European Medicines Agency

Pre-Authorisation Evaluation of Medicines for Human Use

London, 23 July 2009
Doc. Ref. CPMP/EWP/1119/98/Rev. 1

**COMMITTEE FOR MEDICINAL PRODUCS FOR HUMAN USE
(CHMP)**

GUIDELINE ON CLINICAL EVALUATION OF DIAGNOSTIC AGENTS

DISCUSSION IN THE EFFICACY WORKING PARTY (EWP)	May 1998/ January 2000
TRANSMISSION TO CPMP	November 2000
RELEASE FOR CONSULTATION	November 2000
DEADLINE FOR COMMENTS	February 2001
DISCUSSION IN THE EFFICACY WORKING PARTY	June 2001
TRANSMISSION TO CPMP	November 2001
ADOPTION BY CPMP	November 2001
DRAFT REV. 1 AGREED BY EFFICACY WORKING PARTY	April 2008
ADOPTION BY CHMP FOR RELEASE FOR CONSULTATION	26 June 2008
END OF CONSULTATION (DEADLINE FOR COMMENTS)	31 December 2008
AGREED BY EFFICACY WORKING PARTY	June 2009
ADOPTION BY CHMP	23 July 2009
DATE FOR COMING INTO EFFECT	1 February 2010

This guideline replaces guideline CPMP/EWP/1119/98.

KEYWORDS	*CHMP, EMEA, Diagnostic Agents*

<div style="border: 1px solid black; padding: 10px;">

GUIDELINE ON CLINICAL EVALUATION OF DIAGNOSTIC AGENTS

</div>

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The revision of the points to consider on diagnostic agents was decided in order to reflect better the necessary steps in development of diagnostic agents as well as to define the assessment of benefits (technical performance, diagnostic performance, impact on diagnostic thinking and impact on patient management/outcome) and the risks related to the development of these agents. Principal chapters such as possible indications/claims, patient selection, endpoints, standard of truth, strategy and design of clinical trials, statistical considerations and data presentation, have also been reviewed. In addition, a chapter on the requirements for registration of products similar to already authorised products has been added.

The Appendix on the development of imaging agents has also been reviewed. Technical performance has been rewritten as well as the section on different types of blinding.

1   INTRODUCTION

Diagnostic agents are medicinal products used for diagnosis or monitoring of a disease. The evaluation of diagnostic agents is governed by the same regulatory rules and principles as for other medicinal products. The principles used for the evaluation of medicinal products with respect to quality, pharmacology, toxicology, pharmacokinetics and safety apply to diagnostic agents; however, as diagnostic agents are used to diagnose and/or monitor diseases/conditions and not for treatment, clinical development programmes should be adapted for these purposes. As for other medicinal products, benefits and risks related to the use of diagnostic agents are taken into account when granting a marketing authorisation.

2   SCOPE

This note for guidance outlines the principles for the clinical evaluation of diagnostic agents that are intended for *in vivo* administration only; it is restricted to those diagnostic agents administered into or onto the human body. It applies for, but is not limited to, structure delineation, functional assessment including biological and physiological processes, detection or assessment of diseases or pathology as well as prognostic and/or therapeutic management guidance.

Any *in vitro* use or therapeutic use of diagnostic agents will not be covered here.

- Radiopharmaceuticals as defined in European Directive 89/343/EC, intended for diagnostic use.
- Contrast agents for use in imaging techniques, including X-rays, magnetic resonance imaging (MRI) and ultrasonography (US).
- Compounds used in diagnostic tests that do not involve radionuclides or imaging techniques (e.g., allergen extracts for skin prick test, histamine in lung provocation tests, urea (^{13}C) for breath test).
- Various stains/markers, e.g., stains used intraoperatively in detection of malignant mucosal lesions.

This document is presented in two parts, a general part and an appendix. The general part deals with overall principles and applies to clinical trials forming the core of a registration application of diagnostic agents in general. The appendix gives the current classification and methodologic requirements related to the development of imaging agents. Appendices detailing outstanding issues for other groups of diagnostic agents may be added in the future.

The way of performance (technique) of diagnostic procedures will not be covered. However, it is generally recommended that diagnostic agents should be developed and evaluated by using adequate, state-of-the-art techniques and equipments.

3   LEGAL BASIS

This guideline should be read in conjunction with the introduction and general principles (4) and the Annex I to Directive 2001/83 as amended.

Applicants should also refer to other pertinent EU and ICH guidelines, and particularly:

- Good Clinical Practice (ICH topic E6).
- Statistical Principles for Clinical Trials (ICH topic E9).
- Choice of Control Group in Clinical Trials (ICH topic E10).
- Structure and Content of Clinical Study Reports (ICH topic E3).

