Supplementary Table 1 Challenges at each guideline development stage are often interconnected to the domains of methodological issues, resource limitations and awareness/education needs

Guideline Development Stage	Summary of Key Message(s)	Domain(s) impacted illustrated by a representative quotation (ID*)		
		Methodological Challenges	**Resource Limitations**	**Awareness and Educational Needs**
Scoping	Scoping in diagnostic guideline is more extensive and hence resource heavy compared to intervention guidelines	" It takes a fair chunk of an analyst's time over several months to go through scoping, yes. It's absolutely critical that the problem is well defined ...to understand exactly what all the ins and outs of the problem are and it's not something you just throw together. It's one of the things that makes diagnostics different. It requires a vastly, a more complicated problem definition phase that you would typically have for treatment" (ID 1)		
Key Question(s) Formulation	PICO format for question formulation is not very useful for diagnostics. Additionally, the panel needs to be educated and trained on how to develop focused questions that include patient outcomes as often time and money is limited	"PICO is not very relevant for diagnostic questions but still used to try to be consistent with intervention format" (ID 5) "You cannot really apply PICO" (ID 15) "Developing comparative test accuracy questions is complex, its sometimes not just a straightforward diagnostic test accuracy question but involves more the diagnostic pathway ..so that's quite complex" ID 2	"There is a need to be more specific in the guideline question as time and money is often short" (ID 3) "One of the major challenges is actually limiting the PICO to a question that you can handle because we only have about sort of three weeks to do the review once the search has been done so you can imagine a bit of speed work in that" (ID 13)	"The panel still need help with question formulation. There is a lack of appreciation of what's required in a question. They found it quite hard to give us the details of what it actually means to do a PICO because a lot of them don't know that, you know. You can call it education of the clinicians" (ID 13) "Educating the panel on test's downstream

				consequences helps define exact questions to be answered in guideline" (ID5)
Guideline Development Stage	**Summary of Key Message(s)**	**Domain(s) impacted illustrated by a representative quotation (ID*)**		
		Methodological Challenges	**Resource Limitations**	**Awareness and Educational Needs**
Developing a test-treatment pathway	Developing a test-treatment pathway is useful in helping panelists develop focused questions that are patient outcome centered but the awareness of the panel on the importance of developing a test-treatment pathway needs to be raised. There are also methodological and resource issues to be considered especially if the topic is broad and/or a lot of variation in practice exists. At the moment, there seems to be no systematic approach used	"Defining a clinical pathway helps to define diagnostic questions to be addressed. It helps clinicians make the link between test accuracy and clinical outcomes. That's usually when it kind of ticks off in their brain and helps them see the bigger picture." (ID5) "Yes, so it's difficult, of course there are many pathways and then at the beginning to define just one, they first have to have the data to really make clear which way is the best way." (ID 10) "Mostly it's a panel discussion or even just an opinion of the experts which I know is not very evidence-based but that is the	"Looking at a test in its context is useful, however some guidelines such as cancer ones are really broad and deal with the entire p/w of dx to staging to treatment and FU and for that reason including a pathway is not doable due to resource constraints" (ID 7)	"Clinicians see the diagnosis question as being quite straightforward (so) it can be quite difficult trying to get across to a clinician and trying to get them to think about the pathway" (ID 5) "Consideration of the clinical pathway and how tests fit in that pathway, that should have a more central role in the whole process " (ID 9)

		present system that we use" (ID 9) "Variation in practice can make it difficult to get consensus. It's sometimes amazing how much variation there is and how people don't necessarily agree on each different pathway that is sort of proposed" (ID 13)		
Guideline Development Stage	**Summary of Key Message(s)**	**Domain(s) impacted illustrated by a representative quotation (ID*)**		
		Methodological Challenges	**Resource Limitations**	**Awareness and Educational Needs**
Types of outcomes and types of evidence included	Resource is a major consideration as to whether the panel includes outcomes other than test accuracy (Issues relating to patient outcomes and test accuracy evidence is covered under Tables 3 and 4)		"We do not always look for information on costs because there is not always an economist involved. I know that NICE for example is doing it always but we do not have the resources to do it" (ID 7) "Define the budget, get another team to bring the resource and the effectiveness data, the economic data to the table. I don't know any other way to do it, one person can't do this, you know, and I can't just extract diagnostic test	

