The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use

London, 15 November 2001
CPMP/EWP/1119/98

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP)

POINTS TO CONSIDER ON THE EVALUATION OF DIAGNOSTIC AGENTS

DISCUSSION IN THE EFFICACY WORKING PARTY	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

Note:

Points to Consider have been developed to provide advice on selected areas relevant to the development of medicinal products in specific therapeutic fields.

This document will be revised in accordance with the scientific advances made in this area.

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@emea.eu.int http://www.emea.eu.int

POINTS TO CONSIDER ON THE EVALUATION OF DIAGNOSTIC AGENTS

1. INTRODUCTION

1.1 Background and Purpose

This Points to Consider document provides guidance for the evaluation of diagnostic agents. It should be read in conjunction with Directive 75/318/EEC, as amended, and is intended to assist applicants with its implementation. 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).

The evaluation of diagnostic agents is governed by the same regulatory rules and principles as 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 but there are no generally accepted principles for the evaluation of efficacy of diagnostic agents.

This document aims to outline the principles for the evaluation of efficacy of diagnostic agents that are intended for *in vivo* administration. Such uses may include, but are not limited to, structure delineation, functional assessment including biological and physiological processes, detection or assessment of diseases or pathology as well as prognostic/therapeutic management guidance. Any proposed therapeutic use of diagnostic agents would not be covered here.

Medicinal products used as diagnostic agents include:

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

This document is presented in two parts, a general part and an appendix. The general part deals with overall principles and applies to confirmatory trials forming the core of a registration application. The appendix is concerned with radiopharmaceuticals and contrast agents used in imaging. Appendices detailing outstanding issues for other groups of diagnostic agents may be added in the future. Diagnostic procedures are not covered in this document.

1.2 Glossary

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.

Absolute standard: a diagnostic test that has been critically evaluated and documented to identify the true disease state or true value of measurement.

Surrogate standard: 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 new investigational agent is compared. Includes agents/tests approved for the indication (=active comparator), the unenhanced test procedure and placebo

Diagnostic confidence: Degree of subjective certainty in making a diagnosis as measured by predefined scoring systems.

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)
- *Likelihood ratio (LR)*: The LR is 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 is negative. In a suitable experiment 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).
- *Reliability (precision):* the ability of a diagnostic test to reveal the same result when repeatedly performed on the same individual and assessed by different readers.

Test variable: the continuous or categorical observation assessed in a diagnostic test, for example, air flow in a metacholine lung provocation test or imaging signs indicative of presence of disease that is subsequently used as a basis for defining the diagnostic performance of the test.

Endpoint: the efficacy variables in a clinical trial, e.g., test accuracy or impact on final diagnosis, treatment or clinical outcome.

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.

2. FUNDAMENTALS IN THE EVALUATION OF DIAGNOSTIC AGENTS

Diagnosis of disease requires careful clinical assessment. Appropriate diagnostic testing can then assist in making a correct diagnosis and in providing additional information to guide patient management.

In order to establish an indication for a diagnostic agent, it is necessary to assess the diagnostic performance of the new agent as well as to assess the ability of the agent to provide useful clinical information. The diagnostic performance of a new agent or test, however, can in principle only be determined if the true disease-state of the subject under investigation can be defined. Despite considerable technical advancement in the field of diagnostics, this frequently remains a major challenge. (see 3.4 and 3.5)

The assessment of the clinical usefulness of a diagnostic agent includes specifically technical performance, procedural convenience, diagnostic performance and safety of the test procedure, and, more generally, the impact of the use of the agent/test on diagnostic thinking, therapeutic decisions and clinical outcome. Depending on the specific indication, some of the components of clinical usefulness of the agent have to be established directly, some may be established indirectly or historically. The level of directly demonstrated clinical usefulness may be presented in the SPC. The requirements on study data to support a marketing authorisation are detailed in section 6 of this document. Trials aiming at evaluating especially "higher levels" of clinical usefulness, e.g. impact on diagnostic thinking, should be carefully designed to reflect the proper setting and conditions for clinical use and state the diagnostic questions that the agent is intended to resolve.

