



HAUTE AUTORITÉ DE SANTÉ

GUIDE METHODOLOGIQUE

Test compagnon associé à une thérapie ciblée : définitions et méthode d'évaluation

Février 2014

Ce guide méthodologique est téléchargeable sur
www.has-sante.fr

Haute Autorité de santé

Service documentation – Information des publics
2, avenue du Stade de France – F 93218 Saint-Denis La Plaine Cedex
Tél. : +33 (0)1 55 93 70 00 – Fax : +33 (0)1 55 93 74 00

Sommaire

Abréviations et acronymes	4
Glossaire	5
1. Généralités	6
1.1 Avant-propos	6
1.2 Notion de marqueur prédictif	7
1.3 Notion de test compagnon	7
1.4 Notions de thérapeutiques « stratifiée » et « ciblée »	8
2. Principes généraux de validation d'un test diagnostique déterminant le statut des patients pour un marqueur présumé prédictif	10
2.1 Les trois niveaux de validation d'un test diagnostique	10
2.2 Les trois conditions impérativement requises pour démontrer l'utilité clinique d'un test diagnostique	10
3. Principaux schémas d'études cliniques : analyse critique	13
3.1 L'étude d' « interaction marqueur/traitement » : seul schéma permettant de démontrer l'utilité clinique d'un test identifiant un marqueur présumé prédictif et le caractère ciblé d'une thérapie	13
3.2 Les principaux autres schémas d'étude : détermination de l'effet du traitement et identification exploratoire de marqueurs présumés prédictifs	15
4. Evaluation économique	21
Conclusion	22
Références	23
Annexe 1. Liste des tableaux	24
Annexe 2. Liste des figures	25
Fiche descriptive	26

Abréviations et acronymes

ALK..... *anaplastic lymphoma kinase*

AMM autorisation de mise sur le marché

cf...... confer

EGFR *epidermal growth factor receptor*

HER2 *human epidermal growth factor receptor-2*

IHC..... immunohistochimie (détection immunohistochimique)

Mq+..... marqueur positif

Mq-..... marqueur négatif

Glossaire

Biomarqueur : « caractéristique qui est objectivement mesurée et évaluée comme un indicateur de processus biologiques normaux ou pathologiques, ou de réponses pharmacologiques à une intervention thérapeutique » (définition du *National Institute of Health*, US)). cf. terme « marqueur ».

Facteur (marqueur) de risque : facteur (marqueur) associé à une augmentation de fréquence (couramment appelée risque) du critère de jugement considéré (exemples : décès, survenue d'une complication) chez les sujets qui en sont porteurs par rapport aux autres.

Facteur (marqueur) prédictif : facteur (marqueur) influençant l'effet du traitement et dont l'aptitude à modifier le devenir clinique des patients est démontrée. La démonstration de la valeur prédictive d'un marqueur est équivalente à celle de l'utilité clinique du test diagnostique associé.

Interaction : situation statistique où un tiers facteur modifie les effets d'un autre. En l'occurrence, un marqueur prédictif modifie l'effet du traitement.

Marqueur : dans le contexte de ce document, caractéristique des patients pouvant représenter un indicateur de risque (facteur de risque), d'effet du traitement (modificateur de l'effet du traitement) ou d'utilité de traiter (marqueur prédictif). Remarque : les termes biomarqueur et marqueur sont utilisés indifféremment dans le langage courant.

Médecine « personnalisée » : terme consistant en un abus de langage à proscrire et à remplacer par le terme médecine « stratifiée » compte tenu du fait que la validation du couple marqueur/traitement est fondée sur une approche populationnelle classique de validation de thérapeutique. La particularité de cette approche est limitée au fait que la population des patients est stratifiée en sous-populations en fonction du marqueur pressenti comme prédictif. cf. terme (médecine) « stratifiée ».

Médecine « stratifiée » (*stratified medicine*) : correspond à une approche thérapeutique où l'objectif est de sélectionner les patients auxquels administrer un traitement en fonction d'un marqueur prédictif, afin de ne traiter que la sous-population susceptible de recevoir un bénéfice du traitement.

Test compagnon : test diagnostique permettant de sélectionner, en fonction de leur statut pour un marqueur prédictif identifié par ce test, uniquement les patients chez lesquels le traitement est susceptible d'apporter un bénéfice parmi ceux diagnostiqués pour une maladie donnée. Le test est considéré comme « compagnon » d'utilisation du traitement.

Théranostic (ou théragnostic) : néologisme qui dérive de la contraction des termes « thérapeutique » et « diagnostic » ; c'est l'utilisation d'un test diagnostique, identifiant un marqueur, pour orienter la thérapeutique du patient en fonction de son statut pour le marqueur (statut positif ou négatif pour un marqueur binaire).

Thérapie ciblée : traitement qui n'a démontré de bénéfices que chez certains patients identifiés par un marqueur prédictif déterminé par un test compagnon.

