HAUTE AUTORITÉ DE SANTÉ

METHODOLOGICAL GUIDE

Companion diagnostic test associated with a targeted therapy: definitions and assessment method

February 2014

Department of Health Technology Assessment

This methodological guide can be downloaded at
www.has-sante.fr

Haute Autorité de Santé (French National Authority for Health)
Documentation Department - Public Information
2, avenue du Stade de France – F 93218 Saint-Denis La Plaine Cedex
Tel.: +33 (0)1 55 93 70 00 – Fax: +33 (0)1 55 93 74 00

This methodological guide was validated by the HAS Board in February 2014.

Contents

Abbreviations and acronyms

ALK..... anaplastic lymphoma kinase

MA ..... marketing authorisation

EGFR.... epidermal growth factor receptor

HER2 .... human epidermal growth factor receptor-2

IHC..... immunohistochemistry / immunohistochemical detection

Mk+..... marker positive

Mk-..... marker negative

TKI ..... tyrosine kinase inhibitor

TKR..... tyrosine kinase receptor

Glossary

Biomarker: "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." (definition of the National Institute of Health, US). See "marker".

Risk factor (marker): factor (marker) associated with an increase in the frequency (commonly called risk) of the endpoint event considered (for example: death, occurrence of a cardiovascular complication) in subjects who are carriers of this factor compared with others.

Predictive factor (marker): factor (marker) influencing the effect of a treatment and whose ability to change patient clinical outcome has been demonstrated. The demonstration of the predictive value of a marker is equivalent to that of the clinical utility for the associated diagnostic test.

Interaction: statistical situation where a third factor changes the effects of another. In this case, a predictive marker changes the effect of the treatment.

Marker: in the context of this document, patient characteristic that may represent an indicator of risk (risk factor) or treatment effect (treatment effect moderator), or allow patient selection regarding the administration of a treatment (predictive marker). Note: the terms biomarker and marker are interchangeable in everyday language.

Personalised medicine: term consisting of a misnomer to be avoided and replaced by the term "stratified" medicine since the validation of the marker/treatment association is based on a standard populational approach used for therapeutic validation. The particularity of this stratified approach consists of the fact that the population of patients is stratified into subpopulations according to the marker believed to be predictive. See term "stratified" (medicine).

"Stratified" medicine: corresponds to a therapeutic approach where the objective is to select patients to whom to administer a treatment in accordance with a predictive marker, so as to only treat the subpopulation who can benefit from the treatment.

Companion (diagnostic) test: diagnostic test permitting the selection only of patients in whom the treatment is likely to provide a benefit from among those diagnosed with a given illness, according to their status for a predictive marker identified by this test. The test is then considered to be a "companion" to the use of the treatment.

Theranostic (or theragnostic): neologism that derives from the contraction of the words "therapeutic" and "diagnostic"; it consists of the use of a diagnostic test, identifying a marker, to guide patient treatment according to their status for this marker (positive or negative status for a binary marker).

Targeted therapy: treatment that has only shown benefits in certain patients identified by a predictive marker determined by a companion test.

Clinical utility: ability to improve the clinical outcome of patients, and to provide an added value in terms of optimising treatment decisions and, as a corollary, therapeutic strategy. Demonstrating the predictive value of a marker is equivalent to showing the clinical utility of its diagnostic test.

Analytic validity: ability of the diagnostic test to perform the measurement of interest with accuracy and reliability. This validity includes analytical sensitivity and specificity, reproducibility, robustness and satisfaction of quality controls.

Clinical validity: ability of the diagnostic test to precisely and reliably predict the clinical phenotype of interest (for example, overall survival or progression-free survival of patients receiving a given treatment). It includes clinical sensitivity and specificity as well as the positive and negative predictive values of the test. The clinical validity also refers to the so-called "diagnostic performance" of the test.

1. General Information

1.1 Foreword

The objective of this document is to explain the theoretical and methodological foundations for assessing a diagnostic test, called a "**companion" test** for selecting, by identification of a **predictive marker**, only the patients able to benefit from a so-called **"targeted" therapy**.

Since the treatment/test association is an inseparable concept, which is why the test is called "companion", the assessment of the diagnostic test has to be synchronous with that of the treatment. This entails direct consequences regarding study methodologies, which can be appropriate or not, to provide the necessary demonstrations to satisfy the level of requirements for conferring the added value of the terms "companion" test and "targeted" therapy. In terms of **requirements**, the major problem resides actually in providing evidence **that** the **marker** identified by the test being assessed is really **predictive** of the efficacy (or toxicity[1]) of the allegedly "targeted" treatment. On this point, it should be emphasised that "**targeted" treatment** is **an improper term in the absence of a satisfactory level of evidence** of the targeting in question which requires the demonstration of the predictive nature of the marker. Without evidence of this predictive value, **the only thing** that can be asserted is that the treatment is **stratified**, i.e. that its efficacy has been studied and demonstrated in a subpopulation (stratum) of patients. To demonstrate the predictive value of a marker means demonstrating the **clinical utility** of the diagnostic test, i.e. the ability of the test to improve the clinical outcome of patients by assisting in decision making in their therapeutic management. This is this level of validation that must be reached for the test to be **called a companion test** and the **treatment** to be called **targeted**.

The reasoning reflected in this document has been developed from the relatively scarce literature available on the subject, as well as and especially from the methodological fundamentals on which the development and assessment of therapies are more broadly based. This approach has led to proposing **requirements for assessing presumed companion tests** consistent with those currently imposed for drugs. The evidence as well as the uncertainties provided by various study designs have also been specified to that effect.

In general, the **main sources of errors** that can lead to incorrectly concluding clinical utility of a diagnostic test are the same as for drugs. These are essentially:

- a *post-hoc* exploratory approach that can lead to considering an artefact found when exploring an abundance of results to be evidence;
- a random statistical error due to sampling fluctuations;
- a bias in the demonstration. The methodological principles of randomisation, double blind and intention-to-treat analysis prevent bias from being introduced into studies;
- a lack of clinical relevance of the results.

Consequently, the assessment must ensure that the results on which the claim of clinical utility of a diagnostic test is based comply with the **following principles**:

- hypothetico-deductive approach (1);
- control of the risk of making a type I error;
- control of the risk of bias by a strict application of the methodological principles for randomised controlled trials (randomisation, double blind, intention to-treat analysis);
- clinical relevance of results: a tangible and evidential improvement of patient outcome must be provided, on appropriate clinical (or clinical and biological) efficacy endpoints.

[1] We limited the scope of this document to markers predictive of treatment efficacy. Yet, the principles described herein may also be applied to markers predictive of side effects.

1.2 Concept of predictive marker

The definition of a "predictive marker" combines **two conditions**.