The performance of a diagnostic agent should, wherever possible, be expressed in terms of sensitivity, specificity, positive and negative predictive value. Positive and negative likelihood ratios may also be used to present sensitivity and specificity data in an easily interpretable way. In addition, it is of importance to characterise the change between pre-test and post-test probabilities of a correct diagnosis in a study population reflecting clinical practice, especially when new diagnostic fields are entered, or if a test is proposed as an "add-on" to existing tests. Such information is required for selecting an appropriate test to aid clinical decision making and interpretation of test results.

3 METHODOLOGICAL CONSIDERATIONS

A detailed trial or study protocol is required before commencing clinical trials or studies. The protocol should describe the trial objectives, products and methods investigated (including the experimental agent, standard, 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 of an experimental agent obtained from the earlier phases of its clinical development 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 this context it is reminded that separate power calculations may be necessary for success in terms of sensitivity and specificity.

For trials designed to provide data on higher levels of clinical usefulness 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.

3.1 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 an absolute or surrogate standard (="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.

3.2 Patient selection

To the extent possible, subjects included in confirmatory trials should be representative of the population in which the diagnostic agent is intended to be used. The protocol should specify the eligibility criteria for trial participation and the clinical setting where data are to be collected.

3.3 Test variables and endpoints

For a new test, 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, etc. Appropriate primary endpoints include sensitivity and specificity, or impact on diagnostic thinking. Improved clinical outcome may be the ultimate way to demonstrate the clinical usefulness of a new test.

If a new agent is under development with characteristics very similar to existing agents, it may be appropriate to use alternative endpoints to sensitivity, specificity, etc. In these cases a thorough justification is expected, normally supported by data showing that the depicted endpoint, compared with conventional endpoints, improves the difference detecting ability of the study, e.g. based on dose-finding studies. As the use of alternative endpoints in confirmatory trials may have major consequences, regulatory scientific advice might be advisable.

When a diagnostic agent is used in a procedure yielding quantitative results, e.g. blood concentration, ejection fraction, etc. improved precision in the measurement of a relevant parameter might constitute an appropriate endpoint for active comparator comparative studies, provided that other parameters of diagnostic performance remain unaltered.

3.4 "Standard of truth"

In confirmatory studies, a diagnostic agent/test should, if at all possible and ethically defensible, be shown to provide valid information by comparing the results yielded by the new agent with the results of an absolute standard. Clear description of the testing procedures is required and the choice of absolute standard needs to be justified. Likewise, omission of an absolute standard has to be justified also. The absolute standard by definition can truly reflect the presence or absence of the target disease, 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 an absolute standard, a surrogate standard, such as an appropriate combination of tests, clinical data, repeat testing and clinical follow-up may be used if known to provide a very good approximation to the true disease state. The choice of surrogate standard is of major importance for the interpretation of study data and needs to be fully justified.

In the confirmatory trials, the absolute or surrogate standard should be established independently of the investigational test or agent and the standard should not include as component any information obtained with the new test or agent. A component of the surrogate standard may, however, meaningfully serve as a comparator in a blinded comparison (see 3.6.4) of diagnostic performance with the new agent, but this may introduce a bias in favour of the comparator if the surrogate standard is less than perfect. Standards normally do not undergo blinded reading procedures, or independent assessments by separate readers.

3.5 Comparator

In the event that an experimental test or agent is being developed as an alternative or improvement over existing tests or agents, comparative studies are expected and facilitate the assessment of clinical usefulness. It is essential to ensure that the selected comparator is appropriate and reflects current medical practice as the estimation of capability of a new test often depends on the comparator used. The choice must be justified and the test procedures clearly described. The comparison should include an evaluation of both efficacy and safety data.

Placebo can rarely be used meaningfully in the case of diagnostics. Saline, however, has been used as a "placebo"/vehicle during the development of US contrast enhancing agents. Vehicle/placebo-controlled trials may provide important information with respect to acceptance limits in case of non-inferiority, active comparator studies. For imaging agents, the unenhanced procedure may serve as an appropriate comparator. If the clinical usefulness of the unenhanced procedure has not been established, however, the clinical relevance of the findings has to be further justified directly or indirectly.

