European Medicines Agency

London, 3 March 2009
Doc. Ref. EMA/182470/2009

OVERVIEW OF COMMENTS RECEIVED ON
DRAFT GUIDELINE ON CLINICAL EVALUATION OF DIAGNOSTIC AGENTS – REV 1

Table 1: Organisations that commented on the draft Guideline as released for consultation

Add name followed by link to individual received comment (upon publication by Web Services)

	Name of Organisation or individual	Country
1	Industry Task Force and AIPES	
2	EORTC	
3	GE Healthcare LTD	
4	International Society for Clinical Biostatistics (ISCB)	
5	MSD	
6	Novartis Pharma	
7	Jorgen Hilden (Dept of Biostatistics, University of Copenhagen)	
8	Anabel Cortes-Blanco	

7 Westferry Circus, Canary Wharf, London, E14 4HB, UK
Tel. (44-20) 74 18 84 00 Fax (44-20) 74 18 86 13
E-mail: mail@ema.europa.eu http://www.ema.europa.eu

Table 2:Discussion of comments

GENERAL COMMENTS - OVERVIEW
General Comments (1) Firstly, the organization of the document is difficult to follow. Since the target indication drives the development program, the document should be organized along the lines of the target indications, and under each section to outline potential claims and the requirements for each. For example, a diagnostic test may be used to: a) exclude and detect disease or characterize tissues; b) assess and rank severity of disease; c) assess prognosis; d) monitor response to treatment; e) structure delineation; f) functional, physiological, or biochemical assessment. What the test is intended for should drive the assessment. Outcome: Not accepted; interesting proposal, but would complicate the document even more. Secondly, the main focus of the document is on confirmatory trials, so that limited or no guidance is given to earlier phases of clinical development. Thirdly, radiopharmaceuticals are different from contrast agents. Appendix 1 to the Guideline would be more helpful if guidance on radiopharmaceuticals would be separated from guidance on contrast agents. Outcome: Not accepted. Since the requirements for registration are the same, it was decided that a specific appendix on radiopharmaceuticals was not necessary. **Major concerns (1)** The demonstration of clinical benefit should be tailored to the diagnostic test being developed and its potential claims. Unless a patient management indication is sought, to always require a positive impact on diagnostic thinking and/or a positive impact on patient management is inappropriate and places an undue burden on diagnostic tests, for which the demonstration of appropriate diagnostic performance should be enough to obtain approval. Outcome: Not accepted. Could the company give an example of a diagnostic agent for which an impact on patient management does not hold? To avoid bias, the demonstration of efficacy of diagnostic agents requires an artificial setting (inclusion and exclusion criteria, fully blinded off-site reading, systematic adoption of a truth standard, etc.). The artificial setting significantly affects any possible demonstration of impact on patient care. Besides, in most cases, all the investigator can provide about implications on patient care is informed speculation, and this should not be any part of the basis for a regulatory approval. It is recommended that proper demonstration of diagnostic performance constitute the primary, if not the sole, basis for the regulatory approval of new diagnostic agents. The Industry Task Force recommends that no other component of clinical usefulness be a requirement for authorisation of a new agent, and that, when they are included in a dossier, they may be established indirectly or historically. Moreover, proper demonstration of adequate technical performance should be sufficient to obtain marketing authorization for the following situations:

1) agents used to locate and/or outline normal anatomic structures (or variants of normal anatomic structures (for example, an agent that non-specifically enhances the airway lumen to distinguish dilated bronchi from normal bronchi and categorizes the bronchiectasis anatomically, or an oral agent used to delineate the stomach or the bowel loops with CT or MRI, or a contrast agent developed to image the normal parathyroid glands); in such situations, efficacy is immediately obvious from the images themselves; Outcome: this is not correct, in addition of delineating normal structures, the product has to be safe and the information provided meaningful for patient management. 2) agents used in conjunction with an imaging modality that is considered to be the closest approximation to the truth, no real standard of truth is available and follow-up information cannot be used to validate the results of the diagnostic test (for instance, an agent used to detect, define the extent and/or assess brain or spine lesions with MRI, PET or optical imaging, or agents used to detect whole body metastases with PET, or agents intended for the assessment of regional cerebral blood flow with MRI, or agents used for imaging of inflammation); 3) new agents compared to validated comparators, i.e., to agents already approved and widely used to improve the technical performance of imaging modalities. Industry recommends that the requirements for authorisation are limited to the demonstration of an equivalent or superior technical performance and/or safety profile of the new agent compared to the approved comparator. Outcome : Not accepted. Technical and diagnostic performance are always required, demonstration on impact on diagnostic thinking is necessary in most cases.
(2) I am not sure whether the title is correct, the way it is written. There is not much help for an expert in this field, it is more an introductory article to the subject of developing and licensing diagnostic agents. Also, the examples given are rather poor regarding description, field of application, intended use, etc. The article is superficial and not much help for someone developing kits. Also, there is no direct link to clinical applications, and if mentioned, is very brief. All in all, some paragraphs read well, but the article is not more than considerations (or reflections). I would have loved to see an article which is more to the point, e.g. giving discrete help for blood testing, tissue testing, preparation of assays, and the limitations of the test. Outcome: Not accepted (out of scope). Development of kits is out of the scope of this guideline.
(2) I miss the external assessment by independent organisms. No comment is made if the institution or company, specifically concerning the chapter "Methodological considerations when developing and licensing diagnostic agents" (starting on page 6), is audited with respect to GCP (Good Clinical Practice) and in this case specifically GMP- criteria (Good Manufacturing Practice). Also, the centres integrating this new type of agent should also be audited if the integration of patients in their trials is according to GCP-criteria. We are in an era of QA (Quality Assurance) and allowing an external assessment is imperative in any new development. Outcome: Accepted but out of scope for a guideline an clinical development of diagnostic agents
(2) I am surprised that the group of children is not mentioned at any part of the program. Only once has “age” been taken into account (section 3.2). The relevance of a diagnostic procedure or drug for children has to be an independent part of the development program of a new diagnostic test and be addressed separately. In the meantime, this is standard in nearly all drug related guidelines (anticancer, HIV, asthma, anticonvulsants, immuno-globulines…..) and should be handled here in the same way. Outcome: Paediatric addendum is not planned for the moment
(3) A revision of the guidance on Clinical Evaluation of Diagnostic Agents is welcomed; however the organization of the document is difficult to follow. The target indication drives the development program, so it may be more useful to industry to organize the document along the lines of the target indications, and under each section to outline potential claims and the requirements for each. Outcome: Proposal is of interest, but will complicate the document further; current structure will be kept

(3) We view the draft document as requiring an increased level of supporting clinical data in terms of demonstration of the impact on patient management and clinical outcome. The level of supporting clinical data required for marketing approval should be related to the claimed indication. If an imaging agent is not claiming a prognostic indication, then technical/diagnostic performance data from clinical studies should be sufficient for approval. In line with this, the possible indication of Prognostic/therapeutic management guidance' should be reinstated. Outcome: the minimal requirement for registration is an adequate technical and diagnostic performance of a new diagnostic agent in relation to a standard of truth and, when appropriate, to an established comparator in the clinical context in which the diagnostic agent is to be used in well-designed superiority or non-inferiority trials. Demonstration of the impact on patient management and clinical outcome is rarely, if ever, necessary (except if there is a specific claim).
(3) The Appendix would be more helpful if the guidance on radiopharmaceuticals was separated into its own section within the appendix, as radiopharmaceuticals are very different from contrast media in their properties and use. Outcome: Not accepted. Since the requirements for registration are the same, it was decided that a specific appendix on radiopharmaceuticals was not necessary.
(6) This revision to the earlier (2001) EMEA Points to Consider on Evaluation of Diagnostic Agents is welcome as it incorporates some new developments in the field and some new thinkings. For example several concepts are newly added compared to the previous Points to Consider. Although this draft guideline concerns diagnostic agents intended for administration to patients (eg contrast agents, radionuclides), care should be taken to ensure that the principles outlined in this document are aligned with those principles (eg clinical trial design, PPV, NPV, analyses, clinical impact) that may also be relevant for diagnostic agents not intended for administration to patients (eg pharmacogenetic test kits). Furthermore, although the EMEA would not review ex-vivo diagnostic kits for approval as devices, it may increasingly review applications that contain companion ex-vivo diagnostic kits as part of the MAA for a drug. As EMEA is also increasingly likely to give advice on biomarker qualification it is important that principles that are common to in vivo diagnostic agents and in vitro diagnostic agents are aligned. Consultation with the EMEA's Pharmacogenetics Working Party is advised. However, we have some concerns about the new terms such as 'impact on diagnostic thinking', the expectation of a "benefit-risk assessment" for a diagnostic (which will provide no benefit by itself), some internal inconsistencies, and the absence of a distinction between 'detection' claims and 'clinical utility' claims (the former would leave the treatment decision to the physician, the latter would instruct the physician on the treatment decision). Some of the topics could benefit from clearer explanation. Outcome: Not accepted. Diagnostic agents not intended for administration to patients (egp pharmacogenetic test kits) are out of the scope of this guideline. "Impact on diagnostic thinking' is not a new term, it was already well described in the previous guideline. Benefit-risk assessment is new, and stresses the fact that a diagnostic has to be both as safe as possible and has to provide benefit for patient (related to the correct diagnosis).

(7) The term ‘prevalence’ is used in an extended sense about the fraction diseased in a stream of admissions, say. Epidemiologists prefer to reserve the word for the fraction diseased in a snapshot (cross-section) of citizens, such as perimenopausal Irish women. The latter concept is appropriate for health screening, the former extension for clinical case streams. In my own writings I have experienced the difficulty of trying to avoid the extended use. But many facts stated in epidemiology texts, such as those about prevalence (in the strict sense) being equal to incidence rate times disease duration, become meaningless if the reader naively thinks that prevalence in the extended sense is also governed by this equality. Therefore I believe we should acknowledge the epidemiologists’ laboriously developed, clean terminology. No particular comment.
(8) It would may be more appropriate to refer to “diagnostic medicinal products” instead of “diagnostic agents” in the text of the guideline and the title taking into account that these “agents” are “medicinal products” and not any other different entity. Outcome: It is stated in the Introduction that diagnostic agents are medicinal products used for diagnosis or monitoring of a disease and that the evaluation of diagnostic agents is governed by the same regulatory rules and principles as for other medicinal products. The term diagnostic agents is used fro simplicity.