Utilité clinique (*clinical utility*) : aptitude à améliorer le devenir clinique des patients en événements cliniques mesurables, et à apporter une valeur ajoutée en termes d'optimisation de décision de traitement et en corollaire de stratégie thérapeutique. Valider le caractère prédictif d'un marqueur équivaut à démontrer l'utilité clinique de son test diagnostique.

Validité analytique (*analytic validity*) : aptitude d'un test diagnostique à réaliser la mesure d'intérêt avec exactitude et fiabilité. Cette validité inclut la sensibilité et la spécificité analytiques, la reproductibilité, la robustesse et la satisfaction des contrôles de qualité.

Validité clinique (*clinical validity*) : aptitude d'un test diagnostique à prédire avec précision et fiabilité le phénotype clinique d'intérêt (exemple : survie globale ou survie sans progression des patients recevant un traitement donné). Elle inclut la sensibilité et la spécificité cliniques, ainsi que les valeurs prédictives positive et négative du test.

1. Généralités

1.1 Avant-propos

L'objectif de ce document est d'exposer les fondements théoriques et méthodologiques de l'évaluation d'un test diagnostique, nommé **test compagnon**, permettant de sélectionner par l'identification d'un **marqueur prédictif** les patients seuls susceptibles de recevoir un bénéfice d'une **thérapie dite ciblée**.

Du fait de sa relation étroite avec le traitement, dont il est nommément dit « compagnon », l'évaluation de ce type de test est singulière car elle présente la particularité de ne pas pouvoir être dissociée de celle de l'efficacité du traitement. Il ne peut s'agir que d'une **évaluation synchrone** à celle du traitement. De ce fait résultent nécessairement des conséquences directes en termes de méthodologie d'étude permettant de répondre aux exigences de cette évaluation.

En termes d'**exigences**, la problématique majeure réside dans la nécessité d'apporter la **preuve que le marqueur** identifié par le test évalué est réellement **prédictif** de l'efficacité (ou la toxicité¹) du traitement revendiqué comme ciblé. Il faut sur ce point souligner que l'appellation de **thérapie ciblée** constitue une **terminologie impropre en l'absence de niveau de preuve satisfaisant** du ciblage concerné, donc de la démonstration du caractère prédictif du marqueur de sélection des patients. En l'absence de preuve de ce caractère prédictif, il peut **seulement** être affirmé que la thérapeutique est **stratifiée**, c'est-à-dire que son efficacité a été étudiée et montrée dans une sous-population (strate) de patients. En outre, démontrer le caractère prédictif du marqueur équivaut à démontrer l'**utilité clinique** du test l'identifiant, c'est-à-dire l'aptitude du test à améliorer le devenir clinique des patients en aidant à la prise de décision dans le cadre de leur prise en charge thérapeutique. C'est ce niveau de validation qui doit être atteint pour que le test puisse être **qualifié de test compagnon** et la **thérapie de ciblée**.

La réflexion rapportée dans ce document a été élaborée à partir de la littérature disponible sur le sujet, relativement rare, mais également et surtout à partir des principes méthodologiques fondamentaux sur lesquels se fondent plus largement le développement et l'évaluation des thérapeutiques. Cette démarche a permis de proposer les **exigences d'évaluation pour les tests compagnons** en cohérence avec les exigences actuellement imposées pour les médicaments, permettant ainsi de préciser la nature des connaissances apportées par différents schémas d'études, ainsi que les incertitudes qui peuvent leur être associées.

D'une façon générale, les **principales sources d'erreurs** qui peuvent conduire à conclure à tort à l'utilité clinique d'un test diagnostique sont les mêmes que pour les médicaments. Ce sont essentiellement :

- une approche exploratoire *post-hoc* pouvant amener à considérer comme preuve un artefact trouvé à l'occasion de l'exploration d'une abondance de résultats ;
- une erreur « statistique » due au hasard par l'intermédiaire des fluctuations aléatoires d'échantillonnage ;
- un biais dans la démonstration. Les principes méthodologiques de la randomisation, du double aveugle, de l'analyse en intention de traiter permettent de prévenir l'introduction de biais dans les études ;
- un défaut de pertinence clinique des résultats.

Par conséquent, l'évaluation doit s'assurer que les résultats sur lesquels repose la revendication d'utilité clinique d'un test diagnostique respectent les **principes suivants** :

- démarche hypothético-déductive (1);

¹ Dans le document ci-présent, nous nous sommes limités aux tests identifiant des marqueurs prédictifs d'efficacité d'un traitement, mais les principes décrits sont transposables aux marqueurs prédictifs d'effets indésirables.

- contrôle du risque d'erreur de première espèce alpha ;
- contrôle du risque de biais : stricte application des principes méthodologiques des essais contrôlés randomisés (randomisation, double aveugle, analyse en intention de traiter) ;
- pertinence clinique des résultats : une amélioration tangible et probante du devenir des patients doit être apportée, sur des critères de jugement cliniques (parfois clinico-biologiques) adéquats.

1.2 Notion de marqueur prédictif

La définition d'un « marqueur prédictif » réunit **deux conditions**.