4   FUNDAMENTALS IN THE CLINICAL EVALUATION OF DIAGNOSTIC AGENTS

Diagnosis of disease requires careful clinical assessment. A diagnostic test is useful if it can decrease the uncertainty with respect to the diagnosis or to the disease stage / extent or prognosis. A diagnostic test is embedded in a diagnostic work-up that consists of a battery of diagnostic procedures including anamnesis, physical examination, etc. Appropriate diagnostic testing can then assist in making a correct diagnosis and in providing additional information to guide patient management. A correct diagnosis is clearly a benefit for a patient; an incorrect diagnosis (e.g. false negative or false positive results of the diagnostic tests) may be detrimental for a patient and can have important consequences in patient management. In addition, there may be safety concerns with use of some diagnostic agents.

In the assessment of diagnostic agents, the choice of patient population is crucial. If healthy volunteers and patients with a confirmed disease may be acceptable in proof of concept and phase II trials, subjects included in confirmatory trials should be representative of the population in which the diagnostic agent is intended to be used. In general, these are the patients with diagnostic uncertainty, which a new diagnostic agent should decrease (see sections 6.2 and 6.3); the validation of the new diagnostic agent should be performed in this place of the diagnostic work-up that is intended in later diagnostic practice.

Whatever the indication, simple visualisation of an anatomic structure, which does not confer benefits to the patient, is considered insufficient. The demonstration of clinical benefit should be tailored to the diagnostic agent being used and its potential claims. In most cases, clinical benefit of a diagnostic agent may be demonstrated by assessing its technical performance, diagnostic performance and by an appropriate discussion on the impact on diagnostic thinking. Depending on the type of claim, and in some particular situations (e.g. where no standard of truth is available), impact on patient management, and clinical outcome, may also need to be assessed. In addition, the measurement of clinical outcome might also be required if a diagnostic agent has e.g. better diagnostic performance but is less safe than other diagnostic procedures.

Safety of diagnostic agents should be assessed in all cases.

4.1   Assessment of benefit of diagnostic agents

4.1.1   Technical performance

It consists of image quality, and/or procedural advantages/disadvantages of an investigational agent/test and, if applicable, in comparison with a comparator. Technical performance should include evaluation of precision of the diagnostic agent which consists of observer's concordance (if subjective outcome) or reproducibility test-retest (if objective outcome). Technical performance alone is necessary but not sufficient to show clinical benefit of a diagnostic agent and cannot be the only basis for registration. Structure delineation by itself is not sufficient if it does not deliver information beneficial for the patient.

4.1.2 Diagnostic performance

It consists of sensitivity and specificity of a test. In case the test decision is based on a cut-off value the trade-off between sensitivity and specificity requires careful analysis with respect to intended applications of an experimental test and their implications on patient care. The impact of disease prevalence should also be discussed, as co-morbidity, specificity and sensitivity may vary in different study populations.

4.1.3 Impact on diagnostic thinking

It refers to the impact of a test result on post-test versus pre-test probability of a correct diagnosis, in relation to a well-defined clinical context (patient characteristics, prior diagnostic procedures). All diagnostic agents should have an impact on diagnostic thinking (higher probability of correct diagnosis after the test than before the test, or change in diagnosis); impact on diagnostic thinking should be discussed in the marketing authorisation application even if no treatment exists for a disease.

Positive and negative predictive values are important parameters which influence the impact on diagnostic thinking in a given patient. Both negative and positive predictive values depend on the prevalence of a disease in the studied series and may not necessarily reflect the prevalence of a disease in the overall population.

The role of a diagnostic test involving a diagnostic agent or of a combination of diagnostic tests in the determination of the prognosis of a patient (e.g. occurrence of a cardiac event during a given period of time or overall survival in oncology) may represent a major part of its impact on diagnostic thinking. The demonstration of this prognostic value should be brought by means of adequate statistical methods such as multivariate analysis. It is important to demonstrate that the diagnostic test brings information on the prognosis that is independent from other data of the conventional work-up or may replace independent prognostic factors which are more demanding to obtain.

4.1.4 Impact on patient management (in particular therapeutic decisions)

It refers to a description and quantification of impact of diagnostic information gained with the diagnostic agent on patient management. Both impact on diagnostic thinking and impact on therapeutic decisions may be assessed by using an appropriate questionnaire or by sequential unblinding. Patient follow-up data should be available for this purpose. 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.

4.1.5 Impact on clinical outcome

Studies assessing patient outcomes may be required if there is no standard of truth to compare to (see section 5.4.3). In all other cases, these studies are not mandatory, but if performed can be the basis for a specific claim.

4.2 Indications/claims

For each claim, the diagnostic agent may be used alone or in combination with other diagnostic procedures or medicinal products (e.g. furosemide for diuretic urography) necessary for the indication claimed; this should always be specified in the study protocol.

Broadly speaking, the possible indications/claims of diagnostic agents may be grouped as follows:

- Structure delineation for imaging agents or some stains/markers;

- Functional, biological and physiological evaluation;

- Detection and/or assessment of disease, its extent and/or its prognosis.