SPECIFIC COMMENTS ON TEXT

TITLE

Line no. + paragraph no.	Comment and Rationale	Outcome
Title (8)	It is more specific to say that they are intended for human use. Proposed change: “Guideline on the evaluation of diagnostic agents for human use”	Not accepted. All EWP guidelines are intended for human use.

EXECUTIVE SUMMARY

Line no. + paragraph no.	Comment and Rationale	Outcome
Line 2 (8)	It is more specific to say that they are intended for human use, and that they are medicinal products instead of agents. Proposed change: “…in development of diagnostic medicinal products for human use…”	Not accepted. See introduction and answer to General comment 8.
Line 4 (8)	It is more appropriate to specify that outcome refers indeed "clinical outcome", which is one of the items to measure clinical usefulness. Proposed change: “….impact on patient management/clinical	Not accepted. Clinical outcome is not clinical usefulness of a diagnostic agent.

SPECIFIC COMMENTS ON TEXT		
	outcome)…"	
Line 5 (8)	It is more appropriate to say "population" or "trial population" instead of "patient selection" since the participants in clinical trials may be not only patients but also healthy volunteers. Proposed change: " population,…" It is also important to include the mention to the comparator in these items.. Proposed change: "…., standard of truth, comparator, strategy,…"	Accepted.
Line 7 (8)	It is more complete to say that a chapter has been added on the requirements for registration of diagnostic medicinal products that not only of products. Proposed change: "…for registration of diagnostic medicinal products…"	Not accepted. See introduction and answer to General comment 8.
Add more rows as appropriate		
EXAMPLE:		
1 INTRODUCTION		
Line no. + para no.	**Comment and Rationale**	**Outcome**
(8)	This information is missing: "This guideline is restricted to those medicinal products for diagnostic use administered into or onto the human body."	Accepted. See Scope, 1st sentence.
Line 1 (8)	The guideline aims ONLY at the clinical evaluation, not to the development, of diagnostic medicinal products Proposed change: "This note provides guidance for the clinical evaluation…"	Not accepted. Guideline is aimed to help both development and assessment of diagnostic agents.
Line 20 (8)	It is important to specify that this guideline refers ONLY to those radiopharmaceuticals intended for diagnostic use and not every radiopharmaceutical since some of them are intended for therapeutic	Accepted.

SPECIFIC COMMENTS ON TEXT		
	use. Proposed change: "Radiopharmaceuticals as defined in European Directive 2001/83/EC, as amended, intended for diagnostic use."	
Line 24 (8)	A parenthesis was missing breath test with urea (^{13}C)	Accepted.
Section 1 (6)	Should also refer to diagnostic and device regulations & guidelines Please explain if the principles outlined in this document are also relevant for the qualification of biomarkers to be used in diagnostic tests. Such biomarker data may be submitted for qualification via the EMEA's new qualification procedure (EMEA/CHMP/SAWP/72894/2008).	Not accepted. Qualification of biomarkers is out of scope in this guideline.
Throughout (6)	It is unclear how 'benefit' can be assigned to a diagnostic agent that does not deliver a therapeutic action by itself. This could be better clarified & discussed in the guideline. Please better describe what are the expectations for 'benefit' of a diagnostic agent. For example: in terms of sensitivity, specificity, earlier detection etc; reduced risks of false postives and false negatives.	Accepted. Correct diagnosis is clearly a benefit for a patient: wrong diagnosis is detrimental for a patient; in addition there are safety concerns for some diagnostic agents.
2 FUNDAMENTALS IN THE CLINICAL EVALAUTION OF DIAGNOSTIC AGENTS		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Line 15 (8)	It is not necessary to mention about procedural convenience when referring to technical performance since procedural convenience is just one item of technical performance and it is already detailed in section 2.1.1 Proposed change: "…technical performance ~~(including procedural convenience),…~~"	Accepted.
Line 16	Linguistic change: "…impact on diagnostic thinking, on patient	Accepted.

SPECIFIC COMMENTS ON TEXT		
	management and clinical outcome,…"	
Section 2, Lines 61-64 (1)	Lines 61-64: "*In order to establish an indication for a diagnostic agent, it is necessary to demonstrate its benefit by assessing its technical performance (including procedural convenience), diagnostic performance, impact on diagnostic thinking, patient management, and clinical outcome, as well as its safety*." It is almost impossible to establish for any new agent all of the components of clinical benefit indicated at Section 2, i.e. technical performance, procedural convenience, diagnostic performance, impact on diagnostic thinking, patient management, and clinical outcome, especially if clinical benefit has to be established directly in a development setting. Proposal: "The demonstration of clinical benefit should be tailored to the diagnostic test being developed and its potential claims."	Accepted.
2.1 Assessment of diagnostic agents (line 1) (8)	It could be more appropriate to join title 2.1. and this sentence below to result in 2.1. "Assessment of the benefit of diagnostic agents" and to delete the sentence "ASSESSMENT OF BENEFIT" Proposed change: "**2.1 Assessment of the benefit of diagnostic agents** ~~ASSESSMENT OF BENEFIT:"~~	Accepted.
2.1 Assessment of Benefit 2.1.1. and 2.1.2 (3)	The text already says that 'Technical performance alone is necessary but not sufficient to show clinical benefit of a diagnostic agent and cannot be the only basis for registration.' Perhaps the sections 2.1.1. and 2.1.2. may be combined together. In addition, showing diagnostic performance implied technical performance, so there should be no separate need to show technical performance. Proposal: Consider combining sections into Technical and Diagnostic performance	Not accepted. Technical performance and diagnostic performance are two different notions; good technical performance may in a way compensate lower diagnostic performance.

SPECIFIC COMMENTS ON TEXT		
2.1.1. Technical performance (8)	This subsection should mention about precision or reproducibility as part of technical performance. In the current guideline there is mention to Precision as consisting of observer's concordance (if subjective outcome) or reproducibility test-retest (if objective outcome), but this has disappeared from the section regarding assessment of diagnostic agents in the current draft of the guideline. Proposed change: "Technical performance should include evaluation of precision of the diagnostic medicinal product/test which consists of observer's concordance (if subjective outcome) or reproducibility test-retest (if objective outcome)."	Accepted.
2.1.1. Technical performance (line 1) (8)	This change tries to keep the same style as the items below: "It consists of image quality and/or procedural advantages/disadvantages of and investigational diagnostic medicinal product/test….."	Accepted.
2.1.1. Technical performance (line 2) (8)	I suggest deleting this content since comparison with a comparator may be a requirement for approval and should then appear in a different section. Proposed change: "~~….and, if applicable, in comparison with a comparator is required."~~	Partly accepted. Will be added for section 6 (requirements for registration) but will not be deleted from section 2.1.1.
2.1.1. Technical performance (line 3) (8)	This sentence may be more interesting to be included in section 6 (requirements for registration). Proposed change: "~~Technical performance alone is not necessary but not sufficient to show clinical benefit of a diagnostic agent and cannot be the only basis for registration."~~	Not accepted. This relates more to the topic of technical performance.
Para 2.1.1, Lines 70-71 (1)	Lines 70-71: *"Technical performance alone is necessary but not sufficient to show clinical benefit of a diagnostic agent and cannot be the only basis for registration."* That should not be true for: a) agents intended for structure delineation (see above); b) agents is used in conjunction of an imaging modality that is considered to be the closest approximation to the truth, no real standard of truth is available	Not accepted. See answers to general comments. Structure delineation by itself is of no interest. It has to deliver information beneficial to the patient.

SPECIFIC COMMENTS ON TEXT		
	and follow-up information cannot be used to validate the results of the diagnostic test; c) new agents compared to validated comparators, i.e., to agents already approved and widely used to improve the technical performance of imaging modalities. For agents developed to improve disease detection or exclusion or tissue characterization, since technical performance is necessary but not sufficient for diagnostic performance, the demonstration of diagnostic performance implies adequate technical performance, so it should not be a separate requirement. Proposal: "*"Technical performance alone is necessary but not always sufficient to show clinical benefit of a diagnostic agent and cannot be the only basis for registration, unless in case of:*" *a) agents intended for structure delineation; b) agents is used in conjunction of an imaging modality that is considered to be the closest approximation to the truth, no real standard of truth is available and follow-up information cannot be used to validate the results of the diagnostic test; c) new agents compared to validated comparators, i.e., to agents already approved and widely used to improve the technical performance of imaging modalities*."	
Technical/Diagnostic performance 2.1.1/2.1.2 (5)	An analysis of sources of variability (analysis, reading, instrument...) as a top level view of sensitivity to controllable variables is recommended.	Current text will be kept.
2.1.2. Diagnostic performance line 1 (3)	The definition of diagnostic performance is too narrow. Sensitivity and specificity are not appropriate as a measure of the performance of all diagnostic tests, especially when used to assess prognosis rather than to reach a diagnosis. Proposal: 'Diagnostic performance of a test may be demonstrated by sensitivity and specificity, negative and positive predictive values, accuracy and precision, or by ROC curves as appropriate to the situation.'	Not accepted. The current definition of diagnostic performance will be kept. Diagnostic performance represents the performance of the diagnostic agent itself and should therefore be independent of the prevalence of the disease in the studied sample of patients.