			accuracy studies. It's not a complete enough picture yet, it's not the way to do it (ID 11)	
Guideline Development Stage	**Summary of Key Message(s)**	**Domain(s) impacted illustrated by a representative quotation (ID*)**		
		Methodological Challenges	**Resource Limitations**	**Awareness and Educational Needs**
Searching for test accuracy evidence	Searches are time and labor consuming because of the lack of appropriate search filters	"No good search filters for diagnostic studies. The existing search filters have been validated, and these validations show that they don't work making the process time and labor intensive" (ID 15) "A search can come up with 5000-6000 hits. It's usually too much to go through in a short period of time " (ID 12)		
Synthesizing the Evidence	The main challenges in pooling results from test accuracy studies is the poor study quality and heterogeneity of the studies. Guideline developers also lack support methodologically as methods tend to more complex, as well as the time needed to prepare such reviews	"There's not much secondary synthesis you can do if you don't have good information from individual studies" (ID 2) "Yes, we have a lot of them that show very heterogeneous data. That indicates there are a lot of things still to be done and that we should be very cautious of single studies and drawing conclusions" (ID 10) "I'm not sure if the estimates can	"We don't have enough outside resources to have a fully-fledged systematic review done properly" (ID 9)	"We do not do meta-analysis because we are not as familiar on how to do this compared to treatment which is much more common in guidelines and also because the evidence is usually very thin so its better to just stick to narrative summaries " (ID 3)

		be derived as easily as estimates for treatment effects. I see uncertainty even in those meta analysis where such estimates have been reported. I think the very differing populations where diagnostic studies have been performed, make some additional difficulty in this" (ID 4)		
Guideline Development Stage	**Summary of Key Message(s)**	**Domain(s) impacted illustrated by a representative quotation (ID*)**		
		Methodological Challenges	**Resource Limitations**	**Awareness and Educational Needs**
Moving from evidence to making recommendations	Expert opinion is important especially in the face of lack of good quality evidence. However, this makes the process unstructured, not transparent and political.	"There's a discussion about the benefits and harms, about resources and about patient values and preferences. Do we know those? No. Again, people give you their opinions about it, but that's the best we can do at this point" (ID 11) "We're so often limited by the quality of the evidence available, and so we are taking the clinical expertise of a group of highly expert individuals as the next best thing" (ID 6)		

		"Now all systematic work is focused on the hard evidence and the rest is “Oh yes, let’s do some focused interviews or let’s ask the working group" It’s not very systematic and we don’t have the tools or a format for that process” (ID 3) “Indirectness of current evidence and the lack of direct evidence on patient outcomes are reasons why making recommendations for is not transparent” (ID 7) “It’s political and then lack of evidence that’s not there. You have to build on the expert opinion and that can be quite difficult. Some people feel that their way of working is say put aside and it’s not reflected in the recommendation” (ID 10)		

Guideline Development Stage	Summary of Key Message(s)	Domain(s) impacted illustrated by a representative quotation (ID*)		
		Methodological Challenges	**Resource Limitations**	**Awareness and Educational Needs**
	Other structured approaches such as Delphi processes, focus groups or the use of modeling could make the process more systematic and transparent. Some view modeling as an alternative way to making recommendations. Others feel modeling is the *only* way evidence based recommendations can be made in the face of the lack of end to end studies. One interviewee felt end to end studies is the only way forward and modeling is too misleading.	"Other methods should be explored such as Delphi or focus groups for gathering the information and making recommendations" (ID 3) "We use modelling very successfully but people are now starting to do research that gives you information from the field that is well designed and could also contribute. It's not either/or. It's modelling's great too." (ID 11) "Modelling is the only way where we're able to provide that linkage if it doesn't exist in studies. Reality is we're probably not going to see a time for most diagnostics in the near future where we can get end to end results in a reliable timely manner. So, we're going to have		

	Whatever the view, modeling or any other methods require resources	to do something else. Modelling is the only other answer I know of. Is it probably more accurate most of the time than people just making individual guesses in clinical practice? I would say yes" (ID 1) “Models need assumptions and the assumptions cannot be proved, so it’s very uncertain. Especially in the area of genetic and molecular biology, where our understanding of the natural disease is changing, then modeling is too uncertain. We don’t want to accept this uncertainty. We think it’s better to give some pressure on the community to perform such studies” (ID 14)	“The biggest problem with modelling expertise is less the training cost than just finding the people with that skill" (ID 1) "I think it’s (modeling) really valuable information but it’s a lot of work." (ID 7)	

*ID = unique identifying number for interviewee

Supplementary Table 2 Guideline developers recognise the need for patient important outcomes