The **first** is that a predictive marker must **be a moderator of the treatment effect**, i.e. it must predict (in accordance with its “predictive” value) the effect of a given treatment for a given efficacy endpoint (2,3). In statistical terms, this condition corresponds to the existence of a “marker by treatment interaction”, as explained later on in the document.

The second is that selecting the patients to be treated by using a diagnostic test to identify a predictive marker must have demonstrated **clinical utility**. In other words, the predictive marker must allow optimisation of the therapeutic strategy by differentiating between the patients in whom the treatment will have the best efficacy, and those who will have no benefit from it, to whom it is thus pointless to administer it.

Due to recent developments in the field of molecular biology, predictive markers are currently mainly genetic variants. The principal field of development for targeted (stratified) therapies is oncology, where many proposed new candidate markers correspond to genetic variations within proto-oncogenes, leading to changes in protein activities involved in the process of oncogenesis.

Most often, predictive markers are **binary**[1] and define two types of statuses for patients: marker positive (noted marker (+) or Mk+ in the remainder of the document) and marker negative (noted marker (-) or Mk-). **Marker (+)** corresponds to the **modality of interest**, i.e. patients in whom the treatment is effective or who are at risk of a particular toxicity. In the present document, we limited our scope to tests identifying predictive markers for treatment efficacy. However, the principles described herein are completely transposable to validation of predictive markers for adverse effects, marker (+) patients being the ones at the highest risk of adverse effects with the treatment.

1.3 Concept of companion test

A companion test is a diagnostic test permitting **selection of only** patients in whom a treatment is likely to **provide a benefit** among those diagnosed with a given illness, according to their status for a **predictive marker** identified by this test. For this reason, the test is considered a "companion" to the use of the treatment.

As discussed in the foreword to this document, the assessment of this approach can only logically rely on a **joint assessment of the test and the treatment** with two objectives: to show that the treatment is only effective in marker (+) patients and that the companion test improves therapeutic decision making and therefore patient clinical outcome.

It is important to emphasise that the benefit of selecting patients presumed to be the better, or even the only responders to a treatment by a companion test is also to **optimise the clinical efficiency**[2] of this treatment compared with a strategy consisting of using the treatment in all patients. Indeed, the selection of patients on the basis of identification of a marker that is not very or not at all predictive by a diagnostic test, does not optimise clinical efficiency much, if at all. It may cause a loss of opportunity for some patients if the treatment is actually more or less beneficial in everyone. The rest of this document will show that this ethical problem is of major concern in the context of assessing the utility of these tests.

[1] We limited the scope of this document to binary markers. The methodological guide entitled "Companion (diagnostic) test associated with a targeted therapy: scientific appendix" provides some additional information on continuous markers.
[2] Clinical efficiency corresponds to the number of events avoided for a certain number of treated patients.

1.4 Concepts of "stratified" and "targeted" therapies

The therapeutic approach related to the use of predictive markers is sometimes qualified as **personalised medicine** or individualized medicine. These names should be considered misnomers because the approach involving predictive markers remains populational and statistical, rather than individual, and is actually a **stratified medicine** or precision medicine approach (4). It merely refines the estimates by focusing on subpopulations of patients (hence the term stratification) and thus produces an average estimate of the treatment effect in these subpopulations (strata), not an estimate of the treatment effect in an individual.

A targeted therapy is a treatment providing benefits only in certain patients identified by a **predictive marker** detected by a **companion test**.

In practice, it appears that claims that a treatment is targeted may be observed in two situations:

- when the targeting was sought by the developer at the time of treatment design (synthesis): this is called targeting "by construction"
- following exploratory research by a retrospective analysis of the results from one (or more) trial(s) conducted on a population of patients not selected by a marker, followed or not by a prospective confirmatory trial (see section 3.2).

The **targeting "by construction"** approach has historically been the first used. It still is today. Most often, the efficacy of compounds claimed to fall within this concept is only studied in patients carrying a marker designated as a target. The lack of efficacy of the compound in marker (-) patients is considered implicit. However, the predictive value of the marker cannot reasonably rely on implicit foundations, strictly speaking, but only on clinical evidence. Indeed, preclinical data represent only preliminary exploratory data that very roughly reflect the data that will be obtained in humans. In fact, due to the great complexity and peculiarities of the human organism, no type of preclinical study (immortalised cell lines, animal studies, etc.) can precisely and reliably predict the clinical effects of a drug. Moreover, to consider implicit the lack of efficacy of a treatment in marker (-) patients would mean being able to demonstrate an indisputable specificity of the treatment towards its target *in vivo*. Yet, to provide such a demonstration appears very difficult.

As it happens, we now have some experience with the first therapies qualified as targeted. It has revealed **several examples** where, although the treatment had been initially considered to be ***a priori* specific for its target** and the marker to be central in the treatment mechanism of action, the treatment appeared active in a substantial proportion of marker (-) patients. This is in particular the case with trastuzumab in breast cancer, for which efficacy in HER2/neu (human epidermal growth factor receptor 2) negative cancers is envisaged. A second example is cetuximab. This other monoclonal antibody was synthesised to target the EGFR (epidermal growth factor receptor), an oncogenic tyrosine kinase receptor (TKR) detected in 70-85% of colorectal cancers by the immunohistochemical (IHC) method. Yet, currently, the analysis of data from certain trials has shown that up to 25% of patients with metastatic colorectal cancer identified as "EGFR negative" by IHC could respond to cetuximab. Two main explanations were suggested: the poor specificity of the IHC technique and/or an ADCC (antibody dependent cell-mediated cytotoxicity) phenomenon linked to a toxic effect of cetuximab toward tumour cells. Aside from monoclonal antibodies, it is fairly common not to observe a real specificity for a unique target within the other large class of the so-called “targeted” therapies in oncology, represented by the tyrosine kinase inhibitors (TKI),. For example, crizotinib is used as an inhibitor for a TKR called ALK (anaplastic lymphoma kinase) and its oncogenic variants in non small cell lung cancer. However, it also inhibits the receptor for the hepatocyte growth factor (HGFR/c-Met) also involved in oncogenic processes. In its assessment report associated with crizotinib registration, the European Medicine Agency (EMA; EMA/CHMP/497137/2012) clearly stated the only relative specificity of this drug: "*When tested to its selectivity for different kinases, crizotinib was relatively specific to c-Met/HGFR and ALK fusion proteins*". These examples clearly show that "targeting by construction" is not sufficient to justify the absence of clinical study in marker (-) patients. The lack of treatment efficacy in these patients

must be clinically proven to **attest *in vivo* the reality of the targeting**. The risk of wrongly accepting that a new treatment is targeted would indeed directly lead to a loss of opportunity for these patients.