SPECIFIC COMMENTS ON TEXT		
2.1.2. Diagnostic performance (line 3) (8)	This sentence may be more interesting to be included in section 5.2. as the data to be presented regarding diagnostic performance. Proposed change: "~~The impact of disease prevalence should also be discussed.~~"	Accepted.
2.1.2, Line 73 (1)	Line 73: "*It consists of sensitivity and specificity of a test.*" The definition of diagnostic performance is too narrow. Sensitivity and specificity may not be appropriate as a measure of the performance of all diagnostic tests, such as when assessing prognosis or predicting response to therapy. The use of clinical endpoints different from sensitivity and specificity should be justified in the protocol. Proposal: "*"It usually consists of sensitivity and specificity of a test. However, sensitivity and specificity may not be appropriate as a measure of the performance of all diagnostic tests, such as when assessing prognosis or predicting response to therapy. The use of clinical endpoints different from sensitivity and specificity (e.g., diagnostic accuracy or others) should be justified in the protocol."*	Partly accepted. 1) Nothing in the diagnostic performance definition refers to the assessment of prognosis. 2) NPV, PPV and accuracy are highly dependent on disease prevalence in the studied specimen; they aren't correct to characterise the diagnostic performance of the agent itself. But they are obviously important for the impact on diagnostic thinking. 3) Prognostic value of a diagnostic agent is of course of interest; current guideline does not elaborate on such a claim. The following bullet point has been added in the section 4.2.3 Detection of disease and assessment of its extent and prognosis: • Better defining the prognosis of the disease in a given patient (e.g. standardised uptake value of PET agents may be independent predictors of disease-free survival or overall survival).
2.1.3, Lines 77, 78, 79-80 and 83-85 (1)	It would be useful to have confidence in diagnosis mentioned here, as a secondary endpoint, since that, in real life, is what affects patient management later on. Line 77: "*This refers to the impact of a test result on post-test versus pre-test probability of a correct diagnosis.*" Please, describe how the EWP would envision an applicant estimating the pre-test probability of disease in clinical trials. Line 78: "*..in relation to a well-defined clinical context as regards patient characteristics and prior diagnostic procedures.*"	Partially accepted. Confidence in diagnosis is purely subjective, while changing the patient's diagnosis or his treatment is objective. Of course the adequacy of this change has to be fully checked and assessed. This kind of data can be obtained from the literature.

SPECIFIC COMMENTS ON TEXT		
	What is the meaning of "a well defined clinical context"? This is far too vague to be a useful guidance for developing a clinical research program. If it presumes that a diagnostic test will always be used in a narrow, specific clinical setting that can be defined in advance, this is not a realistic expectation. Lines 79-80: *Therefore, positive and negative predictive values are important parameters which influence the impact on diagnostic thinking in a given patient.* Predictive values are still measures of the diagnostic performance of an agent intended for disease detection, exclusion or characterization. Lines 83-85: *"The impact on diagnostic thinking may be particularly important to show clinical benefit of a diagnostic agent when the sensitivity and specificity and predictive values according to a surrogate standard of truth cannot be defined."* If sensitivity and specificity cannot be defined, also predictive values cannot be defined.	The meaning of the sentence is correctly understood. It aims at obtaining a homogeneous study population. Not accepted. The current definition of diagnostic performance will be kept. Diagnostic performance represents the performance of the diagnostic agent itself and should therefore be independent of the prevalence of the disease in the studied sample of patients. Accepted.
2.1.3 Impact on diagnostic thinking P. 5/16 (7)	The final sentence of the paragraph inverts the idea the sentence is meant to express and should therefore be corrected; p. 13/16, Section 5.2.1, the 2nd paragraph expresses the prevalence concern correctly. However, a warning against a popular misconception deserves to be added. I propose the text on the right →. same time specificity will fall due to mimicry. Proposal: "… depending on the prevalence of the target disease in the overall clinical population of interest [the case stream], a prevalence which does not necessarily coincide with the fraction diseased in the studied case series. (When predictive values are calculated on the basis of a spectrum of prevalences reported from various populations (age groups, countries, occupations, etc.), care should be taken not to take for granted that sensitivity and specificity remain unchanged.)"	Accepted.

SPECIFIC COMMENTS ON TEXT		
Section 2.1.3 (and elsewhere) (6)	'impact on diagnostic thinking' is a rather imprecise description. A clearer term is needed. Change to "Impact on probability of correct diagnosis" or "impact on diagnostic certainty"	These two terms are not interchangeable. Impact on diagnostic thinking is larger than that; it can be a probability of correct diagnosis but also a change from on diagnosis to another.
2.1.4. Impact on patient management (line 1) (8)	For the sentence to be more complete: "This refers to a description and quantification of impact of diagnostic information from the investigational agent on management of a patient and clinical outcome."	Accepted.
2.1.4 Impact on patient management etc line 2 (3)	The suggestion that an "appropriate questionnaire" can be used to assess a test's impact on patient management and clinical outcome is only valid once the test has already been accepted as a valid indicator of disease. While the test is still under investigation, it is not realistic to expect the investigator to assess its impact in this manner, since this could only be done as a speculation rather than based upon actual experience. Proposal: DELETE the second sentence 'It is generally assessed by using an appropriate questionnaire.'	Party accepted. The text will read: 'It may be assessed by using an appropriate questionnaire or by sequential unblinding.'
2.1.4 Impact on patient management etc line 5 (3)	It is unrealistic to expect sponsors to provide information on subsequent patient management for all diagnostic indications. The level of supporting data required should be related to the claimed indication. If an imaging agent is not claiming a prognostic indication, then technical/diagnostic performance data from clinical studies should be sufficient for approval. Proposal: ADD: 'If an imaging agent is not claiming a prognostic indication, then technical/diagnostic performance data from clinical studies should be sufficient for approval.'	Partly accepted. Prognostic value of a diagnostic agent is of course of interest; current guideline does not elaborate on such a claim. The following bullet point has been added in the section 4.2.3 Detection of disease and assessment of its extent and prognosis: • Better defining the prognosis of the disease in a given patient (e.g. standardised uptake value of PET agents may be independent predictors of disease-free survival or overall survival).
Section 2.1.4 (6)	Please elaborate more on the use of an appropriate questionnaire to assess the impact of a diagnostic result on patient management and clinical outcome. How should this be conducted if the diagnostic is new **and** there is	The impact of a diagnostic result on patient management and clinical outcome can be simulated in the questionnaire or by sequential unblinding; patient follow-up data is necessary (should be known)

SPECIFIC COMMENTS ON TEXT		
	little or no prior clinical experience of the consequences of changing patient management?	
2.1.4, Lines 87-88, 79-80 and 89-91 (1)	Lines 87-88: *"This refers to a description and quantification of impact of diagnostic information on management of a patient and clinical outcome."* Impact on patient management (can the new technology influence subsequent therapeutic choices?) and impact on clinical outcome (can the new test affect patient's mortality, morbidity, or health-related quality of life?) are different levels of assessment of the efficacy of a new diagnostic test and should not be mixed together. Line 88: "*It is generally assessed by using an appropriate questionnaire*." The suggestion that the use of an "appropriate questionnaire" can be used to a test's impact on patient management may be only valid once the test is widely used in routine clinical practice. While: a) the test is still under investigation, and b) a standard of truth would be used in all the study patients, it is not realistic to expect the investigator to assess its impact in this manner, since it could only be done as a speculation rather than based on actual experience. The assessment of patient outcomes is too high of a burden to place on a diagnostic test because this implicitly assumes uniformity among physicians in the interpretation of test results, uniformity in the impact of test interpretation on patient management, and consistency of patient outcomes for a given course of patient management; however. There are too many variables (related to different countries, different physicians, different patient conditions, different treatments and their availability) between the diagnostic test and the patient outcome. Lines 88-91: *An assessment of potential benefits and risks arising from the impact on therapeutic decisions should be made. In particular, consequences of an incorrect diagnosis (false positive or false negative) must be considered, e.g. a false positive result that leads to unnecessary interventions.*	Accepted. Accepted. Studies assessing patient outcomes are not mandatory but can be basis for a claim if performed.

SPECIFIC COMMENTS ON TEXT		
	Both the new and the reference diagnostic tests may provide incorrect diagnostic information (in imaging, reference tests include a device technology, e.g., unenhanced ultrasonography, computed tomography or magnetic resonance imaging). In case of an improved diagnostic performance over a reference test, the clinical benefit of a new test should be obvious.	Accepted. Not accepted. Clinical benefit should always be demonstrated/justified.
2.2 (1)	Previous guidance version had a clinical indication of "Prognostic/therapeutic management guidance". It should be reinstated, because this is clinically important to patients and physicians. Proposal: Should be reinstated	Prognostic value of a diagnostic agent is of course of interest; current guideline does not elaborate on such a claim. The following bullet point has been added in the section 4.2.3 Detection of disease and assessment of its extent and prognosis: • Better defining the prognosis of the disease in a given patient (e.g. standardised uptake value of PET agents may be independent predictors of disease-free survival or overall survival).
2.2 Indications/ claims Bullet 3 (3)	Detection and/or assessment of disease and its extent. In addition to extent of disease, severity may also be measurable by a test. Proposal: Detection and/or assessment of disease, its severity or its extent.	Severity of disease is linked to prognosis. The following bullet point has been added in the section 4.2.3 Detection of disease and assessment of its extent and prognosis: • Better defining the prognosis of the disease in a given patient (e.g. standardised uptake value of PET agents may be independent predictors of disease-free survival or overall survival).
2.2 Indications/ claims (3)	Previous version has an indication of 'Prognostic/therapeutic management guidance'. We do not understand why this has been removed from the list of possible indications. It should be reinstated. Proposal ADD: Bullet 4 'Prognostic/therapeutic management guidance'	Prognostic value of a diagnostic agent is of course of interest; current guideline does not elaborate on such a claim. The following bullet point has been added in the section 4.2.3 Detection of disease and assessment of its extent and prognosis: • Better defining the prognosis of the disease in a given patient (e.g. standardised uptake value of PET agents may be independent predictors of disease-free survival or overall survival).