La **première** est qu'un marqueur prédictif doit être un « **modificateur** » de l'**effet** d'un traitement, c'est-à-dire qu'il doit prédire en fonction de sa valeur l'effet d'un traitement donné pour un critère de jugement donné (2,3). En termes statistiques, cette condition correspond à l'existence d'une interaction marqueur/traitement, explicitée plus loin dans le document.

La **deuxième** est que la sélection des patients à traiter par l'identification du marqueur prédictif (donc le test diagnostique) doit avoir montré son **utilité clinique** : optimisation de l'efficacité chez les patients portant le marqueur et absence d'intérêt du traitement chez les autres patients, à qui il est donc inutile de l'administrer.

Du fait des développements récents dans le domaine de la biologie moléculaire, les marqueurs prédictifs sont à ce jour très essentiellement des variants génétiques. Le principal domaine de développement des thérapeutiques ciblées (stratifiées) est l'oncologie, où de nombreux nouveaux marqueurs candidats proposés correspondent à des variations génétiques au sein de proto-oncogènes à l'origine de modifications d'activités protéiques impliquées dans le processus d'oncogenèse.

Le plus souvent, les marqueurs prédictifs sont **binaires**² et définissent ainsi deux types de « statuts » pour les patients : marqueur positif (noté marqueur (+) ou Mq+ dans la suite du document) et marqueur négatif (noté marqueur (-) ou Mq-). **Marqueur (+)** correspond à la **modalité d'intérêt**, c'est-à-dire celle des patients chez lesquels le traitement est efficace ou à risque de survenue d'une toxicité. En l'occurrence, dans le document ci-présent, nous nous sommes limités au cadre des tests identifiant des marqueurs prédictifs d'efficacité d'un traitement. Cependant, les principes y étant décrits sont tout à fait transposables à la validation des marqueurs prédictifs d'effets indésirables, les patients marqueurs (+) étant alors ceux les plus à risque d'effets indésirables avec le traitement.

1.3 Notion de test compagnon

On appelle test compagnon un test diagnostique permettant de **sélectionner**, en fonction de leur statut pour un marqueur prédictif identifié par ce test, uniquement les patients chez lesquels le **traitement est susceptible d'apporter un bénéfice** parmi ceux diagnostiqués pour une maladie donnée. Le test est donc « compagnon » d'utilisation de l'administration du traitement.

Comme évoqué en avant-propos dans ce document, l'évaluation de cette approche ne peut reposer logiquement que sur une **évaluation conjointe du test et du traitement** avec un double objectif : montrer que le traitement n'est efficace que chez les patients marqueur (+) et que le test compagnon permet d'améliorer la prise de décisions thérapeutiques et ainsi l'évolution clinique des patients.

Il est important de souligner que l'intérêt de la sélection des patients présumés meilleurs voire seuls répondeurs au traitement par un test compagnon est aussi d'**optimiser la rentabilité cli-**

² Dans ce document, nous nous limiterons aux marqueurs binaires. Le guide méthodologique intitulé « Test compagnon associé à une thérapie ciblée : Annexe scientifique » apporte quelques compléments sur les marqueurs de nature continue.

nique³ de ce traitement par rapport à une stratégie d'utilisation du traitement chez tous les patients. En effet, la sélection de patients sur la base de l'identification d'un marqueur peu ou non prédictif par un test diagnostique, qui ne peut en l'occurrence pas être nommé test compagnon, n'optimise pas ou peu cette rentabilité clinique, et peut occasionner une perte de chance pour certains patients si le traitement est en réalité plus ou moins bénéfique chez tous. La suite de ce document montrera que ce problème éthique constitue une préoccupation majeure dans le cadre de l'évaluation de l'utilité de ces tests.

1.4 Notions de thérapeutiques « stratifiée » et « ciblée »

L'approche thérapeutique liée à l'utilisation de marqueurs prédictifs est parfois qualifiée de **médecine personnalisée** (« *personalized medicine* ») ou individualisée (« *individualized medicine* »). Ces appellations doivent être cependant considérées comme des abus de langage car l'approche impliquant les marqueurs prédictifs reste populationnelle et statistique, et non individuelle, et correspond ainsi en réalité à une approche de type **médecine stratifiée** (« *stratified medicine* ») ou médecine de précision (« *precision medicine* ») (4). Elle affine simplement les estimations en s'intéressant à des sous-populations de patients (d'où le terme de stratification) et produit ainsi une estimation moyenne de l'effet traitement dans ces sous-populations (strates), et non une estimation de l'effet du traitement chez un individu.

On appelle **thérapie ciblée** un traitement n'apportant de bénéfices que chez certains patients identifiés par un **marqueur prédictif** déterminé par un **test compagnon**.

En pratique, il apparaît que la revendication du caractère ciblé d'un traitement peut être observée dans deux situations :

- un ciblage recherché par le développeur au moment même de la conception (synthèse) du traitement : on parle alors de ciblage « par construction » ;
- une recherche exploratoire par analyse rétrospective de résultats d'un (ou plusieurs) essai réalisé sur une population de patients non sélectionnés par un marqueur, suivie ou non par un essai prospectif de confirmation (cf. section 3.2).