Quote	ID*
“What are the implications of this further downstream? So that’s another challenge that we try to tackle. To try to get them to think a bit further and to also think about what we call clinical utility of the tests, rather than just purely accuracy"	ID 2
“PICO is based on the fact that they just want a good instrument and attention is not paid to other outcomes at the start of the process. That question should be more in the beginning and less about the accuracy of the instrument , because my experience with the working groups is that they are also a bit disappointed that there is too little time left to answer the question what it (the test) means for me as a clinician and what it (the test) means for the patient?”	ID 3
“Linking accuracy to patient outcomes would be useful and should be done more often”	ID 5
“Does it (the test) really help me?” Let’s say the patient, the doctor and then the outcome at the end is becoming more important. Let’s say the link between imaging and outcomes, that’s something that I think is very important. I think there should be more attention for that. For me that would be the most important”	ID 10
“There’s a gap about the downstream consequences on patient health. The piece that’s missing that would be helpful for these meetings would be information on patient important outcomes”	ID 11
"When you administer the test, it might not be for you that important how many false-negatives there are but for the patient it’s quite a different story. So I think steps carrying on from the two by two tango, getting away from that to actually a meaningful translation that could be better, better sort of trained"	ID 12

*ID = unique identifying number for interviewee

Supplementary Table 3 The lack of evidence on patient outcomes is a major challenge guideline developers face

Quote	ID*
“Test accuracy studies available do not measure patient outcomes hence there is no evidence available. The panel is interested to know what it (the test) means to the patient but they are not sure how to address it in the face of lack of evidence”	**3**
"We know that sensitivity, specificity is not enough. We must have an idea of the impact, (but) the studies that measure directly the impact on outcome is very rare “	**15**
"We try to use the outcomes that are also used for therapeutic interventions, so mortality and things like that. It’s hard to find in the literature”	**7**
“Outcome based studies are rather scarce”	**9**
“ The combination of medical testing and the corresponding consequence is rarely done in literature. The awareness just isn’t there”	**14**

*ID = unique identifying number for interviewee

Supplementary Table 4 Including patient outcome data in medical testing is challenging due to issues that are inter connected across the domains of methodological, resource limitations and awareness/education

Domain/Summary of the main issues	Quote (ID*)
Methodological & Resource Challenges: Direct evidence on testing and patient outcome is rare due to methodological challenges Guideline panels should use different methods for looking for patient outcome data but this requires resources	"These studies are usually very long and many of these companies cannot survive for that long waiting for the results" ID 1 "Technological obsolescence" where test technology evolves much more rapidly than for drugs hence studies and tests being studied can become easily obsolete " ID 1 "For questions which the panel feel there will be very little or no evidence, other methods should be explored such as Delphi or focus groups for gathering the information and making recommendations but currently there is no systematic way of doing this, or formats to use" ID 3 "If you want to be comprehensive and look at these patient outcomes you need to do it as a separate project by a different team, because the methodology might differ" ID 11 "Qualitative research is usually not included or available, but may shed more light on the harms…Doing qualitative research as part of a guideline would be really useful, but is not possible because of time constraints" ID 15
Awareness & Resource Challenges: Lack of awareness within the panel, and amongst other stakeholders results in difficulties getting funding to conduct end to end studies	"One of the difficulties, of course is the cost of the study and the profits of some of these diagnostic tests are not, at least we're always told by industry, are not in the range for what they get for pharmaceuticals and many of these c companies tend to be small companies" ID 1 "We couldn't find the funding for looking to an outcome measured based analysis of two imaging strategies. So it's not always accepted as the way to evaluate it. It's so natural that the

	drug industry has to finance these studies, but in the medical test industry it's not. Take for example imaging and the very very big companies, Lehman's or GE or Phillips which make a billion per year, they don't finance corresponding studies that allow them to sell their product" ID 14 "Industry's influence is also a factor that inhibits end to end studies being conducted. We had to have strong discussions with the government to exert pressure on the need to have studies which link CT testing to patient outcomes" ID 14 "We couldn't find the funding for an outcome measured based analysis of two imaging strategies. So it's not always accepted as the way to evaluate it. Yes partly it's funders who should change. Of course there are grants given but for imaging it's often the accuracy that's found interesting and outcome.., yes, that's more difficult because the link is not so clear always" ID 10
Domain/Summary of the main issues	**Quote (ID*)**
A general lack of awareness on the need for direct evidence on testing and patient outcomes	"More primary studies assessing a test's impact on patient outcomes would be helpful in this regard although these types of studies may not always be feasible for reasons of lack of awareness on the need to assess a test's impact on patient outcomes" ID 7