SPECIFIC COMMENTS ON TEXT		
2.2. Indications/ claims (line 1) (8)	It is incomplete to restrict it to imaging agents when also non-imaging diagnostic medicinal products are included in this guideline. It is also important to take into consideration that a diagnostic medicinal product may be used with other diagnostic procedures and with other medicinal products. Proposed change: "For each indication/claim, the diagnostic agent may be used alone or in combination with other diagnostic procedures or medicinal products (such as furosemide for diuretic urography and renogram, vasodilator for stress pharmacological myocardial perfusion imaging) necessary for the indication claimed,"	Accepted.
2.2.1 Structure delineation (3)	Delete the word "normal" as tests can sometimes also outline abnormal structures. Proposal: 'A common example is the need for outlining anatomic structures (e.g. ……..'	Accepted.
2.2.2. line 4 (3)	The term reference product is not defined in the glossary	Not accepted. This term is not specific to diagnostic agents.
2.2.2 Functional, physiological or biological evaluation (line 2) (8)	It is not necessary to mention that the information is compared to the reference product or the standard of truth, since the purpose is to provide with functional, physiological or biological information. The comparison to the reference product or standard of truth is a requirement for authorisation but not for clinical practice. Proposed change: "The purpose is to provide clinically useful information on functional, physiological or biological evaluations of a tissue, organ or body region ~~when compared to the reference product or the standard of truth~~."	Not accepted.
2.2.2 Functional, physiological or biological evaluation	These studies are of particular importance in some other context, not only test-retest reproducibility but also observer's concordances. Then, it is not important only for this context. Proposed change: "~~In this context, studies of reproductibility are of particular importance.~~"	Not accepted.

SPECIFIC COMMENTS ON TEXT		
(line 4) (8)		
2.2.3. Detection of disease and assessment of its extent (line 7) (8)	FDG is an acronymous which should be written as in European Pharmacopoiea when mentioning FDG-PET. Proposed change: "PET with fludeoxyglucose (^{18}F)"	Accepted.
2.2.3. Detection of disease and assessment of its extent (line 7) (8)	It seems as if monitoring is missing; Proposed change: "…for detection and monitoring the response to treatment…"	Accepted.
3 METHODOLOGICAL CONSIDERATIONS WHEN DEVELOPING AND LICENSING DIAGNOSTIC AGENTS		
Line no. + para no.	**Comment and Rationale**	**Outcome**
3. METHODOLOGICAL (line 1) (8)	Linguistic change: "objectives/claims"	Accepted.
3 Methodological considerations Line 1 (3)	In the first paragraph of this section add a clarification that the text is related to phase III confirmatory trials. (Note that the second paragraph refers to data being obtained from earlier phases in the development (phase II studies)). Proposal: 'In phase III studies the protocol should describe the trial objectives…..'	Accepted.
3 Methodological	It is unclear how the assessment of the trade-off between sensitivity and specificity would be judged, especially since there is a tendency to expect that a minimum value for each will be achieved. A real-world	ROC analysis is meant to help on the assessment of trade off between specificity and sensitivity.

SPECIFIC COMMENTS ON TEXT		
considerations Line 12 (3)	scenario where it is only possible to achieve either a high sensitivity or a high specificity, but not both, needs to be considered. Proposal: ADD: 'In some scenarios it is only possible to achieve either a high sensitivity or a high specificity, but not both. In such cases, there may be a relatively high false positive or false negative rate, and the potential risks of these should be discussed in light of the potential benefits of a true positive or true negative test result.'	Accepted.
Section 3 (6)	Further explanation / examples on the trade-off between sensitivity and specificity could be provided, as this section is rather brief.	ROC analysis is meant to help on the assessment of trade off between specificity and sensitivity.
3. 1. Trial objectives (title) (8)	It has been referred as objectives/claims previously in the text: "Trial objectives/claims"	Accepted.
Section 3 General comment + specific comment on Lines 133-136 (1)	In general, this section appears to focus almost exclusively on one indication only, i.e. detection or assessment of disease, and the main focus of the document is on confirmatory trials, so that limited or no guidance is given to earlier phases of clinical development. Lines 133-136: "*Special attention should be put on the trade-off between sensitivity and specificity, taking the intended clinical use into considerations, and to justify power calculations and acceptance limits in terms of clinical relevance. In this context it is reminded that separate power calculations are necessary for success in terms of sensitivity and specificity*." The regulatory assessment of the trade-off between sensitivity and specificity is left to a significant degree of subjectivity. Besides, there is a tendency to expect that a minimum value for each would be exceeded. Of note, as previously mentioned, sensitivity and specificity are not appropriate as a measure of the performance of *all* diagnostic tests. Proposal: The artificial setting of off-site blinded reading and its effect on the absolute values of sensitivity and specificity, and the possibility that it	Partly accepted. The text will read:

SPECIFIC COMMENTS ON TEXT		
	may only be possible to achieve either a high sensitivity or a high specificity, but not both, needs to be considered and specific guidance given.	'In some scenarios it is only possible to achieve either a high sensitivity or a high specificity, but not both. In such cases, there may be a relatively high false positive or false negative rate, and the potential risks of these should be discussed in light of the potential benefits of a true positive or true negative test result.'
P. 7/16, 2nd parag within Section 3 (7)	Rather than the phrase "justify power calculations" I would prefer to read "provide input to sample-size deliberations"; cf. later remark on this aspect Proposal: "… provide input to sample-size deliberations …"	Accepted.
Section 3.2 (6)	The meaning of the following statement is unclear: "In the eligibility criteria, baseline data obtained with the use of the comparator in the planned study should not be used." Please rephrase to clarify. Please also replace 'extension' with 'extent' (bullets 3 and 6)	Not accepted. The sentence is considered clear enough. Accepted.
3. 2. Patient selection (title) (8)	A clinical trial may include not only patients but any participant (both patients and healthy volunteers). Proposed change: "Study participants or trial population"	Accepted.
3. 2. Patient selection (line 6) (8)	Linguistic change: "…e.g., the effect of different levels of impaired hepatic function.."	Accepted.
3.2. Patient selection Line 9 (3)	Rather than mention study phase, it would be better to mention the study goals, as proposed by ICH guidance E8. Proposal: 'In proof-of-concept studies, both healthy volunteers and patients ….'	Accepted. Healthy volunteers and patients may be included both in proof-of-concept and phase II studies.
P. 7/16, 6 lines before the end of Section 3.2	Patients with other conditions should be included, it is said. I agree, but please add "… separate samples of …" to avoid suggesting that one could mix them into an otherwise duly consecutive sample of cases without destroying its conclusiveness. (The examination of a fan of other conditions is really a Phase II activity, not Phase III.)	

SPECIFIC COMMENTS ON TEXT		
(7)	Proposal: "… separate samples of …"	Accepted.
3. 2. Patient selection (line 24) (8)	Linguistic change: "…patients with other conditions…"	Accepted.
3.2. Patient selection Line 25 (3)	To ensure a "sufficient number" of such subjects would increase the cost and complexity of typical trials and is an overly burdensome requirement. If the inclusion/exclusion criteria for a trial are appropriately chosen, this type of requirement should be unnecessary. Proposal: DELETE: 'In addition, sufficient number of patients with other conditions which could affect the interpretation of the imaging results (e.g. inflammation, trauma) should be included.'	Not accepted.
3. 2. Patient selection (line 26) (8)	It is important to consider not only any concomitant disease but also any concomitant treatment administered at baseline or given during the test procedure to know any potential drug interaction. Proposed change: "In general, any concomitant disease at baseline which may affect the interpretation of results should be well described. Any concomitant treatment at baseline or used for the test should be described and drug interactions with the experimental diagnostic medicinal product evaluated."	Accepted.
3.2. Patient selection Line 28 (3)	This last paragraph in this section is not clear. Is the point being made that patient selection should not be made based on a pre-selected patient group using the comparator? Suggest: 'The protocol should always specify the eligibility criteria for trial participation and the clinical setting where the data are to be collected. Use of a comparator group should not introduce bias into patient selection in any way.'	Partly accepted. The current text will be kept and the proposed text will be added.
3.3	Improved clinical outcome is mentioned but this is too high a burden to place on a diagnostic agent because there are too many variables	Not accepted. This will happen in very rare cases. The current text is sufficiently clear.

SPECIFIC COMMENTS ON TEXT		
Endpoints Line 6 (3)	(related to physician, patient, and treatment) between the diagnostic test result and the patient outcome. Proposal: Delete last sentence of first paragraph in 3.3	
Section 3.3 (6)	The statement at end of paragraph 1 “Improved clinical outcome may be the ultimate way to demonstrate the clinical benefit of an investigational agent.” seems to contradict the earlier statement (section 2.1.4) that the “impact of diagnostic information on management of a patient and clinical outcome… is generally assessed by using an appropriate questionnaire.” Please clarify if a questionnaire would be sufficient to assess clinical benefit, or in which circumstances an improved clinical outcome may be the appropriate way to demonstrate the clinical benefit of an investigational agent	Clarification: A questionnaire or a sequential unblinding will be enough to assess clinical benefit. The assessment of clinical outcome is not mandatory; in some very rare cases, the assessment of clinical outcome may be the only way to assess clinical benefit of a diagnostic agent..
3.4.1. Definitions Line 5 (3)	Hypothesis testing should not always be required. For example, a study could have as the goal the estimation of sensitivity and specificity to within pre-specified confidence limits, and base the sample size on that. Proposal: ADD: ‘Hypothesis testing may not always be required, for example, when a study is designed to estimate sensitivity and specificity to within pre-specified confidence limits.’	This is acceptable for phase I and II studies, but not for phase III studies.
3.4.2. Standard of truth is used (paragraph 2 and 3) (8)	It might be interesting to join sentence "In the absence of standard of truth…..disease state." and "The choice of the surrogate…. Justified." since paragraph 3 mentions about the surrogate standard of truth which appears in paragraph 2. Proposed change: “In the absence of standard of truth, a surrogate standard of truth, such as an appropriate combination of tests, clinical data, repeat testing and clinical follow-up may be used to provide a good approximation to the true disease state. The choice of the surrogate standard of truth is of	Accepted.