The approach of claiming that a treatment is targeted following a *post-hoc* **exploratory search** for predictive markers is detailed in section 3.2. It is interesting to note that, in a certain way, this avenue of research **has broadened the concept of "targeted" feature of a drug**. Indeed, this concept originally referred, in connection with "targeting by construction", to the synthesis of treatments oriented toward molecular targets involved in their mechanism of action and implying a direct interaction between the drug and its target. However, *post-hoc* searches for predictive markers have led to situations where the initial "physical" target of the drug, supposed to be targeted by construction, and the presumed predictive marker differ. It led to a broadening from the concept of "physical" pharmacological target to the more general concept of "predictive marker". .This extension of the concept poses actually no real concern of consistency since molecular targets, in theory, are also supposed to be predictive markers (although this is not always the case, see section 3.2.3 "Search for a predictive marker after an inconclusive trial"). However, it still must be noted that with the broadening of the concept to predictive markers in general, the comprehension of the **mechanisms of the involvement** of a marker in the efficacy of a given treatment, as well as the overall mechanism of action of the treatment, has become more complex.

2. General principles for validating a diagnostic test determining patient status for a presumed predictive marker

2.1 The three levels of validation for a diagnostic test

The assessment of a diagnostic test determining the status of patients for a presumed predictive marker consists of three levels: the analytic validity, clinical validity and clinical utility of the test (5,6).

- **The analytic validity** of a diagnostic test is its ability to perform *in vitro* the measurement of interest with accuracy and reliability. In other words: does the test really measure what it is supposed to measure, and does it do so correctly? This validation includes studying analytical sensitivity and specificity, reproducibility, robustness and satisfying quality controls.

- **The clinical validity** of a diagnostic test is its ability to accurately and reliably predict the clinical phenotype of interest (for example: overall survival or progression-free survival of patients receiving a given treatment): is there a relationship and what is it, between the results of the test and the phenotype of interest? Clinical validity includes clinical sensitivity and specificity, as well as the positive and negative predictive values of the test. These parameters may be combined under the name "diagnostic performance" of the test.

- **The clinical utility** of a diagnostic test is its ability to improve patient clinical outcome, and to provide an added value in terms of optimising treatment decisions and, as a corollary, therapeutic strategy.

The study of analytical validity and clinical validity are general concepts for diagnostic tests, whether or not a marker is involved. They are beyond the scope of this document, which is dedicated to tests that identify predictive markers.

The **demonstration of clinical utility** is an **indispensable element** in the field of companion testing. This demonstration is needed because it attests to the **added value provided by the test assessed**, compared with a diagnostic test only permitting selecting a subpopulation by identification of a marker. The very concept of companion test implies that its clinical utility has been demonstrated since it must allow avoiding the pointless treatment of patients in whom the treatment would be ineffective, and conversely, not lead to a loss of opportunity by identification of a marker that is not predictive or not sufficiently predictive, thereby excluding patients from a treatment that could actually benefit them.

2.2 The three conditions absolutely required to demonstrate the clinical utility of a diagnostic test

Demonstrating the clinical utility of a diagnostic test involves **demonstrating the following three elements** (7–11):

- the marker identified by the test must change the effect of the treatment (existence of a "marker by treatment" interaction);
- the treatment must be effective in marker (+) patients;
- the treatment must have no clinical benefit in marker (-) patients.

2.2.1 Existence of a marker by treatment interaction

Treatment effects are most often quantified by relative risks. In therapeutic trials, the relative risk is the factor by which the treatment increases the risk, that is to say the frequency of the study

endpoint for patients receiving it compared with that of control patients. It is calculated as a risk ratio. A clinical trial provides estimates of the true relative risks associated with confidence intervals.

There is a "marker by treatment" interaction when the value of this relative risk is influenced by the marker. However, the values of the observed treatment effect can be different without this being a reality; it may be just an artefact due to random fluctuations. It is therefore only possible to talk about interaction if, in the case of a binary marker, a statistically significant difference of the effect of treatment is observed between the subgroup of marker (+) patients and that of marker (-) patients.

Moreover, it is important to note that if a statistically significant interaction test attests a difference of treatment effect in the marker (+) and marker (-) subgroups, this does not mean that the treatment has no effect in marker (-) patients. In fact, the treatment may be only less effective in marker (-) patients. The approached concept is that of "pseudo-targeting" discussed in the methodological guide attached[3] to this one.

2.2.2 Efficacy of the treatment in marker (+) patients

The trial should logically demonstrate the efficacy of the treatment in marker (+) patients, by following the **same criteria as for a conventional therapeutic trial**: prospective trial, free from bias, statistically significant, clinically relevant and having a good external consistency (biological plausibility and consistency among available studies).

A specific aspect relative to a conventional trial concerning demonstrating the efficacy of a new treatment is alpha risk control. Indeed, latent multiplicity of statistical tests is inherent in this trial design: testing the treatment effect in marker (+), marker (-) and/or all patients. Each time, a 5% risk of yielding false conclusive results exists in the absence of precaution. **Alpha risk inflation** can be controlled in several ways. The first is a hierarchical sequential analysis where marker (+) patients are tested first, then all patients. The second is to pre-specify explicitly that the co-primary objectives of the trial are to validate the interaction and efficacy of treatment in marker (+) patients by concentrating the alpha risk on the analysis relating to these co-primary objectives. However, the analyses in all patients and marker (-) patients will raise the problem of multiple statistical comparisons. Finally, a third possibility is to adjust the alpha risk taking into account the multiple potential comparisons using a Bonferroni correction, or the like.

2.2.3 Lack of treatment efficacy in marker (-) patients

The concept of predictive marker also implies a lack of treatment effect in marker (-) patients.

A statistical demonstration of the lack of effect in this subgroup must be provided. This demonstration must make use of a non-inferiority study based on the confidence interval of the relative risk in these patients. It must show that it is possible to exclude with a 97.5 % degree of certainty that the effect of the treatment could be greater than a certain value corresponding to the minimal clinically useful value. In practice, it is actually a **non-superiority reasoning** (with a clinical limit of non-superiority) mirroring that of a non-inferiority reasoning. Non-superiority may be concluded when the lower limit of the confidence interval for treatment effect in the stratum of marker (-) patients is greater than the non-superiority limit. In fact, the objective is to determine if it is reasonable to rule out the possibility that the treatment would be effective (i.e. that it would have an effect substantially superior to placebo) in marker (-) patients. The **limit** of non-superiority must be **set *a priori*** in the protocol and justified clinically. A narrow confidence interval obtained in marker (-) patients is more favourable for obtaining this demonstration, which implies that the specifically required number of subjects had been calculated *a priori* in order to guarantee this precision.

[3] Methodological guide entitled "Companion test associated with a targeted therapy: Scientific appendix".