3.6 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 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 an increase in the precision of the estimates for the accuracy of the diagnostic test. Whenever feasible and ethically defensible, within subject comparison of tests is preferred. A parallel group design may be needed ifthe number of diagnostic tests that can be performed in a subject must be limited due to, for example, their invasive nature.

3.6.1 Absolute or surrogate standard is used

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

3.6.2 Standard cannot be used

When no absolute or surrogate standard exists, the within subject comparison of the investigational agent and the comparator is complicated by the fact that there is no immediate way of deciding on the correct diagnosis when different results are obtained. Under these circumstances it is of special importance to select the best possible comparator and, based on available data, to define the prior distribution of sensitivity and specificity data for the comparator as fully as possible.

If a comparator has been shown to provide a good approximation to the truth and if its accuracy is better than that likely with the diagnostic procedure being tested, the comparator may in principle be used as a standard (implying open read, etc. see 3.6.4) and the results analysed accordingly. The uncertainty involved in the diagnosis from the comparator, however, must be adequately allowed for in the analysis. Evaluation of the blinded inter-reader reliability of the comparator within the studies may also provide a check on the validity of the underlying assumptions.

In some instances it can be demonstrated that the accuracy of a comparator approximates that of a standard for a definable range of its diagnostic performance. With qualification, the comparator can than serve as a standard in clear-cut cases and study results can be analysed accordingly, while for the remaining cases other methods as discussed below have to be employed. Estimates of overall performance, however, should be based on results from all individuals whether clear-cut or not.

If a comparator with reasonably good accuracy is available, but, based on prior experience and pharmacological considerations, the accuracy is unlikely to be superior to that of the diagnostic procedure being tested, concordance in test results may serve as an endpoint. In that case, discordant findings should be further characterised as fully as possible. If discordant findings are well characterised in an unbiased manner, these results may be used to draw conclusions, e.g. with respect to superiority of the test product.

Histology results from biopsies may be used to confirm the presence of a disease and the specificity of the procedure can be set to 100%, but sensitivity is frequently hard to estimate due to limitations in the sampling procedures. Similarly, surgery combined with histology may be used to confirm positive findings. In case of negative findings for the new test and the comparator, however, confirmatory surgery may be unfeasible.

If all comparators have poor accuracy, the results of conventional studies are likely to be impossible to interpret and evaluation of clinical outcome should be considered. For within patient trials and if effective therapy is available, long-term follow-up may not be useful. A parallel group study design with clinical outcome as measure of efficacy may therefore be necessary.

3.6.3 Randomisation

Random allocation of patients should be used for parallel group comparisons of diagnostic tests.

Despite blinding, it may be impossible to eliminate the possibility of "information-carry-over". For within subject comparisons, the sequence of investigational agent and comparator should therefore be randomised unless inappropriate.

It may also be desirable to randomise the standard in order to test for interaction. When blinding is not possible and if information carry-over cannot be eliminated, however, it may well be preferable to use the standard only after the test and comparator have been used. This is also the case if the standard involves invasive procedures, e.g. biopsy or surgery which may alter the area of interest

3.6.4 Blinding

Whenever the evaluation of test result involves some "subjective" element, the readers should be blinded for the results of the other tests under concurrent investigation. If possible, masking for test results should also include trial participants and those performing the investigations.

"Blinded reader studies" refer to instances where readers have little or no knowledge of the patient's characteristics or prior history, e.g. in the evaluation of imaging results. These studies are in principle not undertaken to estimate the discriminatory accuracy or to directly establish the clinical utility of a certain diagnostic test, but to assess the reliability of a test result. The blinded reader design, however, provides the opportunity for an assessment of the diagnostic performance of the investigational agent/test and comparator *per se* in a demanding, albeit artificial setting. If clinical usefulness has been demonstrated for a comparator in the population and clinical circumstances of interest, a blinded reader design may be appropriate also for confirmatory comparative studies. If the new test and the comparator differ in mode of action, care should be taken, however, to demonstrate that pre-test diagnostic information available in clinical practice will not differentially affect the added diagnostic value of the new test and the comparator.