SPECIFIC COMMENTS ON TEXT		
	major importance for the interpretation of study data and needs to be fully described and justified."	
3.4.3 General comment (1)	This section needs more clarity to be helpful. It is currently vague and everything is left to the discretion of the regulatory authorities. Proposal: More detailed guidance needed	Not accepted. The proposed text is considered clear enough.
3.5. Comparator (paragraph 2) (8)	The sentence in red link appears in the current guideline: "For imaging contrast agents, the unenhanced procedure may serve as an appropriate comparator in order to evaluate the added value of the contrast agent. If the clinical usefulness of the unenhanced procedure has not been established, however, the clinical relevance of the findings has to be further justified. Still, comparison with a marketed active comparator is highly recommended."	Accepted.
3.5, Lines 252-252 (1)	3.5. Line 251 – 252: "*Still comparison with a marketed active comparator is highly recommended"* Proposal: It should be added here, "*but other comparators (e.g. different imaging methods) may be used instead, if clinically justified. This should be discussed in the protocol"*More detailed guidance needed	Accepted.
3.7. Bias (line 3) (8)	The Clinical Overview is currently a part of the dossier in a registration procedure but this name "Clinical Overview" may be changed in the future and the context of this paragraph may be misunderstood. Proposed change: "…must be critically appraised in the provided documentation for the registration procedure.	Accepted.
3.8.1, Lines 282-284 (1)	*Line 282- 284: "Randomisation in parallel groups is rarely necessary. It is performed if the diagnostic burden in a trial is so high, that not all tests (investigational agent, comparator and other diagnostic agents/procedures necessary for the determination for the standard of truth) may be performed in the same patient. In these situations, patients are randomized to two parallel groups, one group receiving the test and the standard of truth….."*" A parallel-group design may be also used in earlier phases of development (e.g., in dose-response studies) or in other situations, as	Accepted.

SPECIFIC COMMENTS ON TEXT		
	appropriate. The use of a crossover design should not be a requirement, and the choice of the study design should be properly justified in the study protocol. Proposal: "*"The use of a parallel-group study design instead of a within-patient, crossover design may be necessary in some cases, e.g, if the diagnostic burden or the risk in a trial is so high, that not all tests (investigational agent, comparator and other diagnostic agents/procedures necessary for the determination for the standard of truth) may be performed in the same patient. In this case, patients are randomized to two parallel groups, one group receiving the test and the standard of truth….."*	Accepted.
Section 3.8.2 (6)	The meaning of the following statement is unclear: "Blinded evaluation concerns imaging agents in most cases…" It maybe was intended to be "~~Blinded evaluation concerns~~ in most cases imaging agents undergo blinded evaluation ~~in most cases~~ …". Please rephrase to clarify.	Accepted.
4 STRATEGY AND DESIGN OF CLINICAL TRIALS		
Line no. + para no.	**Comment and Rationale**	**Outcome**
4.2. PHASE II (line 3)(8)	Linguisting change: "In addition, phase II studies…"	Accepted.
4.2 (1)	Phase II dose-response studies are generally not required for radiopharmaceuticals. Proposal : Should be explicitly stated	Accepted.
4.3 Phase III studies Paragraph 6 (3)	The text says: 'There are different possible claims in the phase III studies (see section 2.2). Whatever the claim, a clinical benefit of the investigational agent should be demonstrated. At the end it should be clearly shown if the investigational agent contributed to an accurate patient diagnosis, was a useful diagnostic adjunct and/or led to a change in diagnosis and/or patient management.'	

SPECIFIC COMMENTS ON TEXT		
	The requirement to demonstrate clinical benefit cannot be met in a Phase 3 trial. This entire approach places an undue burden on any new diagnostic test that provides physiological or anatomic information that assists the clinician but does not confirm a diagnosis in the absence of other relevant factors. Also, as noted earlier, results from a Phase 3 trial should not have to lead to a change in diagnosis or management. All the investigator can provide is informed speculation, and this should not be the basis for a regulatory approval. Suggest: 'There are different possible claims in the phase III studies (see section 2.2). The level of supporting clinical data required for marketing approval should be related to the claimed indication. If an imaging agent is not claiming a prognostic indication, then technical/diagnostic performance data from clinical studies should be sufficient for approval.'	Not accepted.
Section 4.3 (6)	The statement "Whatever the claim, a clinical benefit of the investigational agent should be *demonstrated*." seems to contradict the earlier statement (section 2.1.4) that the "impact of diagnostic information on management of a patient and clinical outcome…is generally assessed by using an appropriate *questionnaire*.". 'Demonstration' implies that a clinical benefit is directly measured. If a questionnaire approach is used, as recommended in section 2.1.4 then a clinical benefit is reliably inferred but not directly demonstrated. Furthermore if a diagnostic test makes no clinical utility claim, it is inappropriate to demand that a diagnostic agent has to demonstrate clinical benefit or utility. Detection is a neutral method of establishing presence or degree of a signal, whereas clinical benefit is dependent on medical practice. The post-test clinical intervention will determine benefit/risk and is dependent on the skill and knowledge of the treating physician, as well as additional available clinical information on the patient. If clinical outcome is not part of the claim for the diagnostic agent, an application should not be compelled to demonstrate clinical	Accepted. Not accepted.

SPECIFIC COMMENTS ON TEXT		
	benefit. For example, if there are multiple potential therapies or even more combinations of therapies it may be impossible to measure a consistent reliable clinical benefit of the post-test clinical intervention. In another example the diagnostic may identify that a patient will be ***non***-responsive to therapy, or ***not*** able to take therapy. How is benefit to be measured in such a case? Lastly, where it is already known that intervention following detection (structure delineation, functional evaluation or assessment of disease) leads to a clinical benefit, it should not be required to repeat this for every subsequent diagnostic agent in the same setting. For comparison, in the USA a 'clinical utility' claim summons up the FDA's Premarket Approval (PMA) pathway, which is distinct from the 510(k) pathway required for a 'detection-only' claim. Proposal: "~~Whatever the~~ For a clinical utility claim (i.e. instruction on a post-test clinical intervention), evidence or adequate reasoning that a clinical benefit ~~of~~ is expected from the specific intervention following diagnosis using the investigational agent should be ~~demonstrated~~ provided."	It is beneficial not to get an ineffective treatment. Accepted. Indirect assessment is possible Partly accepted. The following will be added: "evidence or adequate reasoning that a clinical benefit ~~of~~ is expected from the specific intervention following diagnosis using the investigational agent should be ~~demonstrated~~ provided."
4.3 Phase III studies Paragraph 10 (3)	Under Statistical considerations Paragraph 2 makes no reference to a trial hypothesis which shows that a product is safer than a comparator Proposal ADD: '…… or to show an enhanced safety profile.'	Partially accepted. This is correct, but not enough; product should also have similar efficacy (specificity/sensitivity).
4.3 Phase III studies	Under Statistical considerations Paragraph 3 - negative and positive predictive values are also relevant Suggest:	Not accepted (NPV and PPV).

SPECIFIC COMMENTS ON TEXT		
Paragraph 11 (3)	'In phase II and III diagnostic clinical trials, the primary hypotheses will in most instances concern sensitivity, specificity, negative or predictive values.'	
4.3 (1)	The requirement to demonstrate clinical benefit cannot be met in a phase III trial. Phase III studies should confirm the principal hypotheses developed in earlier studies and validate the instructions for use in the target population. The Draft EMEA Guideline indicates that multiple primary endpoints should be avoided. In Section 3, lines 135-136 of the document, however, it is reminded that "*separate power calculations are necessary for success in terms of sensitivity and specificity.*" That looks like a contradiction. Besides, it is expected that "*At the end, it should be clearly shown if the investigational agent contributed to an accurate patient diagnosis, was a useful diagnostic adjunct and/or led to a change in diagnosis and/or in patient management.*" Therefore, the same trial(s) should show that the agent: a) improves the diagnostic performance of the test, b) has a positive impact on diagnostic thinking and c) positively affects patient management. This entire approach places an undue burden on any new diagnostic test. Proposal: The recommendations about their design and endpoints should be tailored to the indication being pursued. What the test is intended for should drive the assessment. The Industry Task Force recommends that: a) guidance is given along the lines of the target indications, with detailed requirements for each; b) unless a patient management indication is pursued, a relevant impact on diagnostic thinking or on patient management may be established indirectly or historically.	Not accepted. This would generate too complicated guidance.
4.3. Phase III (line 3) (8)	It is supposed that this text must be a heading (and then in bold letter); if not the sentence is incomplete at all: "**Non-inferiority and superiority trials"**	Accepted.

SPECIFIC COMMENTS ON TEXT		
4.3 Phase III studies Page 11, 5th[h] paragraph (4)	It is acknowleged that the justification of a non-inferiority margin is not necessarily straightforward and usually requires intensive discussions. However why is it explicitely forbidden to switch from superiority to non-inferiority? The EMEA PtC on switching between non-inferiority and superiority allow that, at least in principle.	Accepted.
Statistical considerations Page 11, 11[th] paragraph (4)	Other possibilities should at least be acknowledged. The results of many diagnostic procedures are not necessarily binary, as they result in more than one of 2 states of disease. In these cases, clinical efficacy may be better described by measures other than sensitivity and specificity. Consider an MRA study of a contrast agent in which percent stenosis in an artery or some set of arteries is the primary quantity of interest. The extent to how closely evaluations from contrast and non-contrast MRA match the gold-standard can be done on a continuous scale or a multicategory scale. To use sensitivity and specificity to compare the 2 MRAs, the 0-100 stenosis scale must first be dichotomized into "diseased/serious stenosis" and "non-diseased/no serious stenosis" based on a single cutoff value such as 50%. This seems to be nearly the crudest comparison which can be made, and of necessity ignores potentially important diagnostic characteristics. Categorical accuracy using a scale such as: ≤ 10% (no stenotic disease) 11-49% (mild to moderate disease) 50-99% (severe disease) 100% (occluded) seems a better alternative to describe the state of the disease. This would be more informative for the patient and doctor, who would be better able do discuss treatment options, prognostic issues, and disease progression.	Partly accepted. Usually the decision of whether or not to treat or to give a more aggressing treatment is dichotomous.