3. Main clinical study designs: critical analysis

This section sets out HAS (French National Authority for Health) principles for assessing a diagnostic test identifying a presumed predictive marker associated with a stratified therapy. A pragmatic synthesis of these principles is presented herein, but the concepts implemented are not detailed. These are presented and explained in the attached methodological guide entitled "Companion diagnostic test associated with a targeted therapy: Scientific appendix".

The fundamental principle for assessment of this type of test by HAS resides in the fact that the recognition of the terms "companion test" and "targeted therapy" is based on whether or not the clinical utility of the test has been demonstrated. To that end, in this section, **the evidence required to demonstrate the clinical utility of a diagnostic test associated with a stratified therapy is presented**. This demonstration simultaneously proves the predictive value of the marker relative to treatment efficacy, the targeting of the therapy and the designation of companion test.

The clinical utility of a diagnostic test is sometimes claimed from arguments or data coming from studies that do not permit demonstrating it, but which are more easily obtained than rigorous evidence. In this case, recognising or not recognising this claim **depends directly on the design used for the study** supporting this request. It is moreover important to emphasise that a rigorous methodological framework is necessary. The standards are expected to meet the current assessment standards for therapies given the direct involvement of these companion tests in the decision of whether or not to treat patients.

To be able to guarantee the clinical utility of a companion test and concomitantly the targeting of the therapy is an important **healthcare issue** since there are many risks in the absence of this assessment:

- the risk of wrongly recommending a treatment that is actually ineffective (biased result or type I error/alpha risk);
- the risk of wrongly recommending a treatment that has in reality no clinical benefit (lack of clinical relevance).

These two risks are identical to those that are sought to be ruled out in the analysis of a conventional therapy dossier.

- the risk of wrongly recommending a stratified strategy involving performing an additional useless diagnostic test in all patients;
- the risk of wrongly not using in marker (-) patients a treatment that could have some clinical benefit for them, thereby resulting in a loss of opportunity.

Ultimately, not ensuring the clinical utility of a test identifying the subpopulations concerned by a stratified treatment may lead to recommending a discriminatory therapeutic strategy, not optimal in terms of population benefit.

3.1 The "marker by treatment interaction" study design: the only design to demonstrate the clinical utility of a test identifying a presumed predictive marker, and the targeted nature of a therapy

► Characteristics of the approach

A single study design, called "marker by treatment interaction design", allows determining concomitantly the effect of the treatment in marker (+) and marker (-) patients, and as a corollary, permits the existence of a “marker by treatment” interaction to be tested. This design is the only one that can **simultaneously demonstrate the three conditions** validating **the clinical utility** of

a diagnostic test identifying a presumed predictive marker (2,3,12–14). As a corollary, it is also the only one that can demonstrate that a treatment can be actually qualified as targeted, and not merely stratified.

Inasmuch as only this study design can demonstrate the clinical utility of a test, it leads to results with the highest level of evidence and represents **the gold standard** in its field.

This approach **can be envisioned in the majority of situations**:

- before any demonstration of a benefit of the treatment. A joint development of the treatment and the test must be logically envisaged. The focus is on the treatment and the purpose of the dossier is to claim the benefit of the treatment in the subpopulation selected for the marker;
- after a *post-hoc* search in subgroups for a predictive marker following an inconclusive trial on the benefit of a new treatment. These exploratory *post-hoc* results must be prospectively confirmed. For that purpose, an interaction design is appropriate.

However, confirming the presumed predictive value of a candidate marker identified in a *post hoc* search in subgroups following a study with conclusive results cannot be envisaged for ethical reasons. Indeed, it is not reasonable to envisage randomising again patients *versus* the comparator of the first trial after the new treatment has demonstrated greater efficacy in all patients independently of their status for the marker (see section 3.2.3).

The **ethical issues** raised by the investigation of a potential lack of efficacy of the treatment in marker (-) patients are similar to those raised by randomisation in clinical trials in general**.** The justification of this randomisation from an ethical point of view relies on the **principle of equipoise**: conducting a randomised clinical study is justified when the medical community cannot determine, according to the fundamentals of evidence-based medicine, the best option between two treatments. Randomisation is only acceptable for as long as this uncertainty remains (15–17). Marker (-) patients should not be exposed to adverse effects without any hope of being able to benefit from the treatment they receive. Therefore, it must be kept in mind, in the context of stratified therapies, that the **lack of efficacy** of the new treatment in marker (-) patients is not **a result known *a priori***. This is a necessary condition for demonstrating the targeted nature of the therapy. **The hypothesis** tested is **the non-superiority of the new treatment to the comparator**, placebo or current gold standard treatment and, like any hypothesis tested by a clinical study, it will ultimately be rejected or not. This is why it is recommended to provide, in the statistical analysis plan, for the analysis of the treatment efficacy in all patients (Mk+ and Mk-) in addition to the subpopulation analysis. This position is reinforced by the fact that **there are currently counter-examples** among the treatments that were *a priori* claimed to be "targeted by construction" and for which the lack of efficacy in marker (-) patients has been challenged. The lack of efficacy of a treatment in marker (-) patients, which would be predicted by a marker identified by *post hoc* analyses, is also very unreasonable, given the exploratory nature of the research and the purely theoretical nature of the **mechanism of involvement** of the marker in the efficacy of the treatment considered (see section 1.4). In the end, it appears that not conducting this research would generally pose a risk to marker (-) patients by promoting **wrongly accepting markers that are not predictive or not sufficiently predictive** which might then cause a loss of opportunity and deprive them of a treatment that would be effective for them.

► Study design

This design consists of a **double randomisation**, one in each of the subpopulations identified by its positive or negative status for the marker. This approach allows comparing the new treatment N with a comparator C for both marker (+) and marker (-) patients since two estimates of the effect of N *versus* C are produced in marker (+) and (-) patients (see Figure 1). Comparing them allows searching for a statistically significant difference indicating, if applicable, the existence of a marker by treatment interaction.

