuring a diagnostic test aligns with existing laboratory workflows and infrastructure and meets the necessary quality and safety criteria related to sample preparation, specimen volume, handling, storage and transport requirements. The same applies to imaging expertise for imaging-based tests. In addition, networks and networked organisations such as cancer alliances (and, where applicable, leads on cancer-integrated care boards in integrated care systems) and potentially professional royal colleges may also have a role as a voice for healthcare professionals. Some consideration must also be given to the clout and credibility of experts approached to be on the core working group.

3. Industry expertise

Why it matters: Industry will be a core stakeholder group that directly uses demand-signalling TPPs to guide their product developments and respond to the criteria set out in the TPP. An industry perspective is vital for ensuring the specifications are feasible, especially regarding a novel test's technical performance criteria, associated infrastructure, equipment and supplies requirements and commercial (i.e. payer and price) considerations. Industry will also be aware of other products on the market and can input that knowledge into early efforts in TPP development to confirm an unmet need.

Which types of expertise might be relevant: Industry consultation to inform TPP development should seek to gain a diversity of views, as it is essential that a TPP is usable for industry and enables a competitive market and level playing field for innovation. This includes consulting different types and sizes of companies (e.g. small and medium enterprises [SMEs], not just large companies). The industry engagement needed will depend on the test type for which a TPP is being developed (e.g. in-vitro diagnostic or imaging). Although all individuals and groups can be sources of bias (and TPPs can and should seek a diversity of views but not statistical representativeness), insights from the project suggest exercising particular caution to minimise the risk of bias when consulting industry to inform TPP specifications. It is important to consider diversity, potentially involving umbrella organisations such as trade associations (e.g. The British In Vitro Diagnostic Association [BIVDA] and Association of British Healthtech Industries [ABHI]). The timing of industry consultation also matters. Some participants in this project felt that industry should be consulted in later TPP development stages only once drafting is complete (to elicit views on the appropriateness of a draft and explore agreement) rather than formally engaging industry in specification consensus. Others felt it important to engage industry from the very beginning because they know what is on the market and what is commercially viable. However, industry must declare any conflicts of interest.

4. Patient and carer representation

Why it matters: Consulting patients and/or their representatives early in the TPP development process can help understand the unmet needs a novel test must meet regarding end-user experience, accessibility, and acceptability. The insights obtained through our work highlight that the patient-and-public voice has been a lower priority than it could and should be in efforts to develop diagnostic TPPs in other disease areas. This area is ripe for improvement because patient experiences are vital in understanding the needs and inequalities a novel test can respond to, mitigate or inadvertently exacerbate. Our work has also identified the features in a TPP where a patient/patient representative perspective would be helpful (as discussed in Section 4.1.6).

Which types of expertise might be relevant: It is important to seek diversity in who contributes to the patient-and-carer voice and on what issues, including through efforts to engage underrepresented groups in a relevant way for the specific TPP use case in question. It is important to try engaging voices that may be particularly underserved or inaccessible (e.g. different cultural and ethnic groups or people with learning disabilities), as patient and public involvement often involves self-selecting groups that are insufficiently representative of patients overall. It is also essential to be mindful that not all relevant contributors may view themselves as patients, especially in earlier stages of diagnosis, and that the views of patients' carers and families may also be necessary to consider. Relevant insights can come from expert or lay patient representatives and patient networks (panels convened by charities, healthcare patient panels and peer support groups), carers and charities. It can be challenging to gather patient and public representative views due to time restrictions involved in TPP development and a need to create effective ways for patients to communicate and articulate their needs. The type of patient, carer, and public voice representation most appropriate for the working group will be context-dependent (on cancer type and test use case). Engaging the patient, carer, and public perspectives must be commensurate with the issues they can speak to. However, decisions about what patients/carers can contribute should be co-produced with this stakeholder group's representatives. Efforts should be made to convey technical content in accessible ways, avoid jargon, and consider needed training needs for effective engagement. The role of patient, carer, and public-voice contributors must be clear to participants from the start, e.g. their role, commitment, and compensation, as well as how they can provide the most beneficial input (see Annex C).