SPECIFIC COMMENTS ON TEXT		
Statistical considerations Page 12, 1st paragraph (4)	It should be clarified, that only non-inferiority margins have to be specified in the protocol and not superiority margins are meant here. Proposal: '... likewise superiority or non-inferiority hypotheses for the differences in sensitivity and specificity need to be pre-specified in the protocol. For the case of non-inferiority hypotheses a non-inferiority margin has to be pre-specified as well.'	Accepted.
Statistical considerations Page 12, 2nd paragraph (4)	This would rule out the possibility to define for both sensitivity and specificity a non-inferiority margin and to use a procedure that requires one (out of the two) variables to be non-inferior and the other to be superior. Proposal: 'The overall error rate of the chosen testing procedure has to be controlled so that it is less than alpha. For the case when two hypotheses are specified and both have to be rejected no adjustment for multiplicity is necessary.'	The following text has been retained: Adjustment for multiplicity resulting from the assessment of two primary endpoints (sensitivity and specificity) is usually not required because both primary hypotheses need to be rejected, but the impact on the overall power of the trial needs to be drawn into consideration.
Statistical considerations Page 12, 2nd paragraph (4)	We would prefer a statement that the efficacy population only consists of subjects having a valid standard of truth (at least for trials where the SOT is used).	In some cases standard of truth dose not really exist or can not be used (e.g. data post-mortem). Therefore, the proposed statement can not be put as it is not valid for all cases.
Statistical considerations **P. 11/16, 2nd parag.** **(7)**	The paragraph begins with a statement that I find counterproductive because it does not match realities and thus invites lip service and sets a wrong standard: "As with any other clinical trial, clear hypotheses need to be specified in order to justify the sample-size calculation." Let me spell out my objection, proceeding from the general to the specific. Please see items (i)-(iv) in the APPENDIX below. Taken together these features make it impossible to adhere to the quoted recommendation, and, assuming the Guideline cannot go into all the concerns in (i)-(iv) , I suggest that the statement together with the sentence that follows it be replaced with →	

SPECIFIC COMMENTS ON TEXT		
	'As with any other clinical trial, the added-value measure that the planned investigation aims to estimate must be carefully and rationally chosen in the light of the clinical problem at hand; and an analysis of required sample sizes must document that a satisfactory degree of precision of the estimate can be obtained without undue risks, costs and delays.' (Let me add that, as with any clinical trial, the results, if conclusive and trustworthy, should not be discarded just because the investigators failed to make a detailed sample-size analysis beforehand or failed to report their sample-size deliberations.)	Not accepted. We are here in regulatory setting.
Statistical considerations **The two bullets that follow (p. 11 bottom, p. 12 top) (7)**	With that proposal of mine the two bullets that follow (p. 11 bottom, p. 12 top) on the choice of null and alternative hypothesis lose their anchoring and can be made less prominent or even deleted. The remark on <u>there being no need to adjust for multiplicity</u>, on the other hand, is well taken, although I would argue differently, stressing that the two separate tests are subordinate tests to an overall assessment of diagnostic merit; see → "… is not required once the [main] statistical test has documented a positive gain on a univariate overall-merits scale (Overall Added Value, or Overall Equivalence, in equivalence studies)."	The following text has been retained: Adjustment for multiplicity resulting from the assessment of two primary endpoints (sensitivity and specificity) is usually not required because both primary hypotheses need to be rejected, but the impact on the overall power of the trial needs to be drawn into consideration.
4.3, Line 341 (1)	Line 341: *"However if superiority fails to be shown, the switch to non-inferiority is not possible".* It should be explained why this is not possible, although the CHMP Guidance "Points to consider on switching between superiority and non-inferiority" does not preclude this. Explanation needed.	Accepted.
4.3, Statistical considerations	Line 362 – 364: *"When comparing a new agent with a comparator (where sensitivity and specificity for each of these is determined versus the standard of truth) likewise superiority or non-inferiority margins for the differences in sensitivity and specificity need to be pre-specified*	

SPECIFIC COMMENTS ON TEXT		
Lines 362-364 (1)	*in the protocol."* The Industry Task Force recommends making clear that superiority, non-inferiority or both should be specified in the protocol. It should also be clearly stated, that margins for non-inferiority should be pre-specified (i.e. non-inferiority margin), but that no margin is required for superiority. The current version of the guideline can also be interpreted to mean that a superiority margin is required.	Accepted.
4.3, Statistical considerations Lines 365-366 (1)	Lines 365 -366: "*Adjustment for multiplicity resulting from the assessment of two primary endpoints (sensitivity and specificity) is not required because both primary hypotheses need to be rejected"* This statement does not take into account that there are several statistical testing procedures available that control the type-I-error, but allow to conclude superiority for one of the two variables and non-inferiority for the other without pre-specifying which variable should be demonstrated to be superior and which one to be non-inferior. Proposal : *"The overall error rate of the chosen testing procedure has to be controlled so that it less than α. For the case when two hypotheses are specified and both have to be rejected no adjustment for multiplicity is necessary."*	The following text is retained: Adjustment for multiplicity resulting from the assessment of two primary endpoints (sensitivity and specificity) is usually not required because both primary hypotheses need to be rejected, but the impact on the overall power of the trial needs to be drawn into consideration.
5 DATA PRESENTATION		
Line no. + para no.	**Comment and Rationale**	**Outcome**
P. 12, penult. sentence in 'Roc curves'(7)	This statement introduces the univariate test variable, i.e., the ROC summary measure that serves as an overall-merits variable. By adding the parenthesis "(area under the ROC curve)" the passage leaves the impression that the AuROC is the only, or the only recommended, ROC-associated summary measure of diagnostic merits. In the next sentence it is admitted that reliance on the AuROC is not without its problems. I see the parenthesis as unfortunate for two or three reasons: (1) depending on context another measure may be more appropriate;	

SPECIFIC COMMENTS ON TEXT		
	(2) the area has been criticized for being an unreliable guide [my simple counterexample from Medical Decision Making 1991 (see reference following the APPENDIX!) is still the most convincing one I know]; (3) inspection of the detailed course of the two ROCs being compared may reveal that the two diagnostic procedures would supplement each other but perform too differently to be real rivals, rendering any single-scale comparison misleading. I suggest one writes →: "…converting a bivariate (sensitivity / specificity) to a univariate test variable (such as the much-used Area under the ROC, or another ROC summary measure, as deemed appropriate)."	Accepted. The sentence will read: However, in comparative trials with an active comparator, ROC curves may serve as effective illustration of diagnostic performance converting a bivariate (sensitivity/specificity) to an univariate test variable (area under the ROC curve).
Section 5.1 (6)	The meaning of the following statement (under Reproducibility of the results of the test) is unclear: "This is should be considered at large, including reproducibility of all quantitative measurements…" Please rephrase to clarify	Comment not understood.
5.1. Diagnostic performance (8)	The following information is missing in the current draft and regards to data which should be presented to evaluate diagnostic performance of a diagnostic medicinal product in some specific cases: • interaction studies between the diagnostic medicinal product and some other products whose administration would be useful for the evaluation of the first one (such as captopril for renogram, a PET radiopharmaceutical and a CT contrast for PET/CT studies) should be investigated. • or establishing diagnostic performance of a test with a diagnostic medicinal product in some suspected probable diseases, it may be necessary a long-term follow-up to check how disease progresses and this should be studied in groups of patients with different probabilities (low, medium and high) (e.g. Parkinson's disease).	These comments are generally agreed upon, but if accepted, will put too high a burden in the development and registration of diagnostic agents. Not accepted.

SPECIFIC COMMENTS ON TEXT		
	• actors influencing the diagnostic performance of the technique such as patient's preparation and other factors (such as ECG triggering, CT scanning in PET/CT, etc..) should be analyzed. Proposed change: "Interaction studies between the diagnostic medicinal product and some other products whose administration would be useful for the evaluation of the first one (such as captopril for renogram, a PET radiopharmaceutical and a CT contrast for PET/CT studies) should be investigated. For establishing diagnostic performance of a test with a diagnostic medicinal product in some suspected probable diseases, it may be necessary a long-term follow-up to check how disease progresses and this should be studied in groups of patients with different probabilities (low, medium and high) (e.g. Parkinson's disease). Factors influencing the diagnostic performance of the technique such as patient's preparation and other factors (such as ECG triggering, CT scanning in PET/CT, etc..) should be analyzed."	
5.2.1. Test performance in relation to specific patient population (title) (8)	The test performance refers really to the diagnostic performance which is an item defined and used in this guideline: "***Diagnostic performance in relation to specific patient population***"	Accepted.
5.2.2. Test performance when there is more than one lesion per individual (title) (8)	The test performance refers really to the diagnostic performance which is an item defined and used in this guideline: "***Diagnostic performance when there is more than one lesion per individual***"	Accepted.

SPECIFIC COMMENTS ON TEXT		
Test performance more than one lesion 5.2.2. (5)	Although oncology states are the most obvious application of this section, there are also spatially distributed "lesions" (atherosclerosis, multiple sclerosis) that will vary in number and quality (vulnerability for atherosclerotic plaque, inflammatory status for multiple sclerosis). Does the lesion language in this section also apply to quality differences across lesions within a person?	Not accepted.
5.2.2. Test performance when there is more than one lesion per individual Line 1 (3)	The clinical significance of the number of lesions is important Suggest: 'If more than one lesion can be detected in an individual, and if the number of lesions is clinically relevant, test performance has to be expressed ……'	Accepted.
5.2.3. Impact on diagnostic thinking (3)	Is it possible to elaborate on the meaning of 'impact on diagnostic thinking'?	Not accepted. There has already been done in great extent.
5.2.3. Impact on diagnostic thinking Line 1 (3)	The text says: 'Ideally, the impact on diagnostic thinking should be presented numerically: …. There is no realistic way to accomplish this "numerical" assessment in a standard Phase III trial.	Not accepted.
5.2.3 Impact on diagnostic thinking Page 13, 5th paragraph 4)	Ideally, the impact on diagnostic thinking should be presented numerically **and graphically**	Graphic presentation of the impact on diagnostic thinking might be done but is not mandatory.