*ID = unique identifying number for interviewee

Supplementary Table 5: Main differences between medical tests and intervention guideline development

Category	Summary	Quote (ID*)
Diagnostic test accuracy (DTA) studies	The conduct of medical test studies is considered more complex than intervention research	"Quite often we don't necessarily have a gold standard reference, the reference is all quite often woolly... and in intervention we use one against another intervention, you can kind of ... it's a lot more structured" (ID 12)
	Overall quality of DTA studies tend to be poor usually	"When you come to looking through the actual evidence, there's just poor quality with the majority, a lot of them are still very poor quality; there's not much secondary synthesis you can do if you don't have the good information from individual studies" (ID 2)
Funding of DTA studies	Funding for test accuracy studies that report on patient outcomes are usually harder to obtain	"Firstly manufacturers of tests are usually small scale companies with less financial resources compared to pharma hence are unable to fund such studies. The cost of the study and the profits of some of these diagnostic tests are not, at least we're always told by industry, are not in the range for what they get for pharmaceuticals and many of these companies tend to be small companies that develop these things" (ID 1) "We couldn't find the funding for really looking to a really outcome measured based analysis of two imaging strategies. So it's not always accepted as the way to evaluate it. It's so natural that the drug industry has to finance these studies, but in the medical test industry it's not" (ID 14)

Methods for the synthesis of DTA studies	Methods for synthesis of evidence from primary studies is more complex and guidance for this is less explicit	"Diagnostic guidance is less explicit on when studies are too heterogeneous to pool compared to intervention where there are explicit heterogeneity cut offs "(ID 5) "We do not do meta-analysis of diagnostic studies because we are not as familiar on how to do this compared to treatments which are much more common in guidelines and also because the evidence is usually very thin so it's better to just stick to narrative summaries" (ID 3)
Category	**Summary**	**Quote (ID*)**
Guideline Development	Guideline development in medical tests is more complex and methodological guidance is less explicit	"We can do a very extensive scoping operation in order to understand exactly what all the ins and outs of the problem are and it's not something you just throw together. It's one of the things that makes diagnostics different. It requires a vastly, a more complicated problem definition phase that you would typically have for treatment" (ID 1) "There are no good search filters around which make searches labor and time intensive. The PICO is not directly applicable compared to interventions and cut off points can be difficult to determine and usually depend on expert opinion which can vary from GDG to GDG" (ID 15) "There's less clear guidance than there is with for intervention reviews (when it comes to making recommendations)" (ID 3)
Familiarity and understanding of medical test	Interventions research seems to be of higher priority in the research community and gets more research	"Most attention goes to intervention studies in journals and in guidelines normally and generally in the education of medical professionals there is less focus on dx accuracy so we're not as

diagnostics	attention (jn terms of publications, funding)	used to it" (ID 3) "Clinicians are a bit more used to big, mega, pharmaceutical mega-trials and how to interpret that information" (ID 9) "I think with RCTs it is that they tend to be very familiar to them, maybe because it's just sort of been what they've been exposed to the most I think. There's this perception that it (RCT) is the best design, so sometimes they might even ask for RCTs for questions where it's not actually appropriate to do an RCT. You know, I think RCTs have gotten people used to the idea that you can get a bottom line and that's of course not the case" (ID 13)
Category	**Summary**	**Quote (ID*)**
	Medical test statistics and link to outcomes is harder for clinicians to understand than for interventions	"I think that the statistics behind evaluating a therapeutic intervention are more easily understood by experts. Experts do not find it intuitive to associate a test with impact on a downstream patient outcome unlike with interventions" (ID 7) "Oh this test is a lot better than the other" but what does that mean in real terms? I think that's something that is difficult for us as well as the clinician to appreciate I think" (ID 12) "Clinicians seem to focus quite a lot on the interventions. They understand intervention type studies more easily, and struggle with sensitivity, specificity , multiple tests, tests affecting management strategies and care pathways" (ID 5)
Regulatory controls	Regulatory control for medical tests is less stringent	"Combination of medical testing and the corresponding

	than for drugs	consequence is rarely done in literature. There are no regulations from an overseeing board at least not in Europe. If we change the regulatory process to be similar to drugs, then I feel that we will in a very few years we'll have much better studies" (ID 14) "Evidence is a problem throughout, because we don't have the regulatory requirements that you have for pharmaceuticals when you're dealing with diagnostics for the most part. Even to the extent that you have some regulatory requirements but they're usually quite weak and don't tell you much" (ID 1)

*ID = unique identifying number for interviewee