La démarche de **ciblage « par construction »** a été historiquement la première utilisée. Elle l'est toujours aujourd'hui. Le plus souvent, l'efficacité des molécules revendiquées comme relevant de ce concept de ciblage par construction n'est alors étudiée que chez les patients porteurs du marqueur désigné comme cible, l'absence d'efficacité de la molécule chez les patients marqueurs (-) étant considérée comme implicite. Cependant, la valeur prédictive du marqueur ne peut raisonnablement pas en toute rigueur reposer sur des fondements implicites, mais uniquement sur des preuves cliniques. Les données précliniques ne constituent que des données préliminaires, de statut exploratoire, et ne reflétant que plus ou moins bien celles qui seront obtenues chez l'homme. En effet, du fait de la grande complexité et des spécificités de l'organisme humain, aucun type d'études précliniques (lignées cellulaires immortalisées, études animales...) ne peut prévoir précisément et avec fiabilité les effets cliniques d'un médicament. En outre, revendiquer l'absence d'efficacité du traitement chez les marqueurs (-) comme implicite impliquerait de pouvoir apporter la preuve d'une spécificité indiscutable du traitement envers sa cible *in vivo*. Or, il paraît fort difficile de pouvoir apporter une telle démonstration.

En l'occurrence, à ce jour un certain recul sur les premières thérapies qualifiées de ciblées a fait apparaître **plusieurs exemples** où, bien que le traitement ait été initialement considéré comme **a priori spécifique de sa cible** et le marqueur comme central dans le mécanisme d'action du traitement, ce dernier s'est en fait parfois révélé actif chez des patients ne portant pas ce marqueur pour diverses raisons possibles. Cela est notamment le cas avec le trastuzumab dans le cancer du sein, pour lequel une efficacité chez les cancers du sein HER2/neu (*human epidermal growth factor receptor-2*) négatifs est envisagée. Un deuxième exemple est le cetuximab. Cet autre anti-

³ La rentabilité clinique correspond au nombre d'événements évités pour le traitement d'un certain nombre de patients.

corps monoclonal a été synthétisé pour « cibler » le récepteur EGFR (*epidermal growth factor receptor*), récepteur oncogénique détecté dans 70-85 % des cancers colorectaux par technique immunohistochimique (IHC). Or, à ce jour, l'analyse des données de certains essais a montré que jusqu'à 25 % patients atteints d'un cancer colorectal métastatique identifié « EGFR négatif » par IHC pourraient répondre au cetuximab. Les explications avancées seraient la mauvaise spécificité de la technique IHC et/ou un phénomène d'ADCC (*Antibody-dependent cell-mediated cytotoxicity*) lié au cetuximab envers les cellules tumorales. Par ailleurs, au sein de la deuxième grande classe de thérapies dites ciblées en oncologie, représentée par les inhibiteurs de récepteurs à activité tyrosine kinase (RTK), il est assez fréquent de ne pas observer, dès les résultats précliniques, une réelle spécificité pour une cible unique. Par exemple, le crizotinib, utilisé en tant qu'inhibiteur du RTK nommé ALK (*anaplastic lymphoma kinase*) et de ses variants oncogéniques dans le cancer du poumon non à petites cellules, inhibe aussi le RTK du facteur de croissance des hépatocytes (c-Met) également impliqué dans les processus oncogéniques. Dans son rapport d'évaluation associé à l'enregistrement de la molécule, l'EMA (*European Medicine Agency*; EMA/CHMP/497137/2012) évoque clairement les limites de spécificité du crizotinib : « *When tested to its selectivity for different kinases, crizotinib was relatively specific to c-Met/HGFR and ALK fusion proteins* ». Ces exemples montrent bien qu'un « ciblage par construction » n'est pas suffisant pour justifier de l'absence d'étude clinique chez les patients marqueur (-). L'absence d'efficacité du traitement chez ces patients doit être cliniquement prouvée pour **attester de la réalité *in vivo* du ciblage**, le risque d'accepter un ciblage à tort d'un nouveau traitement ayant pour conséquence directe une perte de chance pour ces patients.

L'approche de revendication du caractère ciblé d'un traitement faisant suite à la **recherche exploratoire *post-hoc*** de marqueurs prédictifs est détaillée dans la section 3.2. Il est intéressant de noter que, d'une certaine façon, cette voie de recherche **a élargi le concept de « caractère ciblé »**. En effet, ce concept référait initialement, en lien avec le « ciblage par construction », à la synthèse de traitements orientés vers des cibles moléculaires impliquées dans leur mécanisme d'action, et faisant intervenir une liaison entre le médicament et sa cible. Avec les recherches *post-hoc* de marqueurs prédictifs, le concept de caractère ciblé s'est élargi à la notion plus générale de marqueur prédictif. Cette extension du concept ne pose pas de réel souci de cohérence car les cibles moléculaires sont, en théorie, censées être également des marqueurs prédictifs (même si ce n'est pas toujours le cas ; cf. section 3.2.3 « Recherche d'un marqueur prédictif après un essai non concluant »). Cependant, il faut tout de même relever qu'avec l'élargissement du concept aux marqueurs prédictifs en général, la compréhension des **mécanismes d'implication** d'un marqueur dans l'efficacité d'un traitement donné, ainsi que du mécanisme d'action global du traitement, est devenue plus complexe.