Full blinding of readers may not always be necessary and can even be counterproductive. This is particularly relevant when a new diagnostic principle is introduced and the added diagnostic value to available diagnostic tools including clinical assessment is to be demonstrated. These studies should therefore aim to reflect the expected clinical setting for the use of the new diagnostic agent in estimating its discriminatory accuracy. Sequential unblinding, i.e. with readers who evaluate test results with progressively more information on each read, may also be used in these cases. Extreme care, however, is needed in the design of these studies so that the pre-test diagnostic accuracy of the baseline assessment is not deflated, thereby facilitating the demonstration of the added diagnostic value of the investigational product. In addition and when meaningful, a conventional blinded reader assessment designed to provide supportive evidence of efficacy is recommended.

3.6.5 Test reliability

Inter-reader variability and other sources of unreliability are often important sources of error in the interpretation of diagnostic tests. It is recognised that frequently tests cannot be repeated on individuals,

but inter- and within-reader variability should be investigated. In order to assess inter-reader variability, a reasonable number of readers should be engaged, trained and allocated to evaluate the same test results as part of the development plan. Similarly to evaluate within-reader variability, the same test results should be assessed repeatedly by the same reader. Test reliability may be investigated with and/or without masking of clinical data. If non-inferiority in terms of reliability vs. an established active comparator is to be demonstrated, blinded reader data are expected.

It should be noted that the repeated test interpretations required for estimates of inter-reader variability should not be combined to give consensus opinions of diagnostic efficacy.

3.6.6 Dependency on study specific conditions

The competence of readers charged with interpreting diagnostic data can influence the evaluation of the accuracy of a test whenever assessment of test results involves some subjective element. The trial protocol should therefore explain how readers are trained to interpret the experimental test results.

The conduct of multicentre trials designed to provide on-site test results may also help to assess the diagnostic performance in settings closer to clinical practice.

4 DATA PRESENTATION

4.1 Diagnostic performance

Diagnostic performance or accuracy of a test agent can be presented in terms of sensitivity, specificity, negative and positive predictive value as already discussed.

4.1.1 ROC curves

ROC curves are mainly considered to be a means for selecting appropriate cut-off points to be used prospectively in confirmatory trials. In comparative trials with an active comparator, however, ROC curves may serve as sensitive measures of diagnostic performance converting a bivariate (sensitivity/specificity) to an univariate test variable (area under the 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.

4.1.2 Test 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.), the diagnostic performance of an agent may also be affected by, e.g. stage of disease. ROC curves generated from different populations might therefore 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 value of a diagnostic agent in detecting a disease of interest in any population is dependent on the prevalence of the disease; 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.

4.1.3 Test performance when there is more than one lesion per individual

If more than one lesion can be detected in an individual, overall test performance has to be expressed in relation to an individual rather than lesions detected. Evaluation of sensitivity, specificity and other relevant measurements might 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, however, number and distribution of lesions, as measures of diagnostic performance, may be acceptable if validated by a standard and appropriately handled statistically, even if the clinical relevance of the findings *per se* may be questioned.

5 ASSESSMENT OF CLINICAL USEFULNESS

A marketing authorisation application for a diagnostic agent/test, should address the following issues:

5.1 Technical performance

An assessment of technical performance, e.g. image quality, and/or procedural advantages/disadvantages of a new agent/test and, if applicable, in comparison with a comparator is required.

5.2 Diagnostic performance

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 on predictive value of a test and consequently indications of use should also be discussed, e.g. screening versus diagnostic uses.

5.3 Impact on diagnostic thinking

This refers to quantifying directly or showing indirectly or historically the impact of a test on diagnostic thinking, i.e., the impact of a test outcome on post-test versus pre-test probability of a correct diagnosis in relation to a well-defined clinical context as regards patient characteristics and prior diagnostic procedures.

5.4 Impact on therapeutic decisions and clinical outcome

This refers to a description and quantification of impact of diagnostic information on management of a patient and clinical outcome.

An assessment of potential benefits and risks arising from the impact on therapeutic decisions should be made. Consequences of an incorrect diagnosis (false positive or false negative) must be considered, e.g. a false positive result that leads to unnecessary interventions. As for impact on diagnostic thinking, this level of clinical usefulness may be addressed directly, indirectly or historically.