4.2.1 Structure delineation

A common example is the need for outlining anatomic structures (e.g. structures that could not be seen well with other imaging agents such as delineate bowel loops on CT scan).

4.2.2 Functional, physiological or biological evaluation

Some examples are: measurement of regional cerebral blood flow, assessment of cardiac ejection fraction or assessment of bronchial hyper-responsiveness. 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. In this context, studies of reproducibility are of particular importance.

4.2.3 Detection of disease and assessment of its extent and prognosis

Common clinical situations are:

- providing a diagnosis in patients with suspected disease (e.g. scintigraphy or angioscan in suspected pulmonary embolism, detection of thyrotropic function failure with TRH challenge).

- monitoring the extent/rate of progression or response to treatment of disease in patients with previously confirmed disease (e.g. MRI contrast agents for detection of brain metastases, or PET with fludeoxyglucose (18F) for detection, monitoring and response to treatment of whole body metastases, or detection of sentinel lymph node with blue dye). In some cases, the agent can provide good visualisation of adjacent tissues e.g. some dyes for intra-operative imaging which target the normal tissue.

- 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).

4.2.4 Multiple indications/claims

If several indications/claims are planned for one imaging agent, they may have to be assessed in separate clinical situations/trials.

5 METHODOLOGICAL CONSIDERATIONS WHEN DEVELOPING AND LICENSING DIAGNOSTIC AGENTS

In the phase III studies, the protocol should describe the trial objectives/claims, products and methods investigated (including the investigational agent, absolute or surrogate standard of truth, comparator, and other clinical assessments and procedures if used), testing procedures, trial population, sample size calculation, endpoint justification, blinding, randomisation, statistical considerations, principles for data presentation, issues related to collection and analysis of data, safety and any other relevant considerations.

Relevant data on the diagnostic performance (specificity and sensitivity) of an investigational agent obtained from the earlier phases of its clinical development (phase II studies) should be used to design subsequent confirmatory trials. 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 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.

In this context, it is reminded that separate power calculations are necessary for success in terms of sensitivity and specificity. If the disease is rare, obviously the required precision for estimating the sensitivity of the test determines the sample size of the trial.

In any case, it is of particular importance to consider the intended clinical use of the product and design the trials accordingly. For example, whether the agent/test under investigation should be used as add-on in case of insufficient diagnostic information based on established tests, or as an alternative to standard tests.

5.1   Trial objectives

The confirmatory trials of an experimental agent or of a new indication for an approved agent often aim to establish the agent's superiority or non- inferiority relative to an established active comparator and/or to show acceptable levels of inferiority when compared with standard of truth. The principles used for conclusion of superiority, equivalence or non-inferiority in comparison with other tests should be defined and justified in the trial protocol.

5.2   Trial population

While designing early phase clinical trials, the applicant may consider including a broad spectrum of patients with respect to:

- manifestations of the target disease or anatomical condition of interest;

- physical attributes (e.g., age, sex, body fat to muscle mass ratio);

- levels of function of the organ system(s) responsible for the elimination of the diagnostic agent (e.g., the effect of different levels of impaired hepatic functions including cirrhosis on the elimination of the agent by the liver), if patients with impaired drug elimination functions are later to be included in the indication.

In proof-of-concept and phase II studies, both healthy volunteers and patients with a confirmed disease may be included, in order to get preliminary information on technical performance and diagnostic performance of a diagnostic agent.

In the confirmatory trials, the choice of patients and the clinical setting should be appropriate to provide data to support the diagnostic claim(s) for the diagnostic agent. Subjects included in confirmatory trials should be representative of the population in which the diagnostic agent is intended to be used, e.g. if the agent is developed to diagnose a disease, clinical trial may be performed:

- in asymptomatic patients (screening)

- in patients with suspected but not confirmed disease

- in patients with a confirmed disease for evaluation of its extent (initial staging), severity or prognosis

- in patients undergoing treatment to monitor its efficacy

- in previously treated patients to search for recurrence

- in patients with a confirmed recurrence for evaluation of its extent (restaging), severity or re-evaluate the prognosis.

In addition, sufficient number of patients with other conditions which could affect the interpretation of the imaging results (e.g. inflammatory lung lesions in patient suspected to have pulmonary metastases) should be included.

In general, any concomitant disease at baseline which may affect the interpretation of results should be well described.

In the eligibility criteria, patient data obtained with the use of the comparator in the planned study should not be used.

The protocol should always specify the eligibility criteria for trial participation and the clinical setting where data are to be collected.

5.3 Endpoints

For a new claim/indication, appropriate clinical test variables are normally given directly by the rationale for test development, e.g. signs or test data related to presence or absence of a disease or grading of organ dysfunction.

Appropriate primary endpoints may include diagnostic performance (sensitivity and specificity), predictive values, likelihood ratios, evaluation of prognosis, impact on diagnostic thinking and impact on patient management or on clinical outcome.