5. Expertise from research and innovation funders in the public and charity sectors

Why it matters: Charities funding research and innovation activity and public sector research and innovation funders are relevant to consult, offering perspectives useful for informing unmet needs. Our research suggests that funding organisations can have a broad understanding of diverse aspects of unmet needs, whether related to a need for improved technical performance from diagnostic tests or to broader health system issues such as the need for improved diagnosis in primary care settings, accessibility, invasiveness and cultural acceptability issues.

Which types of expertise might be relevant: Expertise from charities funding research on specific types of cancers as well as those with a broader remit (e.g. CRUK, Macmillan Cancer Support) and public-sector funders active in cancer diagnosis-related research (e.g. from The Department of Health and Social Care [DHSC], National Institute for Health Research [NIHR] and NHS England research and innovation funding programmes) can be helpful to consider. Our work did not consider private equity investors as they are more likely to engage in funding companies responding to a TPP than contribute to the TPP development process itself.

6. Broader decision-maker expertise: Regulators, HTA and policymaker perspectives

Why it matters: A broader set of decision-makers play a role in the innovation pathway for novel diagnostic tests, e.g. by providing information about the regulatory requirements new tests must meet, conducting HTA cost-effectiveness assessments that can influence whether tests are paid for by the NHS, and informing priority areas for innovation via policy-related decision making. Those with a role in the purchasing, adoption, and/or uptake of diagnostic tests also matter, as their perspectives can help shed light on the improvements needed in existing tests for which there is a viable value proposition and broader cost considerations. Policymakers and HTA agencies are likely to have some insight into purchasing realities. While it is outside the scope of a TPP to provide detailed information on issues of regulation, HTA criteria, policy priorities or reimbursement pathways, our project insights strongly support the opportunities demand-signalling TPPs present to signpost to sources of such information as added value for innovators. Therefore, it is vital to bring regulatory and HTA considerations into the TPP development process to help innovators consider these needs early in the product-development journey (including potentially ruling out specifications that would lead to cost-ineffective tests using early economic modelling) rather than failing on regulatory or HTA grounds. Regulators and/or regulatory experts can inform health economists about requirements that can feed into early economic modelling to inform TPP specifications. TPPs can also signpost key organisations or information sources that can provide further insights on regulatory and HTA criteria or routes to access, e.g. navigating the NHS adoption and uptake landscape. Examples of such organisations and programmes include The Accelerated Access Collaborative at NHS Innovation Services, which provides a guide to innovation in the NHS, or the non-profit Medilink UK, whose remit is to link SMEs to NHS adoption pathways.

Which types of expertise might be relevant: Regulatory, HTA, and policymaker perspectives can contribute to the information innovators can access via TPPs via signposting or key insights. Representatives of these organisations can be consulted directly in the TPP development process to identify essential information sources and organisations/offices. However, there may also be other individuals in a core working group leading TPP development who are well-versed in and well-sighted on regulatory, HTA, and policymaker considerations (e.g., influential boundary spanners). In a UK context, relevant expertise from an HTA perspective may span NICE Early Value Assessment and Diagnostics Assessment programme experts for England, Wales and Northern Ireland, or experts from the Scottish Health Technologies Group as another example, as well as Health Technology Wales expertise. From a regulatory perspective, the MHRA Innovation Office and Interim Devices Working Group also offer support for innovators in the UK, sometimes with additional engagement from agencies in the devolved nations. Other sources of relevant expertise could come from sources such as the NHS Innovation Service and experts within AHSNs. In addition, engaging the payer perspective is also relevant to understanding and reflecting purchaser realities (e.g. MedTech funding mandate, NHS Supply Chain, regional procurement networks, and trust-level purchasing decision-makers).