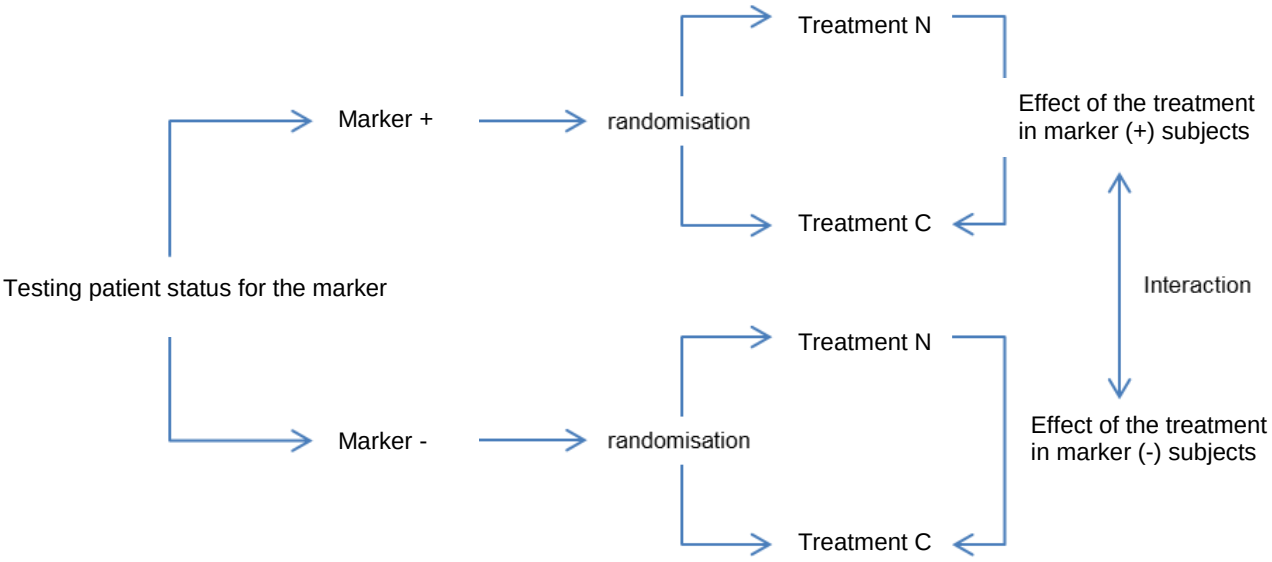

Figure 1 – "Marker by treatment interaction" study design

3.2 The other main study designs: determination of treatment effect and exploratory identification of presumed predictive markers

In practice, the types of methodologies used in the majority of studies were and are still focused on demonstrating the efficacy of new treatments but are not suited to demonstrating the clinical utility of the associated diagnostic tests.

3.2.1 The “targeted” study design

► Characteristics of the approach

The "targeted" study design formed the basis of the first registrations of treatments having prescription limited to a subpopulation identified ("targeted") by the presence of a marker presumed to predict treatment efficacy. It is still commonly used.

This design is completely appropriate for demonstrating the efficacy of a novel treatment in a subpopulation of patients defined by the presence of a marker detected by a diagnostic test. However, as it does not demonstrate **explicitly** the lack of effect in marker (-) patients, it does not validate the clinical utility of the test in question, nor allow claiming the targeted nature of the treatment. (18–20). Indeed, for the reasons explained in section 1.4, it is not reasonable to consider the lack of efficacy in marker (-) patients as **implicit** on the basis of preclinical data and/or a theoretical biological rationale.

When the efficacy of the new treatment has been demonstrated with this design in the subgroup selected by the test, and therefore in marker (+) patients, the **marketing authorisation in the indication** studied is logically **restricted to this subgroup** of patients in whom treatment efficacy has been assessed. In order to comply with the marketing authorisation, the diagnostic test becomes required for the treatment prescription in order to differentiate between marker (+) and marker (-) patients. Indeed, prescribing the therapy to a marker (-) patient would represent an off-label prescription. It must be emphasised on this point that the prescription restriction of the marketing authorisation should not be considered a proof of ineffectiveness of the treatment in marker (-) patients, but only the result of the lack of assessment of this efficacy. In the absence of clinical evidence, the **principal risk** of this design is consequently to cause a **loss of opportunity**

for marker (-) patients by depriving them of a treatment that could benefit them, in addition to performing a pointless test.

Ultimately, this design is **not suitable for demonstrating the clinical utility** of a diagnostic test associated with a stratified therapy. Nevertheless, case by case, the level of conviction accorded to the preclinical and early clinical elements during the assessment may support to some extent this demonstration. This design would basically be useful when the preclinical and/or early clinical data lead to concerns of a particular toxicity of the treatment in marker (-) patients. However, the clinical utility of the test will not be demonstrated. It may therefore be preferable, when possible, to choose a marker by treatment interaction design combined with a protocol ensuring patients the degree of safety necessary for any therapeutic trial.

► Study design

The targeted design, also called “enrichment” design, consists of assessing the new treatment N *versus* a comparator C only in marker (+) subjects (see Figure 2).

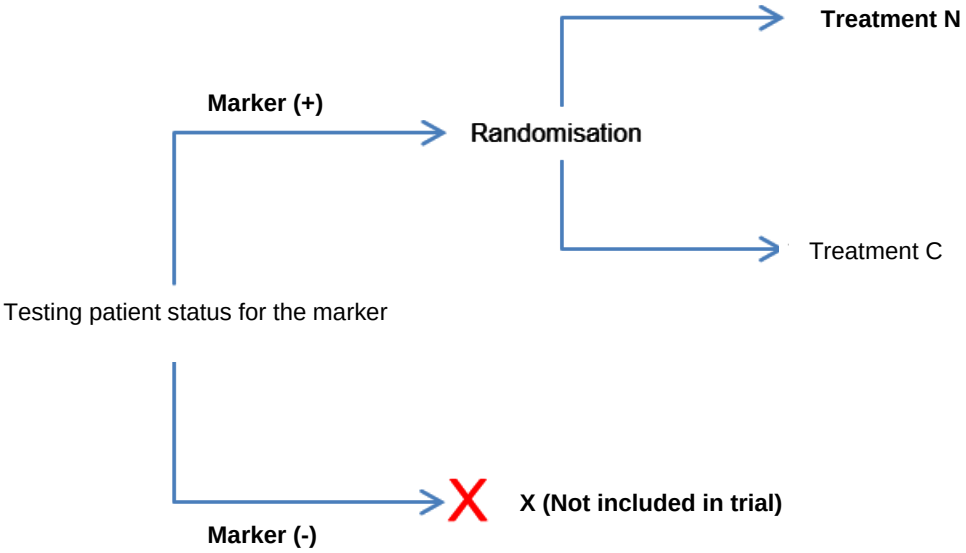

Figure 2 – Targeted study design

3.2.2 The "marker-based strategy” study design

► Characteristics of the approach

The "marker-based strategy" study design consists of a strategy comparison. In fact, the strategy of administering a new treatment on the basis of the patient marker status identified by the diagnostic test is compared with a "standard" strategy where the standard treatment is administered without determination of the marker (see Figure 3).

As with the targeted design, this design does not measure the effect of the new treatment in marker (-) patients. Therefore it does not demonstrate the clinical utility of the diagnostic test or, as a corollary, the targeted nature of the new treatment.

In reality, despite its name, this design does not allow demonstrating in the strict sense the superiority of a stratified "marker-based treatment" strategy - where the choice of the treatment depends on the result of the diagnostic test - in comparison with a standard non-stratified strategy.