For most diagnostic agents, appropriate co-primary endpoints are sensitivity and specificity of an investigational agent; the improvement in specificity, sensitivity and, consequently, in the certainty of diagnosis is reflected in the improvement of a diagnostic thinking.

In some particular situations, when there is no standard of truth (or it can not be accurately determined), impact on patient management or improved clinical outcome may be the ultimate way to demonstrate the clinical benefit of an investigational agent (see section 5.4.2).

When a diagnostic agent is used in a procedure yielding quantitative results, e.g. blood clearance, ejection fraction, etc., improved precision in the measurement of a relevant parameter constitutes an appropriate endpoint for active comparator in comparative studies, provided that other parameters of diagnostic performance remain unaltered. In this case, it is recommended that studies of agents for functional, physiological, or biochemical assessment indications provide a quantitative or qualitative understanding of how the measurement varies in normal and abnormal subjects or tissues. It is critical to identify the range that is normal and the values that indicate the abnormality. When possible, the minimum detectable limits and reproducibility of the measurement should be assessed. Reproducibility assessments are most meaningful when performed within the same subject. However, under some circumstances, this practice might be unethical in which case applicants should consider alternative approaches to testing reproducibility (e.g. in animals).

5.4 Standard of truth

5.4.1 Definitions

Standard of truth is believed to give the true state of a patient or the true value of a measurement. It provides an independent way of assessing the same variable being assessed by the investigational diagnostic agent. Standards of truth are used to demonstrate that the results obtained with the investigational diagnostic agent are valid and to define diagnostic performance. After the standard of truth has been selected (e.g. histopathology after surgery), the hypothesis for the expected diagnostic performance of the investigational agent in reference to the standard of truth should be determined to reflect the intended population and clinical setting for use of the diagnostic agent.

5.4.2 Standard of truth is used

In confirmatory studies, a diagnostic agent/test should be shown to provide valid information by comparing the results yielded by the investigational diagnostic agent with the results of the standard of truth. Clear description of the testing procedures is required and the choice of standard of truth needs to be justified. The standard of truth by definition can truly reflect the presence or absence of the target

disease, or the true value of a measurement, but may not be clinically appropriate outside the setting of a clinical trial, for instance due to cost, complexity or delay in reaching a diagnosis.

In the absence of standard of truth, a surrogate standard of truth, such as an appropriate combination of tests, clinical data, repeat diagnostic work-up 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 major importance for the interpretation of study data and needs to be fully described and justified.

When standard of truth is included in the trial, efficacy analyses are in principle straightforward, especially if a within subject analysis is possible. The diagnostic performance of the investigational diagnostic agent (and the comparator, if applicable), should be expressed according to the objectives of the trial, e.g. in terms of sensitivity and specificity, or change in probability of a correct diagnosis pre-test and post-test with relation to the standard of truth.

In the confirmatory trials, the standard of truth should be established independently of the results of the investigational agent and of its comparator. The standard of truth should not include as a component any information obtained with the investigational agent or the comparator. Standards usually do not undergo blinded reading procedures, or independent assessment by separate readers, but it can be wise to do so in some instances (e.g. multiple blind reading of histopathologic specimens when the histologic assessment is known to be difficult).

5.4.3   Standard of truth cannot be used

As a general principle, a standard of truth is required to assess the diagnostic performance of a diagnostic agent. In special circumstances it might not be feasible or ethical to obtain a standard of truth or to determine it in an accurate manner.

If there is a well documented comparator available, "concordance" in a cross-over study can be used as outcome measure. The study population should be representative for the variability of the condition under investigation. In the case of discordant findings in the individual patient, further investigations such as biopsies or long term follow-up without intervention should be undertaken to establish the true state of the findings. If this is not feasible, it might be necessary to conduct a randomised parallel group study comparing the new test as add-on to the standard procedure versus the standard procedure. Impact of patient management and clinical outcome would in these rare cases provide the necessary information of the benefit of the new diagnostic procedure.

Alternative approaches may be used which must be discussed and justified in the protocol, but in cases where a standard of truth cannot be used, regulatory acceptance through scientific advice procedures is recommended prior to the initiation of confirmatory trials.

5.5   Comparator

In the event that an investigational agent is being developed as an alternative or improvement over existing diagnostic agents, comparative studies are requested where both investigational agent and selected comparator are compared to the standard of truth. It is essential to ensure that the selected comparator is appropriate, widely accepted in the EU for the claimed indication and reflects current medical practice. The choice of a comparator must be justified and the corresponding procedures clearly described. The comparison should include an evaluation of both efficacy and safety data.

For imaging contrast agents, see remarks in appendix 1.

5.6   Need for placebo

For most non-imaging diagnostic agents, the use of placebo is not useful for efficacy assessment, except when the response to a diagnostic test is assessed using subjective evaluation criteria (e.g. skin changes in a skin prick test). For the assessment of tolerability, administration of placebo followed by a dummy procedure can be of interest.

Use of placebo for imaging agents: see appendix 1.