2. Principes généraux de validation d'un test diagnostique déterminant le statut des patients pour un marqueur présumé prédictif

2.1 Les trois niveaux de validation d'un test diagnostique

L'évaluation du test diagnostique déterminant le statut des patients pour un marqueur présumé prédictif comporte trois niveaux : l'évaluation de la validité analytique, de la validité clinique et de l'utilité clinique du test (5,6).

- **La validité analytique (*analytic validity*)** d'un test diagnostique correspond à son aptitude *in vitro* à réaliser la mesure d'intérêt avec exactitude et fiabilité. En d'autres termes : le test mesure-t-il bien ce qu'il est censé mesurer et de façon correcte ? Cette validation inclut l'étude de la sensibilité et la spécificité analytiques, la reproductibilité, la robustesse et la satisfaction des contrôles qualité.
- **La validité clinique (*clinical validity*)** d'un test diagnostique correspond à son aptitude à prédire avec précision et fiabilité le phénotype clinique d'intérêt (ex : survie globale ou survie sans progression des patients recevant un traitement donné) : y-a-t-il une relation, et laquelle, entre les résultats du test et le phénotype d'intérêt ? La validité clinique inclut la sensibilité et la spécificité cliniques, ainsi que les valeurs prédictives positive et négative du test, paramètres que l'on peut réunir sous l'appellation de « performances diagnostiques » du test.
- **L'utilité clinique (*clinical utility*)** d'un test diagnostique correspond à son aptitude à améliorer le devenir clinique des patients en événements cliniques mesurables, et à apporter une valeur ajoutée en termes d'optimisation de décision de traitement et en corollaire de stratégie thérapeutique.

L'étude de la validité analytique et de la validité clinique sont des notions génériques pour les tests diagnostiques, qu'un marqueur soit ou non impliqué. Elles n'entrent donc pas dans le champ de ce document dédié aux tests associés à des marqueurs prédictifs.

La **démonstration de l'utilité clinique** est un **élément indispensable** dans le domaine des tests compagnons. Cette démonstration est indispensable parce qu'elle atteste de la **valeur ajoutée apportée par le test évalué** en comparaison d'un test diagnostique permettant uniquement de sélectionner une sous-population par l'identification d'un marqueur. Le concept même de test compagnon intègre que son utilité clinique soit démontrée puisqu'il doit permettre de ne pas traiter inutilement des patients chez lesquels le traitement serait inefficace, et à l'inverse de ne pas entraîner de perte de chance par l'identification d'un marqueur non ou peu prédictif en excluant des patients d'un traitement qui leur apporterait en réalité un bénéfice.

2.2 Les trois conditions impérativement requises pour démontrer l'utilité clinique d'un test diagnostique

La démonstration de l'utilité clinique d'un test diagnostique implique la **démonstration des trois éléments suivants** (7–11) :

- le marqueur identifié par le test doit modifier l'effet du traitement (existence d'une interaction « marqueur/traitement ») ;
- le traitement doit être efficace chez les patients marqueur (+) ;
- le traitement doit être sans intérêt clinique chez les patients marqueur (-).

2.2.1 Existence d'une interaction marqueur/traitement

Les effets d'un traitement sont quantifiés le plus souvent par un risque relatif. Dans les essais thérapeutiques, le risque relatif est le facteur multiplicatif par lequel le traitement augmente le risque, c'est-à-dire la fréquence du critère de jugement, des patients le recevant comparé à celui des patients du groupe contrôle. On le calcule comme un rapport des risques. Un essai clinique permet d'obtenir des estimations des vrais risques relatifs associées à des intervalles de confiance.

Il y a interaction « marqueur/traitement » quand la valeur de ce risque relatif est influencée par le marqueur. Les valeurs d'effet du traitement observées peuvent cependant être différentes sans que ce soit une réalité mais simplement un artefact dus aux fluctuations aléatoires. Il n'est donc possible de parler d'interaction que si, dans le cas d'un marqueur binaire, une différence statistiquement significative d'effet du traitement est observée entre le sous-groupe des patients marqueur (+) et celui des patients marqueur (-).

Il est néanmoins important de noter que si un test d'interaction statistiquement significatif atteste d'une différence d'effet du traitement dans les sous-groupes marqueur (+) et marqueur (-), cela ne signifie pas que le traitement n'a pas d'effet chez les patients marqueur (-). Il est en effet possible que le traitement soit seulement moins efficace chez les patients marqueur (-). Le concept approché est alors celui de « pseudo-ciblage » abordé dans le guide méthodologique annexe⁴ à celui-ci.