Indeed, a better result with the "marker-based treatment" strategy assessed relative to the standard non-stratified strategy can be observed with this design if the new treatment is more effective than the standard treatment in marker (+) patients, but also if the new treatment is more effective than the standard treatment in all patients, regardless of whether they are marker (+) or (-); in other words, whether or not the test is clinically useful. Consequently, a study using this design and having conclusive results only permits drawing conclusions about the superiority of the overall strategy based on the use of the marker and the new treatment, relative to the use of the standard treatment in all patients. Yet it does not tell us whether or not, in the context of the marker-based strategy, patients should be selected to receive the new treatment depending on the presence of the marker.

Ultimately, since this design does not ensure that the treatment has no benefit in marker (-) patients, its main associated risk is, like for the targeted design, to wrongly exclude these patients from the treatment benefits, therefore causing a loss of opportunity for them in addition to administering a pointless test.

► Study design

In the "marker-based strategy" design ,the standard strategy of treating all patients by the standard treatment is compared with the strategy consisting of treating marker (+) patients with the new treatment and marker (-) patients with the standard treatment (3). The (theoretical) objective is to show that the new strategy using the assessment of the marker permits obtaining better results than the standard strategy. Therefore, the results obtained with the marker-based strategy (with grouping of the patients treated by the new and standard treatments) are compared with those of the standard strategy.

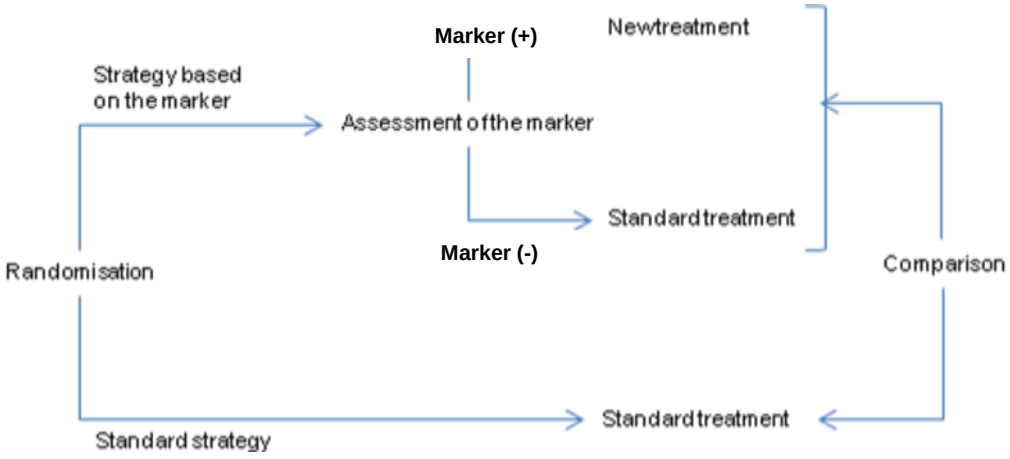

Figure 3 – "Marker-based strategy" study design

3.2.3 *Post-hoc* subgroup analysis

► Subgroup analysis: only exploratory results

Currently, in the literature, it can be observed that the majority of studies conclude the identification of predictive markers from the results of *post-hoc* (also called *a posteriori* or retrospective) subgroup analyses. This type of approach is not methodologically satisfactory, since evidence of the clinical utility of a diagnostic test has to be provided by a hypothetico-deductive approach, as for the demonstration of therapeutic efficacy.

Thus, *post-hoc* subgroup analyses are not appropriate for seeking predictive markers. They are limited to providing exploratory results because of the inflation of the alpha risk they generate, even if these markers were *a priori* listed. This approach may be useful or even necessary for identifying candidate markers of interest. However, the results will only lead to a generation of hypotheses that must be validated by *ad-hoc* (prospective) confirmation studies. To this end, the only design that permits demonstrating the clinical utility of a test is the marker by treatment interaction design as previously explained (see section 3.1).

► *Post-hoc* search for a predictive marker after a conclusive trial

This situation may arise when a *post-hoc* search for potential determinants of the response after a conclusive trial in a non-selected population of patients leads to suspecting that the treatment would in fact really be effective only in certain patients who carry a defined marker.

This situation presents the methodological problems of *post-hoc* approaches, in particular statistical multiplicity and may result in inflation of the beta risk of wrongly not showing evidence of treatment efficacy in marker (-) patients. These patients could then wrongly be excluded from the treatment while they were initially included in the trial that had demonstrated the treatment efficacy without selecting patients according to their status for this marker. An *ad-hoc* trial is therefore needed to confirm these results.

However, an **ethical question** may be raised in this context. Indeed, the initial trial led to conclusive results while conducted according to a standard non-stratified study design. Consequently, the new treatment became the gold standard treatment in the area concerned. To randomise again patients between the new treatment and a placebo (or the previous gold standard treatment) poses thus an ethical problem. A **marker by treatment interaction design cannot** therefore be **recommended**, especially with regard to marker (+) patients who cannot reasonably be randomised again *versus* the initial comparator since the purpose of the confirmatory study would be to show that the treatment is only effective in them. Finally, the search for a predictive marker following a conclusive trial in the general population most often yields exploratory results not confirmed prospectively and therefore a level of evidence unacceptable for demonstrating the clinical utility of the diagnostic test associated with the marker.

However, there can be situations where a **strong rationale** may appear in the sense of a probable lack of treatment efficacy in patients who do not carry a marker, thus ethically justifying conducting a prospective study to confirm (or refute) this presumed lack of efficacy in these patients. In this case, the approach to follow is that of a **prospective non-superiority study**. This study may be designed as a targeted study only including marker (-) patients, or as a "marker-based strategy" study, with a "standard strategy" arm corresponding to the use of the new treatment without identification of the marker. The trial must have a high metrological quality, since the major risk is to wrongly exclude the subpopulation of marker (-) patients from the treatment. If the lack of a clinical benefit from the treatment in marker (-) patients is confirmed, it is possible to conclude the clinical utility of the test, thus providing, in accordance with the definition of clinical utility, optimisation of patient management, particularly by avoiding pointlessly exposing patients to potential adverse effects without any benefit and allowing a potentially more effective therapy, if there is one, to be used instead.

► *Post-hoc* search for a predictive marker after an inconclusive trial

Searching for a predictive marker after obtaining inconclusive results in a trial assessing a new treatment poses substantial methodological problems. It particularly carries a **high risk of artefactual discovery of non-predictive markers**, which may wrongly lead to depriving certain patients of the benefit of a new treatment and complicating the management of all patients by conducting a pointless test.

Indeed, in the context of an inconclusive trial, searching for a predictive marker generally results directly from the achievement of this unsatisfactory result, which leads to proposing the hypothesis that the efficacy is not ubiquitous, but limited to a certain type of patient. This search is then undertaken by a retrospective data analysis, which had not been initially planned. The same data will also be used as a demonstration of the predictive value of the marker, creating a tautological situation. Moreover, following inconclusive results, a *post-hoc* approach generates multiple comparisons, leading to alpha risk inflation on the hypothesis of treatment efficacy, thus providing a second chance of wrongly concluding this efficacy.