5.5 Safety

The safety profile should be characterised, using commonly accepted criteria, e.g., the Common Toxicity Criteria. Clinical safety assessments of diagnostic agents should be tailored based on their characteristics and 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 for imaging agents, evaluation of safety of associated test procedures (e.g., radiation exposure) and possible problems associated with incorrect handling of test procedures must be discussed.

5.6 Patient acceptability and test simplicity in relation to comparators

In addition to considerations concerning safety and side effects, patient's comfort and tolerability of invasive procedures have to be considered and methods of assessment stated in trial protocol. 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 new test should also be discussed.

6 REQUIREMENTS FOR AUTHORISATION

The level of directly demonstrated clinical usefulness may be detailed in the SPC. Similarly, the characteristics of the patient populations included in the confirmatory trials and, if appropriate, circumstances for proper use of the diagnostic agent/test, e.g. add-on to other investigations, should be adequately reflected in the labelling.

6.1 Requirements on study data

- Adequate diagnostic performance and reader/technique dependent precision in relation to a standard and/or a comparator.
- Relevant impact on diagnostic thinking in the clinical context in which the test is to be used, unless, as justified from previous studies and detailed in the Clinical Overview, such impact can be shown indirectly or historically. Comparative trials with established comparators may provide useful information also in relation to this issue.
- Acceptable safety profile.

6.1.1 Clinical relevance of improved diagnostic thinking

If it is not clearly obvious that accurate diagnostic/prognostic information *per se* is beneficial, the application for authorisation should provide support that the information is clinically useful, e.g. for monitoring of therapy. If not evaluated in clinical trials, support for these issues could be drawn historically from published scientific evidence and clarified in the Clinical Overview.

6.1.2 Performance in relation to an active comparator

In cases where there are established comparators for the condition in question, the accuracy-precision/risk balance of the test based on the new agent should be at least equivalent to the justified comparator. Advantages for the new method with respect to, e.g., simplicity, safety and patient and health-care convenience could form a basis for concluding superiority and might also compensate for inferior diagnostic performance if justified from the intended use.

APPENDIX 1

Specific considerations for imaging agents

This appendix should be read in conjunction with the general part. Imaging agents refer to both radiopharmaceuticals and contrast agents in this document.

An analysis of the following issues is required in an application:

- Classification of imaging agents
- Diagnostic claims for imaging agents
- Efficacy criteria
- Methodological issues
- Safety assessment

1. CLASSIFICATION OF IMAGING AGENTS

Imaging agents can be classified according to various principles, e.g.:

- Physical properties (e.g., density, viscosity, osmolality)
- Route of administration (e.g., oral, intravenous, intra arterial, intrathecal, intra-rectal, intra-articular, intra-cavitary).
- Pharmacokinetics
- Imaging modality, e.g. X-ray, MRI, nuclear medicine, etc.

A more clinically relevant classification is the *specificity* of an imaging agent and this should not be confused with the statistical term specificity.

(i) S*pecific* agents can enhance one or more anatomical sites as determined by factors such as functions or biological processes, e.g. imaging agents specific for the lymphatic system.

(ii) *Non-specific* agents are widely distributed in the body before being eliminated through the lungs (e.g., gas-filled microbubbles), kidneys (e.g., iodinated contrast agents), or digestive tract.

The anatomical specificity of an imaging agent depends on both the route of administration and the pharmacokinetics. Local administration induces an anatomical specificity at the site of the injection. Different phases of drug pharmacokinetics can result in varying anatomical specificity over time, and thus different enhancement properties. For example, a non-specific iodinated contrast agent administered as an intravenous bolus enhances the blood vessels (angiographic phase) initially, then the parenchyma (e.g., liver, spleen, kidneys) and finally the urinary tract. Similarly, a liver specific nanoparticulate agent distributes in the vasculature (non-specific angiographic phase) before being taken up by the liver or the lymph nodes.

An imaging agent can therefore be used for different imaging purposes. With the rapid development of imaging techniques (CT, MRI and US), new uses of an already approved drug can emerge with new developments in instruments or techniques.