2.2.2 Efficacité du traitement chez les patients marqueur (+)

L'essai doit logiquement démontrer l'efficacité du traitement chez les patients marqueur (+), ceci en suivant les **mêmes critères que pour un essai thérapeutique classique** : essai prospectif, à l'abri des biais, statistiquement significatif, cliniquement pertinent et ayant une bonne cohérence externe (plausibilité biologique, cohérence entre les études disponibles).

Un aspect spécifique par rapport à un essai classique concernant la démonstration de l'efficacité d'un nouveau traitement est le contrôle du risque alpha. En effet, une multiplicité latente des tests statistiques est inhérente au design dans ce concept d'essai : tests de l'effet du traitement chez les patients marqueur (+), marqueur (-) et/ou tous patients confondus. À chaque fois, un risque de conclure à tort de 5 % peut être pris en l'absence de précaution. Le **contrôle d'inflation du risque alpha** peut être réalisé de plusieurs façons. La première consiste en une analyse séquentielle hiérarchique où l'on teste en premier les patients marqueur (+), puis la totalité des patients. La deuxième est de pré-spécifier explicitement que l'objectif primaire de l'essai est de valider l'interaction et l'efficacité du traitement chez les patients marqueur (+), en concentrant le risque alpha sur l'analyse liée à cet objectif primaire. Cependant, les analyses tous patients confondus et chez les patients marqueur (-) poseront un problème de multiplicité des comparaisons statistiques. Enfin, une troisième possibilité est d'ajuster le risque alpha en tenant compte du nombre des différentes comparaisons possibles à l'aide d'une méthode de type Bonferroni ou autre.

2.2.3 Absence d'efficacité du traitement chez les patients marqueurs (-)

Le concept de marqueur prédictif implique aussi l'absence d'effet du traitement chez les patients marqueur (-).

Une démonstration statistique de l'absence d'effet dans ce sous-groupe doit être apportée. Cette démonstration doit faire appel à une étude de non infériorité basée sur l'intervalle de confiance du risque relatif de ces patients, en montrant qu'il est possible d'exclure avec un degré de certitude de 97.5 % que l'effet du traitement puisse être supérieur à une certaine valeur correspondant au bénéfice minimal cliniquement intéressant. En pratique, il s'agit en fait d'un **raisonnement de non supériorité** (avec une limite de non supériorité clinique), en miroir de celui d'un raisonnement de non infériorité. On peut conclure à la non supériorité lorsque la borne inférieure de l'intervalle de

³ Guide méthodologique intitulé « Test compagnon associé à une thérapie ciblée : Annexe scientifique ».

confiance de l'effet du traitement dans la strate des patients marqueur (-) est supérieure à la limite de non supériorité. On cherche en effet à déterminer s'il est raisonnable d'exclure la possibilité que le traitement soit efficace (c'est-à-dire qu'il présente un effet substantiellement supérieur au placebo) chez les patients marqueur (-). La **limite** de non supériorité doit être **fixée a priori** dans le protocole d'étude et justifiée cliniquement. En outre, un intervalle de confiance étroit obtenu chez les patients marqueur (-) est plus favorable à l'obtention de cette démonstration, ce qui sous-entend qu'un calcul du nombre de sujets nécessaires spécifique ait été réalisé *a priori* afin de garantir cette précision.

3. Principaux schémas d'études cliniques : analyse critique

Ce chapitre propose les principes de conduite par la Haute Autorité de santé de l'évaluation d'un test diagnostique permettant d'identifier un marqueur présumé prédictif associé à une thérapeutique stratifiée. Une synthèse pragmatique de ces principes y est présentée mais les concepts mis en œuvre n'y sont pas détaillés. Ces derniers sont présentés et expliqués dans le guide méthodologique annexe intitulé « Test compagnon associé à une thérapie ciblée : Annexe scientifique ».

Le principe de base de l'évaluation de ce type de test par la HAS réside dans le fait que la reconnaissance des appellations « test compagnon » et « thérapie ciblée » repose sur la démonstration, apportée ou non, de l'utilité clinique du test. Par conséquent, sont présentées dans ce chapitre les **preuves requises pour la démonstration d'utilité clinique d'un test diagnostique associé à une thérapeutique stratifiée**. Cette démonstration prouve conjointement la valeur prédictive du marqueur vis-à-vis de l'efficacité du traitement, le ciblage de la thérapie et l'appellation de test compagnon.

L'utilité clinique d'un test diagnostique est parfois revendiquée à partir d'arguments ou de données issues d'études qui ne permettent pas de la démontrer mais qui sont plus facilement obtenues que des éléments de preuve rigoureux. En l'occurrence, la reconnaissance ou non de cette revendication **dépend directement du schéma utilisé pour l'étude** supportant cette demande. Il est de plus important de souligner qu'un cadre méthodologique rigoureux répondant aux standards actuels de l'évaluation des thérapeutiques est nécessaire compte tenu de l'implication directe des tests compagnons dans la décision de traiter ou non des patients.

Pouvoir garantir l'utilité clinique d'un test compagnon et concomitamment le ciblage de la thérapie constitue un **enjeu de santé** important car les risques en l'absence d'évaluation sont nombreux :

- le risque de recommander à tort un traitement en réalité sans efficacité (résultat biaisé ou erreur de première espèce/risque alpha) ;
- le risque de recommander à tort un traitement en réalité sans intérêt clinique (absence de pertinence clinique).