Consequently, as with all subgroup analyses, the results obtained in this situation must be considered exploratory and require validation by a **prospective confirmation** study. In this case, the demonstration of the predictive value of a presumed predictive marker requires the use of a **marker by treatment interaction design**.

The isolated demonstration of efficacy of the new treatment in marker (+) patients using another study design would not rule out the possibility that the first trial was wrongly inconclusive in all patients because of a lack of statistical power, and then followed by a second trial performed only in marker (+) patients and conclusive only because of better power, whether the marker is predictive or not.

► Search for markers limited to the treated group (treatment-only analysis)

It is not rare to encounter therapeutic trials where the search for predictive markers **resembles an exploration of risk factors** for the event corresponding to the endpoint in the group receiving the new treatment. This approach of not including patients from the control group is called treatment-only analysis. It actually seeks a risk factor for an event rather than evidence of a change in the treatment effect (19). It therefore never provides evidence of predictive markers and carries the risk of false identification of a specific therapeutic effect in a subpopulation of marker (+) patients.

3.2.4 Summary table

The table below presents the summary of the demonstrations that can be provided by the development study designs for a diagnostic test associated with a stratified therapy. This table specifies the level of evidence associated with each design for the test and the treatment to be called a companion test and a targeted therapy, respectively.

The “companion” term brings added value. It does not preclude the possibility of including a diagnostic test for reimbursement in circumstances where the test would not be clinically useful for the purposes of this guide.

Table 1: Demonstrations and level of evidence provided by the development study designs for diagnostic test/stratified therapy associations regarding the terms "companion" test and "targeted" therapy.

Development design selected for the diagnostic test	Potential demonstration(s)	Level of evidence considered for the designations of companion test and targeted therapy
Standard study with *post-hoc* search for "predictive markers" including only patients of the treated group	Search for risk factors for the endpoint rather than predictive markers for therapeutic response	**No conclusion possible**
Post-hoc search in subgroups for a predictive marker in a trial studying the efficacy of a new treatment in the general population	- Efficacy or lack of efficacy of the treatment in a population of patients not selected by a marker - Exploratory search for a predictive marker; the potential clinical utility for a diagnostic test associated with a presumed predictive marker would have to be confirmed in a prospective study	**No conclusion possible** (in the absence of a prospective confirmation study)
"Marker-based strategy" study	- Efficacy of the new treatment in marker (+) patients, or in all patients regardless of their status for the marker, greater than that of the standard treatment in all patients regardless of their status for the marker - Superiority of the strategy based on the marker identified by the diagnostic test assessed and the administration of the new treatment to Mk+ patients, compared with the strategy of using the standard treatment without determination of the marker	**No conclusion possible**
"Targeted" study in marker (+) patients	Efficacy of the new treatment in the subpopulation of patients identified as marker (+) by the diagnostic test assessed	**Low to moderate level of evidence,** depending on the weight given to preclinical and early clinical elements during the assessment
"Marker by treatment interaction" study	- Validation of the clinical utility of the diagnostic test assessed, including the demonstration of the new treatment efficacy in the subpopulation of marker (+) patients. - Validation of the terms companion test and targeted therapy	**High level of evidence**

4. Economic assessment

The medico-economic assessment of a dossier claiming "a companion test associated with a targeted therapy" does not currently require developing a specific methodological guide. While the issue of joint assessment of a test and a therapy is a novelty in the pharmaceutical and biotechnological markets, it is widely acknowledged that the assessment of a diagnostic test or a screening test would not be limited to assessing its ability to identify the presence or absence of a disease or a risk factor. In fact, this assessment must also assess the clinical utility of the test, in other words, its impact on individual health, this impact being associated with the implementation of healthcare procedures. This type of assessment represents a standard for the economic assessment of a diagnostic test or a screening test. Thus, for the joint test/treatment medico-economic assessment in the context of a claim for "a companion test associated with a targeted therapy", **the recommendations of the current version of the HAS methodological guide "Methodological choices for economic assessment at HAS" remain applicable. Nevertheless, updates** relating to the specific features of so-called "companion" diagnostic tests **are conceivable**. These concern **in particular**:

- the formulation of the issue of **joint efficiency** of the companion test and the targeted therapy;
- the types of diagnostic tests (for example, depending on the predictive nature of the marker oriented to the efficacy or toxicity of the treatment);
- formative choices for economic modelling: determination of the target population, choice of comparators;
- the scope of costs associated with so-called "companion" tests[4].

HAS believes that the **demonstration of clinical utility of the diagnostic test is a prerequisite to the joint assessment of the companion test and an associated targeted therapy** (see Table 1). In other words, an economic assessment labelled "Joint analysis of a companion test associated with a targeted therapy" corresponds to a medico-economic study founded on data of a level of evidence sufficient to show clinical utility.

[4] The joint analysis of a companion test associated with a targeted therapy involves considering all the costs for conducting tests (regardless of the result).

Conclusion

The concept of **stratified medicine** consists of a still recent therapeutic approach relying on **"drug/diagnostic test"** associations. Within this approach, the diagnostic test has the theoretical objective of **selecting** only patients in whom the treatment is likely to provide a benefit and thus limit the use of this treatment to these patients. This selection is based on the patient status for a presumed predictive marker detected by the diagnostic test.

The concepts of **targeted therapy** and **companion test** fall within the scope of stratified medicine but are much stricter and therefore more demanding in terms of evidence required. Thus, to claim that a stratified therapy may be classified as targeted or that a diagnostic test may be designated as a companion test, it must be demonstrated that in addition to being effective in patients carrying the marker, the treatment is not effective in patients who do not carry it. These demonstrations are also the foundations for the validation of the **clinical utility of the test**, that is to say its ability to improve patient clinical outcome and provide added value in terms of optimising therapeutic management. The use of the treatment and performance of the test are inseparable in this context. The clinical utility of the test can only be assessed in conjunction with the efficacy of the treatment in the same trial. Only the **"marker by treatment interaction" design** provides these demonstrations simultaneously. Conversely, study designs determining the effect of a new treatment in a selected subpopulation of marker (+) patients only allow concluding the treatment efficacy (or lack thereof) in this subpopulation. They do not demonstrate the clinical utility of the test.