5.7 Bias

It is important to minimise the extent of possible observer bias by determining the true disease-state of subjects using the standard of truth independent of the investigational agent. Possible bias in trial design, patient selection and work-up, conduct and interpretation of results must be critically appraised in the provided documentation for the registration procedure.

5.8 Study design

The appropriate group comparison is either within-subject, or a parallel group design. A within-subject comparison is where the investigational agent, standard of truth and, if appropriate, a comparator are assessed in the same subject. The advantage of this type of trial design is a potential reduction in the variability and consequently a decrease of the variance of the estimates for the diagnostic performance of the diagnostic test. Whenever feasible and ethically acceptable, within-subject comparison of tests is preferred in order to reduce the variability of the experiment and to increase the power of the study.

A parallel group design instead of a within-subject comparison may be needed in some cases, e.g. if the number of diagnostic tests that can be performed in the same subject must be limited due to, for example, their invasive nature or cumulative irradiation, or a carry-over effect of one or both products.

5.8.1 Randomisation of the administration of agents

In general, two kinds of randomisation may be performed in clinical trials with diagnostic agents: randomisation to two parallel groups of patients, and randomisation of the order of administration of diagnostic agents. The most frequently, comparisons are done within-subject (see section 5.8). The sequence of administration of agents (investigational agent, comparator and other diagnostic agents/procedures necessary for the determination of the standard of truth) is randomised unless it is considered inappropriate (e.g. when standard of truth involves invasive procedures, e.g. surgery). The most important design feature of diagnostic trials is that the second test is assessed without knowing the results of the first test. Information carry-over from one diagnostic procedure to the other may result in similar findings in the two tests. Therefore, it is of paramount importance, that this aspect is clearly described in the study protocol.

Randomisation to two parallel groups is sometimes necessary. It is performed if the diagnostic burden in a trial is so high, that not all tests (investigational agent, comparator and other diagnostic procedures necessary for the determination of the standard of truth) may be performed in the same patient. In these situations, patients are randomised to two parallel groups, one group receiving the experimental test and the standard of truth, and the other group receiving the comparator and the standard of truth. Here, also the order of administration of agents may be randomised if appropriate.

5.8.2 Blinding

In most cases, imaging agents undergo blinded evaluation (see appendix on imaging agents for methodological requirements related to blinding).

In addition, blinded evaluation is important whenever the criteria of assessment are subjective (e.g. skin prick test).

6 STRATEGY AND DESIGN OF CLINICAL TRIALS

6.1 Phase I studies

The aim is to obtain pharmacokinetic and first human safety assessments with single mass dose and increasing mass doses of a diagnostic agent. If an agent targets metabolic process or receptors, this

should also be studied in the appropriate pharmacodynamic trials. Phase I studies may be done in healthy volunteers or in patients (e.g. in case of toxicity/irradiation).

6.2 Phase II (dose-response) studies

These studies are aimed to determine the mass dose or dosing regimen in patients, to be used in the phase III studies, and to provide preliminary evidence of efficacy and safety, as well as to optimise the technique and the timing, e.g. for image acquisition or blood sampling. In addition, phases II studies are important for developing methods/criteria by which images and/or test results will be evaluated.

Phase II dose-response studies are generally not required for radiopharmaceuticals (see appendix 1).

To get the preliminary evidence of efficacy, both subjects with known disease and subjects without a disease should be included.

In addition to technical image quality/characteristics and determination of the mass/radioactivity to be used in patients, phase II studies might also aim at assessing diagnostic performance of the investigational agent. Training of readers may be based on images obtained from phase I or II trials. Consistency between readers should be measured quantitatively.

Therefore, in phase II studies, technical imaging quality is recommended as a primary endpoint for imaging agents, and diagnostic performance as a secondary endpoint for all agents.

6.3 Phase III studies

See also sections 4.2 (claims) and 5 (methodological considerations).

The criteria by which the diagnostic agent will be evaluated should generally be developed in phase II studies and then incorporated in the phase III study protocols.

Phase III trials are large-scale trials to establish efficacy of an investigational agent in the well-defined target patient population (e.g. in patients with suspected but not confirmed disease) and in that step of the decision making process, where the test will be used in later clinical practice.

The results of the diagnostic test depend not only on the diagnostic agent itself but also on the techniques and equipments that must be used for the diagnostic procedures (e.g. imaging machines, serum assays for a pharmacologic provocative agent, endoscope for a mucosal stain). It is a general requirement for multicentre phase III trials that homogeneity, reproducibility and quality of the performances of those techniques and equipments must be assessed and, if possible, cross checked within all centres before the start of the assay, when centralised techniques and equipments are used (e.g. serum assays performed in one single reference centre).

Prior to designing the efficacy trial, claims and endpoints (e.g. specificity and sensitivity) should be carefully defined (see section 5.3). The primary efficacy variable should be clinically relevant and evaluable/measurable in all patients. Multiple primary endpoints should be avoided; superiority or non-inferiority for both sensitivity and specificity are considered as one co-primary endpoint. When multiple claims are requested, studies may be done in different clinical settings, each corresponding to the particular claim and intended use.