2. DIAGNOSTIC CLAIMS OF IMAGING AGENTS

The diagnostic claims are determined by the characteristics of the agent itself and by the test variables and design of the trials performed to demonstrate efficacy. As for other diagnostic agents, imaging agents, specific as well as non-specific, can be used for different purposes.

Broadly, the diagnostic claims of imaging agents may be grouped as follows:

- Structure delineation - to characterise normal anatomy and to differentiate abnormal structures from normal structures, e.g., oral contrast agent for CT or MRI.
- Detection or assessment of disease - to detect the presence of a disease or to characterise the extent of pathology, e.g., iodinated contrast agents and bone scintigraphy for detection of brain and bone metastases, respectively.
- Functional or metabolic evaluation – to assess functional and/or metabolism, including biological and physiological parameters, e.g., measurement of regional cerebral blood flow.
- Prognostic/therapeutic management guidance – to provide data to guide clinical management, e.g., tumour staging, or measurement of cardiac ejection fraction during anthracycline treatment.

Approval 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, the assessment of biological or physiological processes may form the basis for an approval.

2.1 Specific imaging agents

Their specificity for an anatomical site, a system of function or a biological process makes the diagnostic claim straightforward. The indication will be limited to the target site (e.g., gastrointestinal tract for an oral agent), system of function (e.g., lymph node MRI for superparamagnetic nanoparticles) or biological function (e.g., receptor imaging scintigraphy).

2.2 Non-specific imaging agents

This category includes iodinated contrast agents, non-specific gadolinium chelates, and $H_2O[^{15}O]$. These non-specific contrast agents can have different indications in different parts of the body. Future contrast agents, in particular US contrast agents might also be incorporated into this category.

A claim for application of a non-specific agent to different anatomical sites would be acceptable provided that the experimental agent has been shown to be clinically useful and, when appropriate, non-inferior to an active comparator belonging to the same group and for which the relevant anatomical territories have been studied.

Major systems that should be systematically included in trials for a whole body indication when appropriate are those in which the imaging agent would be expected to exhibit different pharmacokinetic behaviours. These systems include the brain (because of the blood-brain-barrier), liver, kidneys and blood vessels.

3 EVALUATION OF EFFICACY

3.1 Criteria of efficacy

Various levels of efficacy for diagnostic tests have been described in the general part of this document. Additional considerations for imaging agents are detailed below.

3.2 Assessment of efficacy

3.2.1 Technical performance and practicality

Technical performance relates to the technical characteristics of the image and image quality. This evaluation should be performed during the early phase clinical trials and serves as part of the basis for authorisation. Image evaluation should also be done "off site". Parameters for evaluation should be both quantitative (signal to noise ratio, contrast, number of lesions) and qualitative (sharpness of the details, border delineation, opacification, presence of artefacts).

3.2.2 Diagnostic performance

With respect to diagnostic performance, superiority or non-inferiority of the experimental agent has to be justified also in terms of clinical usefulness based on comparison with a standard, a comparator or other evidence (refer main document).

It is recognised that the standard parameters of diagnostic performance might be unsuitable in some cases, e.g. for the assessment of anatomical delineation. Alternative measures, however, are expected to be justified. If two imaging agents belonging to the same class are compared using similar diagnostic techniques, diagnostic confidence might serve as a suitable endpoint.

3.3 Technological dependence

An imaging agent is only useful in combination with an appropriate device designed for the detection of the physical effect of the agent. There is a strong dependence between the efficacy of the medicinal product and the technical equipment used to create the image. 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.

The efficacy of the imaging agent can sometimes depend on or be enhanced by an interaction between the physical process and the agent. For example, new acoustic emission sequences produced by state of the art US machines are designed to destroy the microbubbles of the ultrasound contrast agent, hence multiplying the signal enhancement properties of the contrast agent.

In summary, the imaging device is a key consideration in the design of clinical trials, and the applicant should pay special attention to:

- Technological choices in the development plan.
- Considerations on whether these technological choices are still valid at the time of the marketing authorisation.
- Any other concerns regarding interaction between the agent and the relevant technology.