With regard to HAS, when a diagnostic test allows screening patients in order to meet the **requirements formulated in the marketing authorisation** of a treatment, the assessment may, depending on the type of demonstrations provided, lead to **two types of general conclusions:**

- if the **clinical utility of the test is not demonstrated**:
  - the test will be considered a **"conventional" diagnostic test** to identify the subpopulation of patients whose selection is reported in the marketing authorisation. Although without demonstrated clinical utility, conducting this test will not be less essential from a regulatory point of view to allow the use of treatment under its marketing authorisation,
  - the qualification "companion test" cannot be claimed, and the therapy will not be considered targeted. In fact, since the predictive value of the marker is not proven, the possibility of efficacy in all patients cannot be ruled out;
- if the **clinical utility of the test is demonstrated:**
  - the test will be considered a diagnostic test for identifying the patients whose selection is reported in the marketing authorisation,
  - the test may in addition be qualified as a **"companion test"** and the **therapy** considered to be **"targeted".**

In all cases, the diagnostic performances of the test must ensure the reliability of the results obtained from the studies conducted. Its implementation must provide quality controls and comply with all current quality standards in the field concerned.

References

1. Sargent D, Allegra C. Issues in clinical trial design for tumor marker studies. Semin Oncol. juin 2002;29(3):222‑230.

2. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points--the challenge of statistical validation. Nat Rev Clin Oncol. juin 2010;7(6):309‑317.

3. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol Off J Am Soc Clin Oncol. 20 mars 2005;23(9):2020‑2027.

4. Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. Nat Rev Drug Discov. avr 2007;6(4):287‑293.

5. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med Off J Am Coll Med Genet. janv 2009;11(1):3‑14.

6. Febbo PG, Ladanyi M, Aldape KD, De Marzo AM, Hammond ME, Hayes DF, et al. NCCN Task Force report: Evaluating the clinical utility of tumor markers in oncology. J Natl Compr Cancer Netw JNCCN. nov 2011;9 Suppl 5:S1‑32; quiz S33.

7. Mandrekar SJ, Grothey A, Goetz MP, Sargent DJ. Clinical trial designs for prospective validation of biomarkers. Am J Pharmacogenomics Genomics-Relat Res Drug Dev Clin Pract. 2005;5(5):317‑325.

8. Mandrekar SJ, Sargent DJ. Design of clinical trials for biomarker research in oncology. Clin Investig. déc 2011;1(12):1629‑1636.

9. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. J Clin Oncol Off J Am Soc Clin Oncol. 20 août 2009;27(24):4027‑4034.

10. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: one size does not fit all. J Biopharm Stat. 2009;19(3):530‑542.

11. Mandrekar SJ, Sargent DJ. Predictive biomarker validation in practice: lessons from real trials. Clin Trials Lond Engl. oct 2010;7(5):567‑573.

12. Buyse M, Michiels S, Sargent DJ, Grothey A, Matheson A, de Gramont A. Integrating biomarkers in clinical trials. Expert Rev Mol Diagn. mars 2011;11(2):171‑182.

13. Lai TL, Lavori PW, Shih M-CI, Sikic BI. Clinical trial designs for testing biomarker-based personalized therapies. Clin Trials Lond Engl. avr 2012;9(2):141‑154.

14. Ma BBY, Britten CD, Siu LL. Clinical trial designs for targeted agents. Hematol Oncol Clin North Am. oct 2002;16(5):1287‑1305.

15. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. 16 juill 1987;317(3):141‑145.

16. Gifford F. Community-equipoise and the ethics of randomized clinical trials. Bioethics. avr 1995;9(2):127‑148.

17. Ferté C, Hollebecque A, Salleron J, Penel N. [Clinical equipoise and randomised clinical trials in oncology]. Bull Cancer (Paris). juin 2009;96(6):727‑731.

18. Food and Drug Administration. Guidance for Industry - Enrichment Strategies for Clinical Trials to Support Approval of Human Drug and Biological Products, dec 2012 [cité 8 août 2013]. Disponible sur: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM332181.pdf

19. Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. Ann Intern Med. 15 févr 2011;154(4):253‑259.

20. European Medicines Agency. Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection. EMA/446337/2011, june 2011 [cité 8 août 2013]. Disponible sur: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/07/WC500108672.pdf

Appendix 1. List of tables

Appendix 2. List of figures

Description sheet

Title	Description
Working method	HAS Methodological Guide
On-line publication date	April 2014
Paper publication date	Only available in electronic format at www.has-sante.fr
Objective(s)	To present a pragmatic synthesis of the theoretical and methodological basis for assessment by HAS of a diagnostic test to identify a presumed predictive marker associated with a stratified therapy. To define and present the demonstrations expected by HAS for recognising the use of the terms "companion" test and "targeted" therapy". To define and provide methodological tools to demonstrate the clinical utility of a diagnostic test associated with a stratified therapy. To inform the level of evidence accorded by HAS to the study designs used for the development of the diagnostic tests concerned regarding the demonstration of clinical utility.
Requested by	Internal HAS initiative
Sponsor	Haute Autorité de Santé (French National Authority for Health, HAS), service évaluation des actes professionnels (Department of Health Technology Assessment, SEAP)
Project steering	SEAP Coordination: Carole GIRAUD, Project Manager, SEAP (Department Head: Michèle MORIN-SURROCA, Deputy Department Head: Marc GUERRIER) Secretary: Christine MAYOL, Assistant, SEAP External Coordination: Michel CUCHERAT, Consultant
Document search	Carole GIRAUD, Project Manager, SEAP Michel CUCHERAT, External Consultant
Authors of the rationale	Carole GIRAUD, Project Manager, SEAP, under the responsibility of Marc GUERRIER, Deputy Department Head, SEAP Clémence THEBAUT, Project Manager, Department of Public Health and Economic Assessment (SEESP) and Salah GHABRI, Project Manager, SEESP, under the responsibility of Catherine RUMEAU-PICHON, Department Head, SEESP Michel CUCHERAT, Consultant
Consultation - Validation	Interim version of the document submitted for opinion (October - November 2013): - Internal: Commission d'évaluation économique et santé publique (CEESP, [Committee for Public Health and Economic Assessment]), Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNEDiMTS) [National Committee for Medical Devices and Health Technologies assessment], Commission de Transparence (Transparency Committee) - External: INCa (National Cancer Institute), ANSM (The French National Agency for Medicines and Health Products Safety), DGOS (General Directorate for Healthcare Provision), DGS (General Directorate for Health), DSS (Directorate for Social Security), SFPT (French Society of Pharmacology and Therapeutics), SNITEM (National Association of Medical Technologies Industry), LEEM (French Pharmaceutical Companies Association), SIDIV (In Vitro Diagnostics Industry Union), CNAM (National Health Insurance Fund), UNCAM (National Union of Health Insurance Funds), CEPS (Economic Committee for Healthcare Products) Validation by the HAS Board: February 2014
Other formats	The only format is the electronic format available at www.has-sante.fr
Supporting documents	Methodological appendix guide entitled "Companion test associated with a targeted therapy: Scientific appendix" available at www.has-sante.fr