There are different possible claims in phase III studies (see section 4.2). Whichever the claim, a clinical benefit of the investigational agent should be demonstrated in the registration file, e.g. it should be clearly shown that the investigational agent contributed to an accurate patient diagnosis or evaluation of the prognosis, and/or led to a change in diagnosis and/or in patient management.

In cases where it is already known that intervention following the use of diagnostic agent leads to a clinical benefit, it will not be required to repeat this proof for every subsequent diagnostic agent in the same setting. However, evidence of an adequate reasoning that a clinical benefit is expected should be provided.

6.3.1   Non-inferiority and superiority trials

A trial may be designed to show that an investigational agent is not inferior to a comparator (and thus could be an alternative to this comparator), usually by means of a within-patient comparison. For example, if the study endpoint is the presence or absence of disease, the sensitivities and specificities of the investigational agent and the comparator will be compared (both values are obtained by reference to the standard of truth). The statistical hypothesis may be non-inferiority, superiority or both. However, if superiority fails to be shown, the switch to non-inferiority is not possible, unless a non-inferiority margin has been pre-specified and justified in the study protocol (see "PtC for switching between superiority and non-inferiority).

The choice of whether superiority or non-inferiority is an acceptable objective of the trial deserves careful justification at the primary stage. The justification of the non-inferiority margin will depend not only on the clinical setting, but include also the anticipated diagnostic performance and impact on management of other available diagnostic agents/procedures in the relevant clinical setting and on its context (emergency, existence of alternative therapies, etc.)

6.3.2   Statistical considerations

Sensitivity and specificity of a new diagnostic agent are considered as one single co-primary endpoint in phase III trials and need to be determined with precision (confidence) so that depending on disease prevalence, positive and negative predictive values can be computed.

As with any other clinical trial, clear hypotheses need to be specified in order to justify the sample-size calculation. These hypotheses need to be in line with the chosen objectives of the trial and of the diagnostic test. Phase III trials should clearly outline the added value of a new diagnostic agent in the diagnostic workup of patients, e.g. is it supposed to replace a more invasive test or procedure or to show a better sensitivity and/or specificity than an existing comparator.

Target values for specificity and sensitivity should be prospectively formulated (e.g. sensitivity greater than 0.8) and sample size calculation should be based on the assumed values for sensitivity and specificity and their confidence intervals or by formulating the respective one-sided hypotheses.

When comparing a new agent with a comparator, where sensitivity and specificity for each is determined versus the standard of truth, superiority or non-inferiority hypotheses for the difference in sensitivity and specificity, respectively, need to be pre-specified in the protocol. For the case of non-inferiority hypothesis, non-inferiority margins for the differences in sensitivity or specificity need to be pre-specified.

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.

7   DATA PRESENTATION

7.1   Technical performance

The evaluation of the technical performance of a diagnostic agent should comprise the procedural aspect and the reproducibility of the results obtained from the diagnostic test.

Procedural aspect:

-    The potential advantages or disadvantages of the agent should be evaluated with respect to convenience and material safety for product preparation, reconstitution, handling (for example a plastic bottle is less dangerous for the technologist than a glass bottle with metallic seals), mode of administration (for example an intravenous route seems less risky than an intra arterial one) and/or sampling (e.g. arterial blood, venous blood, air breathed) and timing of the procedure.

- The procedural convenience and safety aspects should be evaluated from both the patient and technologists perspective.

Reproducibility of the results of the test:

- This should be considered at large, including reproducibility of all quantitative measurements assessed on criteria similar to that recommended for *in vitro* diagnostics (e.g. coefficients of variation within observers and between observers, between several serial measurements if ethically feasible) and of qualitative information derived from the diagnostic agent (e.g. coulour of a structure induced by a dye, analysis of the content of an image obtained with a contrast medium by several readers).

- The analysis of reproducibility of diagnostic tests including imaging will be further discussed in appendix 1.

7.2   Diagnostic performance

Specificity and sensitivity are generally calculated on a per patient basis; for some applications, they may be also calculated on per site or per lesion basis. The confidence interval of those values should be given. For quantitative results or semi-quantitative results (e.g. grades of a scale indicating the relative confidence in the interpretation), cut-off values between a negative and a positive result should be specified in the protocol, according to experience of phase II studies, prior to any evaluation of the current data. The impact of disease prevalence should also be discussed.

7.2.1   ROC curves

ROC (receiver operating characteristics) curves (variation of sensitivity according to 1-specificity) are mainly considered to be a means for selecting appropriate cut-off points to be used prospectively in confirmatory trials. The cut-off point may be either a quantitative threshold value for quantitative continuous variables (e.g. left ventricular ejection fraction, maximal expired volume per second …) or a level in a gradual scale of certainty in imaging interpretation. 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). Observed differences may be hard to interpret in terms of clinical relevance, but "better than" results can, for example, be used to support clinical non-inferiority conclusions.