3.4. Patient selection

In designing early phase clinical trials, the applicant should 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);
- the levels of function of the organ system(s) responsible for the elimination of the diagnostic agent (e.g., the effect 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 the confirmatory trials the choice of the types of patients and the clinical setting should be appropriate to provide data to support the diagnostic claim(s) and formulation of indications for the diagnostic agent.

4 METHODOLOGICAL ISSUES

4.1 Need for placebo

For most imaging agents, the effect of the agent is so obvious on the post contrast images that the value of use of a placebo is very limited. In certain cases where the vehicle might have a contrast effect (e.g., saline in US), a vehicle controlled study may be appropriate to demonstrate that the imaging agent has an effect beyond that of the vehicle. For the assessment of tolerability, administration of placebo followed by a dummy imaging procedure can be useful.

4.2 Bias

It is important to minimise the extent of possible observer bias by determining the true disease-state of subjects using the gold standard independent of the experimental agent. Refer to Section 3.6 in the general part. Possible bias in trial design, conduct and interpretation of results must be critically appraised in the Clinical Overview.

4.2.1 Image evaluation

Training of readers may be based on images obtained from phase I or II trials. Consistency between readers should be measured quantitatively.

The "on-site" evaluation of images in any clinical trials may be biased by lack of blinding to comparator test results and should not be presented as sole proof of efficacy even though this approach mirrors routine clinical practice.

"Off-site" or external evaluation is considered to be the best way to minimise observer bias in the assessment of the efficacy of imaging agents.

These off-site assessment readings should be done by:

- Independent readers (who do not participate in the study at the site of origin of the readings).
- Blinded readers (means that the reader is unaware of the clinical context and the imaging agent used). Readers external to participating centres might also be blinded for inclusion/exclusion criteria for the study, as well as which agent was administered first.
- A representative sample of readers. The reader is an intrinsic part of the diagnostic process in the same way as, e.g., the imaging equipment.

The sequence of external evaluation of the imaging data should be randomised and the evaluation of images should be done in pairs (non-enhanced and enhanced) as well as separately to provide maximal information.

For the assessment of impact on diagnostic thinking, appropriate clinical information including the results of previously performed diagnostic tests needs to be provided. Depending on the agent, and the proposed indication, sequential unblinding may in certain circumstances be appropriate allowing the readers to form a judgement first without and then with prior relevant clinical information.

4.3 Anatomical sites

4.3.1 Non-specific agents

In the case of non-specific agents that can enhance different anatomical sites after intravenous injection, confirmatory trials with an already approved comparator using similar conditions of administration are recommended.

4.3.2 Specific agents

In the case of an agent specific for a system with multiple anatomical sites (e.g., lymph node imaging, arthrography), the choice of the representative anatomical site(s) for the clinical trials should be discussed and justified by the applicant. The extension from one anatomical site to another (e.g., shoulder to knee arthrography) should be discussed based on physiological, anatomical and imaging technique similarities, and might need justification (e.g., lymph node scintigraphy in the thorax versus the abdomen)

5. SAFETY ASSESSMENT

Refer to Section 5.5 in the general part. Specifically, the safety of the imaging procedure itself has to be addressed.

5.1 Radiation exposure when radiopharmaceuticals are used

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 calculations should take into account factors like disease and age and must include the contribution of radionuclidic impurities to the radiation dose, the long-term elimination of the radiopharmaceutical and the radioactive degradation products.

Calculations of absorbed dose to organs should preferably be performed in accordance with the MIRD schedules. If other methods of calculation are used, details must be given with references to the original reports.

5.1.1 Posology and method of administration

The route of administration and the recommended “dosage” for patients based on the dosage findings studies and clinical trials should be presented.

5.1.2 Absorbed dose to organs and whole body

The absorbed dose to the organ receiving the highest radiation exposure must be stated as well as absorbed doses to all organs included in the calculation of the effective dose.. The unit must be milligrays per unit activity administered

(mGy/MBq).

5.1.3 Effective dose

The estimation of the radiation dose must be summarised in terms of the effective dose, using the weighing factors given by ICRP. The units should be millisieverts per unit of activity (mSv/MBq). In very young children a minimum dose, necessary to obtain images of sufficient quality, should be given when appropriate.