SPECIFIC COMMENTS ON TEXT		
5.2.3 Impact on diagnostic thinking Page 13, 5th paragraph (4)	Confidence intervals for secondary endpoints are explicitly mentioned in this section, but not in other sections.	True. No change foreseen.
Section 5.2.3 (6)	See Section 2.1.3 Change section title to "Impact on probability of correct diagnosis" or "impact on diagnostic certainty"	Not accepted.
5.2.3. Impact on diagnostic thinking (8)	The impact on diagnostic thinking is an item not included in the diagnostic performance. Then, it should be numbered apart from the diagnostic performance and in the same level of it and technical performance: "***5.3. Impact on diagnostic thinking"***	Accepted.
P. 13/16, Sections 5.2.3 & 5.2.4 (7)	In these two subsections it may be appropriate to suggest a more quantitative approach (by how much does the diagnostic uncertainty change?, etc.) along with a graphical analysis (a scatterplot of before-against-after probability assessments, say). In fact, the draft guideline has very little on the use of graphs.	Not accepted. This is a quantitative approach; it is not usually done in a guideline.
Same sections (7)	Moreover, the two subsections do not go into the possibility that a novel diagnostic procedure may redefine disease stages (or even diseases) rather than just improving diagnoses within an existing categorization or taxonomy of patient conditions. A new diagnostic procedure may for instance serve to single out for special treatment a subclientele (whose members might undeservedly count as false either negatives or positives in a routine analysis). Nor do the subsections go into the related topic of the use of RCTs in assessing diagnostic advances.	Accepted.
5. 2.4. Impact on	The impact on patient management is an item not included in the diagnostic performance. Then, it should be numbered apart from the	Accepted.

SPECIFIC COMMENTS ON TEXT		
patient management (8)	diagnostic performance and in the same level of it and technical performance: ***5.4. Impact on patient management***	
5.2.4. Impact on patient management (3)	Paragraph one It is unrealistic to expect Phase III studies to provide information on subsequent patient management for all diagnostic indications. The level of supporting data required should be related to the claimed indication. If an imaging agent is not claiming a prognostic indication, then technical/diagnostic performance data from clinical studies should be sufficient for approval. Proposal ADD: 'The level of supporting data required should be related to the claimed indication. Patient management data is not required to support all diagnostic indications. If an imaging agent is not claiming a prognostic/therapeutic management guidance indication, then technical/diagnostic performance data from clinical studies should be sufficient for approval.'	The minimal requirement for registration is an adequate technical and diagnostic performance of a new diagnostic agent in relation to a standard of truth and, when appropriate, to an established comparator in the clinical context in which the diagnostic agent is to be used in well-designed superiority or non-inferiority trials. Impact on diagnostic thinking should also be discussed. Demonstration of the impact on patient management and clinical outcome is rarely, if ever, necessary (except if there is a specific claim).
Section 5.2.4 (6)	Under the title "*Impact on patient management*", the statement "Where appropriate, it is assessed prospectively by using appropriate questionnaires and presented in the application quantified by the rate of change in patient management" seems inconsistent with the earlier statement in section 5.2.3 that "in some protocols, it may be specified …that the referring clinician may not change the diagnosis and the management based only on the results of a new agent.". Quantification of an actual change is not feasible if a) clinical management would be assessed by a questionnaire, and if b) the referring clinician may not change the diagnosis nor the management based only on the results of the new agent. Proposal "Where appropriate, ~~it~~ the impact on patient management is assessed prospectively by using appropriate questionnaires. If the protocol allows the clinician to change the diagnosis and the patient management, such additional data can be ~~and~~ presented in the application quantified by the	Not accepted. The current text is considered to be clear enough: Where appropriate, impact on patient management is assessed prospectively by using appropriate questionnaires and quantified by the rate of change in patient management pre- and post-test. All elements to

SPECIFIC COMMENTS ON TEXT		
	rate of change in patient management"	be taken into account to establish the scheduled management of a given patient should be clearly defined in the study protocol. These generally include the state-of-the art diagnostic work-up and may include data obtained with the comparator. The consequences and adequacy of the induced changes to the scheduled management should be assessed by using follow-up data.
6 REQUIREMENTS FOR REGISTRATION		
Line no. + para no.	**Comment and Rationale**	**Outcome**
6 Requirement foe authorization Line 5 (3)	Diagnostic imaging agents may be molecular imaging agents intended to be used for pathophysiological investigations and disease detection. These agents should be approvable with technical and diagnostic performance data supported by literature data providing evidence for impact on diagnostic thinking. Proposal: ADD: 'Molecular imaging agents intended to be used for pathophysiological investigations and disease detection should be approvable with technical and diagnostic performance data supported by literature data providing evidence for impact on diagnostic thinking.'	Not accepted. Impact on diagnostic thinking of such an investigation should always be discussed. The current text will be kept: "In functional imaging or pathophysiological explorations, the assessment of biological or physiological processes may form the basis for an approval."
6.1 Requirements on study data for new diagnostic agents Paragraph 4 (3)	The guidance should include a statement that for new products where the therapeutic consequences of the new product are not obvious, published literature may be used to support a benefit or impact on diagnostic thinking. Proposal Paragraph 4 ADD as second sentence of this paragraph: 'Published literature may be used to support a benefit or impact on diagnostic thinking for the product.'	Accepted.
6.1 Requirements on study data for new	Regarding the last sentence of paragraph 4 which currently states: 'Relevant impact on diagnostic thinking should be demonstrated even if no treatment exists yet for a disease.' In cases where there is no treatment there is still benefit to the patient	

SPECIFIC COMMENTS ON TEXT		
diagnostic agents Paragraph 4 (3)	and society, The diagnostic workup stops, the patient and family may obtain closure, there may be a clear prognosis, the patient and family may better be able to plan their lives, and expenditure of health resources stops (they can now be used for other patients). Suggest change to: 'Impact on diagnostic thinking should be discussed even if no treatment exists yet for a disease .'	Accepted.
Section 6.1 (6)	The section seems unclear and contradictory: **Detection claim:** performance = diagnostic certainty), since there is no intrinsic clinical benefit to the agent and the outcome of any intervention will depend on medical practice and competence. **Clinical utility claim:** questionnaire suffice? If "therapeutic consequences of the diagnosis obtained with a new agent are not obvious," how should "a benefit in patient management and/or clinical outcome" be demonstrated? Recommend to separate this section into 2 subsections: o Where the diagnostic makes a clinical utility claim, the term 'benefit/risk' could remain. o Where the diagnostic makes no clinical utility claim but simply a 'detection' claim, replace references to 'benefit/risk' with 'performance/risk'. Except where there are established comparator diagnostic agents, the impact on diagnostic certainty should always be assessed, whether the claim is for clinical utility or for detection.	Not accepted. There is always an intrinsic clinical benefit, otherwise diagnostic agent should not be used/developed.
6.2. Requirements on study	It is more complete to say diagnostic medicinal products that not only of products: **6.2. Requirements on study data for diagnostic medicinal products**	Not accepted.

SPECIFIC COMMENTS ON TEXT		
data for products similar to already approved products (title) (8)	**similar to already approved products**	
6.2. Requirements on study data for products similar to already approved products (line 5) (8)	The fact that the technical and diagnostic performances have to be evaluated in relation to a standard of truth or acceptable comparator is in the current guideline and has been deleted in this draft guideline. Both agents (the experimental and the approved ones) should also have a similar or better pharmacodynamic profile. Proposed change: "Non-inferiority comparative trials versus a similar already approved agent to demonstrate similar technical and diagnostic performances (sensitivity and specificity) in relation to a standard of truth or acceptable comparator as well as similar or better safety and pharmacodynamic profiles in the same patient population/indication are recommended."	Not accepted. Products similar to already approved products have similar pharmacodynamic profiles, but these are not necessary to compare in a non-inferiority trial. Non-inferiority between the two products with respect to technical and diagnostic performances (sensitivity and specificity) as well as similar or better safety profile in the same patient population/indication is considered sufficient.
6.2. Requirements on study data for products similar to already approved products (8)	It is important to include that the Applicant should provide with studies updating the conditions of use in case that the approved product has outdated specifications of use. Proposed change: "The Applicant should provide with studies updating the conditions of use in case that the approved product has outdated specifications of use."	Not accepted. Out of scope in a clinical development guideline.
6.2. Requirements on study data for products	The fact that the technical and diagnostic performances have to be evaluated in relation to a standard of truth or acceptable comparator is currently in the guideline and has been deleted in this draft guideline. A mention to their similarities on safety and pharmacodynamic profiles should also be included.	Not accepted. Products similar to already approved products have similar

SPECIFIC COMMENTS ON TEXT		
similar to already approved products (line 10) (8)	Proposed change: "….to better technical and diagnostic performances in relation to a standard of truth or acceptable comparator, better safety and pharmacodynamic profiles, the impact on diagnostic thinking and patient management may need to be shown…"	pharmacodynamic profiles, but these are not necessary to compare in a non-inferiority trial. Non-inferiority between the two products with respect to technical and diagnostic performances (sensitivity and specificity) as well as similar or better safety profile in the same patient population/indication is considered sufficient.
6.2. Requirements on study data for products similar to already approved products (line 12)(8)	The fact that the technical and diagnostic performances has to be evaluated in relation to a standard of truth or acceptable comparator is currently in the guideline and has been deleted in this draft guideline. A mention to their similarities on safety and pharmacodynamic profiles should also be included. Proposed change: "However, if the impact of the diagnosis is well known for the comparator (the approved similar product), better technical and diagnostic performances and better safety and pharmacodynamic profiles may be sufficient to support the superiority claim."	Not accepted. Products similar to already approved products have similar pharmacodynamic profiles, but these are not necessary to compare in a non-inferiority trial. Non-inferiority between the two products with respect to technical and diagnostic performances (sensitivity and specificity) as well as similar or better safety profile in the same patient population/indication is considered sufficient.