7.2.2   Factors to consider for data presentation

− **Diagnostic performance in relation to specific patient population**

Not only may the optimal trade-off between sensitivity and specificity vary in relation to population and purpose (population screening, diagnosis, treatment follow-up, etc. see section 5.2), the diagnostic performance of an agent may also be affected by the stage of disease. Therefore, ROC curves generated from different populations or sub-groups in the study sample might be of value in the optimisation of the test, e.g. for patients with different grades of severity of the target disease, or patients who are treated versus untreated for the target disease.

The predictive values of a diagnostic agent in detecting a disease of interest in any population is dependent on the prevalence of the disease in the studied sample of patients; e.g., the positive and negative predictive values for the general population are different from those for the population at high risk for developing the disease. This has to be taken into account when using predictive values derived from investigation in specialised centres when applied to general practitioner's practices. Predictive values of a diagnostic agent must be reported with caution, as the predictive values of a diagnostic agent in detecting a disease of interest is dependent on the prevalence of the disease in the population studied. In principle, the validation of the new diagnostic agent should be performed in the setting of the diagnostic work-up that is intended in later diagnostic practice (see section 4) providing the

predictive values for this population. These predictive values are recommended not to be transferred to another setting /population.

− **Diagnostic performance when there is more than one lesion per individual**

If more than one lesion can be detected in an individual, test performance has to be expressed both in relation to an overall individual (per patient basis) and to lesions detected and/or organs or sites involved (per lesion or per site basis). Evaluation of sensitivity, specificity and other relevant measurements may still be applicable if they are based on justified cut-off limits for the number of lesions. In this context, additional criteria might need to be introduced, e.g., criteria for comparing the clinical relevance of two metastatic lesions in the same segment of the liver versus two metastases in different liver segments.

If patients with known disease are recruited for established comparator controlled studies, number and distribution of lesions if demonstrable by the investigational agent may be considered if validated by a standard of truth and appropriately handled statistically, even if the clinical relevance of the findings *per se* may be questioned.

7.3   Impact on diagnostic thinking

The impact on diagnostic thinking may be presented numerically; the rate of cases where diagnostic uncertainty with a new agent has decreased as compared to pre-test diagnosis should be reported (percentage, and confidence intervals). Positive and negative predictive values may help clinicians modify diagnostic thinking if reasonable thresholds have been reached.

The impact on diagnostic thinking may influence patient management (e.g. change in a stage of a disease may induce a change in treatment) or not. In some protocols, it may be specified (and must be justified) that the referring clinician may not change the diagnosis and the management based only on the results of a new agent. In these cases, the possible impact on diagnostic thinking and patient management should be discussed with the help of supportive and academic published studies or simulated e.g. by performing sequential unblinding.

7.4   Impact on patient management

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 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.

8   REQUIREMENTS FOR AUTHORISATION

In general, the approval of a diagnostic agent is usually based on clinical indications rather than the general properties of a specific molecule. Nevertheless, these general properties should still be described in the application.

In functional imaging or pathophysiological explorations, the assessment of biological or physiological processes may form the basis for an approval.

Acceptable safety profile should be demonstrated in all cases.

In practice, the requirements for authorisation may differ for completely new diagnostic agents and products similar to already approved products.

8.1 Requirements on study data for new diagnostic agents

In all cases, it is required to demonstrate 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.

In cases where there are established and widely accepted comparators for the condition in question, and when it is already known that intervention following the use of diagnostic agent/comparator leads to a clinical benefit, it will not be required to re-demonstrate the impact on diagnostic thinking for each subsequent use of a diagnostic agent in the same setting. However, relevant impact on diagnostic thinking and/or patient management in the appropriate clinical context should be demonstrated, if therapeutic consequences of the diagnosis obtained with a new agent are not obvious, or the benefit/risk balance is unclear, and if the diagnostic agent itself may have immediate therapeutic implications. It may be useful to refer to published literature.

In addition to considerations concerning efficacy and safety of diagnostic agents, patient's comfort and tolerability of invasive procedures and methods of assessment stated in trial protocol have to be considered. Similarly, presence or absence of obstacles to introduction of a new test/agent into clinical practice and how that might affect the technical performance of the investigational test should also be discussed.

8.2 Requirements on study data for products similar to already approved products

Some products (e.g. contrast agents as non-targeted iodinated monomers or gadolinium chelates) share several similarities with an already approved agent (chemical structure/class, pharmacokinetic profile, dose and dosing regimen of the active moiety), but frequently differ in the chemical structure of the carrier molecules.