Ces deux premiers risques sont identiques à ceux que l'on cherche à exclure dans l'analyse d'un dossier classique de thérapeutique.

- le risque de recommander à tort une stratégie stratifiée impliquant la réalisation chez tous les patients d'un test diagnostique supplémentaire inutile ;
- le risque de ne pas utiliser à tort chez les patients marqueur (-) un traitement qui pourrait avoir un certain intérêt clinique chez eux, et d'induire ainsi chez eux une perte de chance.

In fine, ne pas s'assurer de l'utilité clinique d'un test identifiant les sous-populations concernées par un traitement stratifié peut conduire à recommander une stratégie discriminante d'utilisation d'un nouveau traitement et non optimale en termes de bénéfice populationnel.

3.1 L'étude d' « interaction marqueur/traitement » : seul schéma permettant de démontrer l'utilité clinique d'un test identifiant un marqueur présumé prédictif et le caractère ciblé d'une thérapie

► Caractéristiques de l'approche

Un seul schéma d'étude (ou design), appelé schéma d' « interaction marqueur/traitement », d' « essai de recherche d'interaction » ou « *marker by treatment interaction design* », permet de déterminer de façon concomitante l'effet du traitement chez les patients marqueur (+) et marqueur (-), et de tester en corollaire l'existence d'une interaction marqueur/traitement. Ce schéma est ainsi

le seul à pouvoir **démontrer conjointement les trois conditions** permettant la démonstration de l'utilité clinique d'un test diagnostique identifiant un marqueur présumé prédictif (2,3,12–14). Il est en corollaire également le seul à pouvoir démontrer qu'un traitement peut être réellement qualifié de ciblé, et non simplement de stratifié.

En tant que seul plan d'expérience permettant de démontrer strictement l'utilité clinique d'un test, ce schéma d'étude conduit aux résultats de plus haut niveau de preuve et constitue la **référence** dans son domaine.

Cette approche **peut être envisagée dans la majorité des situations** :

- avant toute démonstration de l'intérêt du traitement. Il s'agit alors d'un développement conjoint du traitement et du test. Le centrage est celui du traitement et la finalité du dossier est de revendiquer l'intérêt du traitement dans la sous-population sélectionnée par le marqueur ;
- à la suite d'une recherche *post-hoc* en sous-groupes d'un marqueur prédictif suite à un essai non concluant portant sur l'intérêt d'un nouveau traitement. L'étude d'interaction correspond dans ce cas à une étude prospective de confirmation de résultats exploratoires.

Elle ne peut cependant pas être envisagée pour confirmer la valeur prédictive présumée d'un marqueur candidat lors d'une recherche *post-hoc* en sous-groupes à la suite d'une étude aux résultats concluants, ceci pour des raisons éthiques. Il n'est en effet pas raisonnable d'envisager de randomiser à nouveau les patients contre le comparateur du premier essai, le nouveau traitement ayant démontré sa plus grande efficacité pour tous les patients indépendamment de leur statut pour le marqueur (cf. section 3.2.3).

Les **enjeux éthiques** soulevés par l'analyse de l'absence d'efficacité du traitement chez les patients marqueur (-) sont similaires à ceux que soulève la randomisation dans les essais cliniques en général. La justification de cette randomisation d'un point de vue éthique repose sur le **principe d'équipoise** : il est justifié de conduire une étude clinique randomisée lorsque la communauté médicale ne peut pas déterminer, selon le principe de la médecine basée sur les preuves, la meilleure option entre deux traitements. La randomisation n'est acceptable que tant que l'incertitude demeure (15–17). Les patients marqueur (-) ne doivent pas être exposés à des effets indésirables sans espoir escompté de bénéfice. Il paraît alors important de garder à l'esprit, dans le contexte des thérapies stratifiées, que l'**absence d'efficacité** du nouveau traitement chez les patients marqueur (-) n'est pas un **résultat connu a priori**, mais une condition nécessaire à la démonstration du caractère ciblé de la thérapie. L'**hypothèse** testée est celle de la **non-supériorité du nouveau traitement au comparateur**, placebo ou traitement de référence en cours, et comme n'importe quelle hypothèse testée par une étude clinique, elle sera susceptible d'être *in fine* rejetée ou non. En l'occurrence, c'est pour cette raison qu'il est recommandé de prévoir l'analyse de l'efficacité du traitement tous patients confondus (Mq+ et Mq-) en plus de l'analyse sous-populationnelle, dans le plan d'analyse statistique du schéma d' « interaction marqueur/traitement ». Cette position est renforcée par le fait qu'à ce jour des **contre-exemples existent** parmi les traitements qui étaient revendiqués comme *a priori* « ciblés par construction » et pour lesquels l'absence d'efficacité chez les patients marqueur (-) a été remise en question. L'absence d'efficacité d'un traitement chez des patients marqueur (-) qui serait « prédite » par un marqueur identifié par des analyses *post-hoc* est aussi très déraisonnable compte tenu de la nature exploratoire de la recherche et du caractère globalement purement théorique du **mécanisme d'implication** du marqueur dans l'efficacité du traitement considéré (cf. section 1.4). Au final, il apparaît que ne pas réaliser cette recherche ferait globalement courir un risque aux patients marqueur (-) en favorisant l'**acceptation à tort de marqueurs non ou insuffisamment prédictifs** susceptibles par la suite de leur causer une perte de chance en les privant d'un traitement efficace pour eux.