7 CLINICAL SAFETY ASSESSMENT		
Line no. + para no.	**Comment and Rationale**	**Outcome**
7 CLINICAL SAFETY ASSESSMENT (line 3) (8)	This information is already included in other sections of the guideline, and it is not related to safety. Then, it is not necessary to iinclude it here. Proposed change: "~~in some cases, a diagnostic agent will have to show a positive impact on diagnostic thinking and patient management in order to support MA claim."~~	Not accepted. It will stay to keep the context.
Section 7 (1)	"*Safety follow-up of patients should not be limited to the duration of the diagnostic procedure but extended to a longer time period corresponding at least to the pharmacokinetic and pharmacodynamic properties of the product.*" It is commonly agreed (and feasible) to have a safety follow-up	

SPECIFIC COMMENTS ON TEXT		
	extended to a longer time period, e.g. 24 or 48 hrs after administration. This 24 or 48 hrs period corresponds also to a complete elimination for **non-specific** diagnostic agents which have an elimination half-life of 90-120 min in Man. However, some specific products (e.g. USPIO) have a longer residence time in the body (several weeks) and it is clearly not feasible to have a safety follow-up of patients on such a long period of time in a clinical trial. Risk of accumulation of such products in the body is avoided due to the single administration of diagnostic agents. "*Not only short-term but also long-term safety (when appropriate) should be properly addressed. An appropriate RMP should be established for agents accumulating in the organism (e.g. deposits of Gd in bones & skin).*" Establishing long-term safety (e.g. in the case of the NSF issue described in this sentence) is not feasible in a "regular" clinical trial (NSF may takes several years to develop in some patients). This should be done appropriately after approval on a large-scale population, on the basis of a RMP agreed with the Authorities during the review process. Proposal: "*Safety follow-up of patients should not be limited to the duration of the diagnostic procedure but extended to a longer time period corresponding at least to the pharmacokinetic (time for elimination) and/or pharmacodynamic properties of the product. A maximum of 48 hrs follow-up after administration will be considered appropriate, unless specific concern.*"	Not accepted. This can not be specified.
7 Clinical safety assessment Line 10 (3)	Since pharmacodynamics is related to pharmacokinetics it is unnecessary to mention it. Pharmacokinetics is enough to set the duration of safety assessment. Suggest: 'More specifically, safety follow-up of patients should not be limited to the duration of the diagnostic procedure but extended to a longer time period corresponding at least to the pharmacokinetic properties of the product.'	Not accepted. The current text will be kept.

SPECIFIC COMMENTS ON TEXT

_8 GLOSSARY (DEFINITIONS)

Line no. + para no.	Comment and Rationale	Outcome
8 GLOSSARY (line 4) (8)	The second sentence is already included in the introduction where it may be more appropriate. It is not part of the definition, then there is no sense to include it in the definition. A diagnostic agent is specifically a medicinal product. Proposed changes: "*Diagnostic agent*: any medicinal product used as part of a diagnostic test (i.e. together with the equipment and procedures that are needed to assess the test result). ~~In this document, the discussion on diagnostic agents is restricted to those administered into or onto the human body.~~	Not Accepted.
8 GLOSSARY (line 11) (8)	The accronymous EU has not been detailed previously in the guideline. Proposed change: "*Comparator*:…. European Union"	Accepted.
8 GLOSSARY (8)	The definition of endpoint, already included in the current guideline, is missing in this draft guideline. It is necessary to include taking into account that there is a section titled "endpoint" Proposed change: "*Endpoint* (definition included in the current guideline)."	Not accepted. Endpoint definition is not specific for diagnostic agents.
8 Glossary (3)	Add definition for reference product.	Not accepted. Reference product definition is not specific for diagnostic agents.
Likelihood (and likelihood ratio) (7)	Margaret Sullivan Pepe has recently made the very nice proposal that the "likelihood (-s, ratios)" of the present contexts should be called diagnostic likelihood (-s, ratios), abbr. DLR, in order to avoiding confusion with the statistical likelihood concept (cf. statements like 'the positive and negative likelihoods are estimated by maximum likelihood'). I would like to see these more precise terms used.	Currently used terms are considered adequate.
Nega-/positive PV	Incidentally, the jargon terms posi-/negative predictive value and likelihood should be avoided in an authoritative text as this. One	Current text is considered adequate.

SPECIFIC COMMENTS ON TEXT		
& likelihood ratio (7)	should be speaking of the 'predictive vale of a negative test [result],' abbr. PV_{neg}, and the '[diagnostic] likelihood ratio associated with a negative test,' abbr. DLR_{neg}, a.s.o. After all, neither a "likelihood" nor a PV can be negative!	
Prevalence (7)	Prevalence: the slight problem with this term has been discussed already (see General above)	No specific comment.
APPENDIX		
Line no. + para no.	**Comment and Rationale**	**Outcome**
(7)	APPENDIX (i) Textbook-style alpha-beta formulae are an aid to choosing a reasonable sample size but cannot stand alone. In the narrow context of hypothesis comparisons, what is missing is an analysis of the necessary trade-off between formal Type I/II error risks and the circumstances of the investigation (constraints on total project duration, patient flow, manpower, costs); a surprisingly small beta increase can often make a project manageable. Missing is also a sensitivity analysis concerning a guessed standard deviation and concerning the choice of alternative ("positive added value") hypothesis. An additional concern is differential reaction in subgroups, whether suspected a priori or not. As a result, it does not make sense to calculate a sample size, and textbooks ought to replace the term 'sample-size calculation' with 'sample-size considerations (or: analysis).' (ii) Diagnostic evaluations are essentially measurements rather than tests of benefit (= added value), so a reasonable sample-size is one that provides a suitably low standard error (SE) for an appropriately chosen measure of benefit. Hypothesis testing often plays a secondary role. (iii) Whether benefit is a function of sensitivity, specificity, correlation coefficients, cost, utility or other parameters being estimated, it is generally so that the SE depends on the unknown parameter(s). For	No specific comment.

SPECIFIC COMMENTS ON TEXT		
	instance, all evaluations that involve one or more binomial parameters close to 0 or 1 possess this feature. Again a range of plausible values must have their sample-size requirement assessed in the light of a hoped-for precision (hence SE). (iv) The latter, too, may itself depend on the unknown parameters being estimated, for the following reason. If the truth is that the hoped-for benefit of implementing work-up policy A instead of B is palpably negative, there is no need to know precisely how negative it is; if the benefit is moderately positive, it is typically desirable to estimate it rather precisely in order to be able to weigh it against secondary concerns such as installation costs; if, finally, the benefit is overwhelming, it may again be a luxury to ask for great precision.	
Appendix 1 Para 2.1.3 (1)	"*An imaging agent is only useful…The fast evolving technological progresses can be such that an imaging drug developed over several years could become obsolete by the time of marketing application*." One concern is that due to the time it takes to conduct the clinical development of a new contrast agent and due to the methodological constraints of a clinical protocol, it is difficult to follow the evolution of the equipments & sequences which are sometimes evolving very rapidly. Proposal: The basis of the evaluation of the dossier, in terms of techniques and equipments used during clinical trials, should be what has been agreed during scientific advice, unless a consensus has been reached within the EU radiological/medical community for the use of a new technique or procedure which makes the new diagnostic agent obsolete.	Accepted.
Appendix - 2. Evaluation of efficacy (8)	There is no sense to include a section 2.1. whenever there is no more sections 2.2., 2.3….. In fact section 2 and section 2.1. have similar titles. Proposal: ~~**2.1. Assesment of efficacy**~~	Accepted.
Appendix -	If deleting 2.1. Assesment of efficacy, this section should be	Accepted.

SPECIFIC COMMENTS ON TEXT		
2.1.1. *Technical performance and practicality (8)*	renumbered.: "**2.1. *Technical performance and practicality"***	
Appendix - **2.1.1. *Technical performance and practicality (line 25)(8)***	Both sentences should be joined as referring to the same context: "Quantitative assessment consists on measuring the ability of the new agent to alter the image density/signal intensity of target structures (normal structures, abnormal structures, etc.). This can be done by measuring signal intensities during time course after administration of the imaging agents, to calculate parameters like target to noise ratio, percent enhancement."	Current text will be kept.
Appendix - **2.1.2. *Diagnostic performance (8)***	If deleting 2.1. Assesment of efficacy, this section should be renumbered.: **2.2. *Diagnostic performance***	Accepted.
Appendix - **2.1.3. *Technological dependence (8)***	If deleting 2.1. Assesment of efficacy, this section should be renumbered.: **2.3. Technological dependence**	Accepted.
Technological dependence Appendix 1 2.1.3 (5)	It is recommended that an imaging agent state the minimum performance characteristics necessary to achieve the stated specificity and sensitivity. This would make the imaging agent more "future proof" and highlight conditions of mismatch where the full diagnostic potential is unlikely to be realized.	Current text is considered sufficient.
Appendix - 3.2. Bias (8)	The Clinical Overview is currently a part of the dossier in a registration procedure but this name "Clinical Overview" may be changed in the future and the context of this paragraph may be misunderstood. Proposal: "…must be critically appraised in the provided	Accepted.

SPECIFIC COMMENTS ON TEXT		
	documentation for the registration procedure."	
Appendix - Safety assessment (8)	Only the radiation exposure is required to be addressed when radiopharmaceuticals are used. However, the radiation exposure when other diagnostic medicinal products are used in combination with X-rays or Computed Tomography or other test involving exposure to ionising radiation should be addressed since the patient is exposed to radiation in all these cases. Then, a new paragraph requiring such an information is mandatory. Proposal: "**4.2. Radiation exposure when diagnostic medicinal products are used in combination with tests requiring patient's exposure to ionising radiation** Information about absorbed radiation doses in various body tissues should be presented by the applicant and the estimation should preferably be based on studies in patients. The units of measurement should be those established by the International System."	Not accepted.