Well-designed comparative trials versus a similar already approved agent to show at least non-inferiority with respect to technical and diagnostic performances (sensitivity and specificity) as well as similar or better safety profile in the same patient population/indication may be sufficient. Such a limited evidence for assessing clinical benefit of these products is based on the claim(s) for the same indication as that already granted to the similar approved agent. In case of a superiority trial, this limited evidence may not be sufficient and, in addition, to better technical and diagnostic performance, the impact on diagnostic thinking and patient management may need to be discussed in the submission. However, if the impact on the diagnosis is well known for the comparator (the approved similar product), better technical and diagnostic performance may be sufficient to support the superiority claim (see section 8.1).

If an indication not granted to the approved product is claimed by the similar product, then the requirements for the registration of this indication are those of a new product (see section 8.1).

9 CLINICAL SAFETY ASSESSMENT

If a diagnostic agent is aimed at diagnosing a disease only, a serious safety concern might preclude marketing authorisation if there are alternative (and safer) diagnostic methods. Marketing authorisation for a diagnostic agent will always be based on its benefit/risk balance; in some cases, a diagnostic agent will have to show a positive impact on diagnostic thinking and patient management in order to support marketing authorization claim.

More generally, clinical safety assessments of diagnostic agents should be tailored based on their characteristics, intended uses (including dose, route of administration, frequency of use, biological half-life, pharmacology and toxicology, etc.) and results of other relevant clinical studies. 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 and pharmacodynamic properties of the product.

Not only short-term but also long-term safety (when appropriate) should be properly assessed. An appropriate risk management plan should be established for agents accumulating in the organism (e.g. deposits of gadolinium in bones and skin).

Risks related to the agent itself (e. g. immunogenicity, allergic reactions) and also to incorrect diagnosis induced by its use should be taken into consideration while assessing benefit/risk balance of the new diagnostic agent.

Evaluation of safety of associated test procedures (e.g., radiation exposure) and possible problems associated with incorrect handling of test procedures must be discussed.

DEFINITIONS

Certain terms that are referred to in this document are described below.

Diagnostic test: any procedure performed to increase the probability of a correct diagnosis. The result of a diagnostic test (test result) can be dichotomous, ordinal or continuous.

Diagnostic agent: any pharmaceutical 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.

Standard *of truth:* a diagnostic tool that has been critically evaluated and documented to identify the true disease state or true value of measurement.

Surrogate standard of truth: a diagnostic test or a combination of tests or follow-up which has been shown to provide a very good approximation to the true disease state or value of measurement.

Comparator: Test or agent to which the investigational agent is compared with. Includes agents/tests approved in the EU for the indication (=active comparator), the unenhanced test procedure and/or placebo.

Diagnostic decision matrix: When a diagnostic test yields a dichotomous result, four combinations of test result and disease state are possible:

Table 1: Diagnostic decision matrix

True disease state

		Present	Absent
Test result	Positive	TP True Positive	FP False Positive
	Negative	FN False Negative	TN True Negative

- *Sensitivity:* the probability that a test result is positive given the subject has the disease. In a suitable experiment the sensitivity can be estimated by: TP/(TP+FN).
- *Specificity:* the probability that a test result is negative given a subject does not have the disease. In a suitable experiment the specificity can be estimated by: TN/(TN+FP).
- *Diagnostic performance:* comprises sensitivity and specificity of a test. It 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.
- *Likelihood ratio (LR):* the likelihood that a given test result would be expected in a patient with the target disorder compared to the likelihood that that same result would be expected in a patient without the target disorder.
  - ➢ *Positive LR:* refers to the LR in case of a positive test: Sensitivity/(1-Specificity)
  - ➢ *Negative LR:* refers to the LR in case of a negative test: (1-Sensitivity)/Specificity.
- *Negative predictive value:* the probability that a subject does not have the disease given that the result of the test is negative. In a suitable experiment, where the prevalence of disease is representative of the average prevalence of disease in the population, the negative predictive value can be estimated by: TN/(TN+FN).
- *Positive predictive value:* the probability that a subject has the disease given that the test result is positive. In a suitable experiment the positive predictive value can be estimated by: TP/(TP+FP).

- *Accuracy or probability of a correct test result:* the probability that the test result reflects the true disease state. In a suitable experiment the probability of a correct test result is estimated as the proportion of cases for which the test result is correct: (TP+TN)/(TP+FP+TN+FN).
- *Reproducibility (precision):* the ability of a diagnostic test to reveal the same result when repeatedly performed on the same individual, assessed repeatedly by the same reader (intra-observer reproducibility) and assessed by different readers (inter-observer reproducibility).

ROC curve (receiver operating characteristics curve): a graphical presentation of the relationship between the sensitivity and specificity of a diagnostic test as threshold value of the test variable is changed.

A priori or pre-test probability of a correct diagnosis: the probability of a correct diagnosis based on the information available before performing the diagnostic test.

A posteriori or post-test probability of a correct diagnosis: the probability of a correct diagnosis after addition of the test result to the information already available.

Within subject comparison of tests: this refers to at least two different tests being performed in a subject for assessing the same set of possible diagnoses in order to compare the diagnostic performances of the respective tests.