► Schéma d'étude

Ce schéma comprend une **double randomisation**, une dans chaque sous-population identifiée par son statut positif ou négatif pour le marqueur, ce qui permet alors de comparer le nouveau

traitement N à un comparateur C aussi bien chez les patients marqueur (+) que chez les patients marqueur (-) (cf. Figure 1). Deux estimations de l'effet de N *versus* C sont produites, chez les patients marqueur (+) et (-). Leur comparaison permet la recherche d'une différence statistiquement significative signifiant le cas échéant l'existence d'une interaction marqueur/traitement.

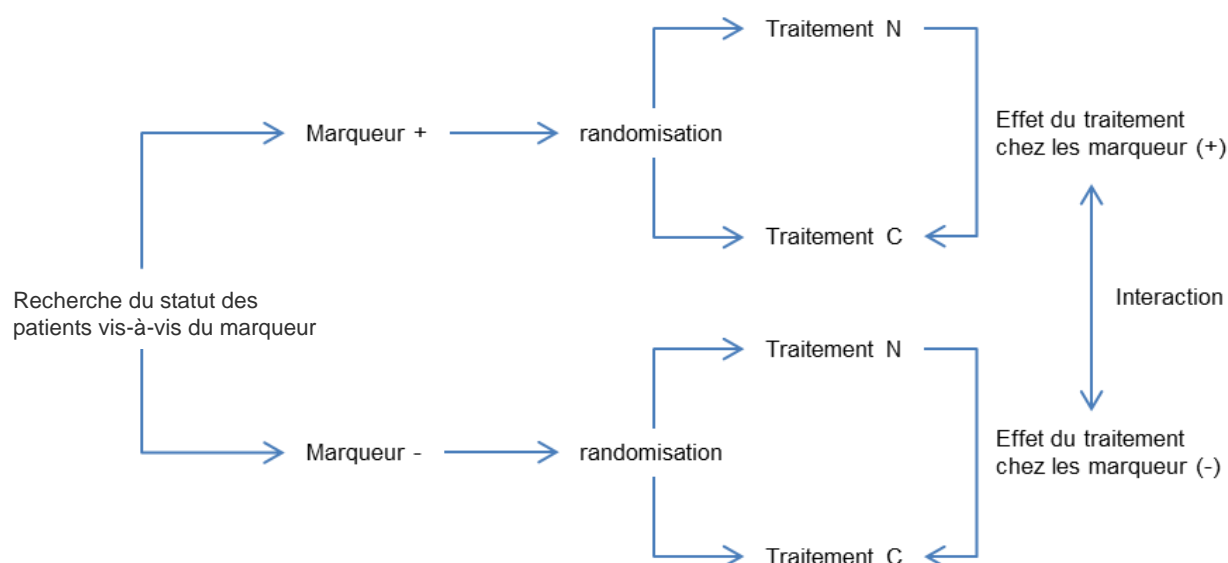


Figure 1 – Schéma d'étude d' « interaction marqueur/traitement »

3.2 Les principaux autres schémas d'étude : détermination de l'effet du traitement et identification exploratoire de marqueurs présumés prédictifs

En pratique, le type de méthodologie utilisée dans la majorité des études a été et reste à ce jour centré sur la démonstration de l'efficacité du nouveau traitement mais n'est pas adapté à celle de l'utilité clinique du test diagnostique associé.

3.2.1 L'étude « ciblée »

► Caractéristiques de l'approche

Le schéma d'étude de type « ciblée » a été à la base des premiers enregistrements de traitements de prescription limitée à une sous-population identifiée (« ciblée ») par la présence d'un marqueur présumé prédictif de l'efficacité de ce traitement. Il est toujours couramment utilisé.

Ce schéma est tout à fait approprié pour démontrer l'efficacité d'un nouveau traitement dans une sous-population de patients définie par la présence d'un marqueur détecté par un test diagnostique. Cependant, en toute rigueur, il ne permet pas de valider l'utilité clinique du test en question, ni par conséquent de revendiquer le caractère ciblé du traitement, car il ne permet pas de démontrer **explicitement** l'absence d'effet chez les patients marqueur (-) (18–20). En effet, pour les raisons explicitées dans la section 1.4, il est peu raisonnable de considérer l'absence d'efficacité chez les patients marqueur (-) comme **implicite** sur la base de données précliniques et/ou d'un rationnel biologique théorique.

Lorsque l'efficacité du nouveau traitement s'avère démontrée par ce schéma dans le sous-groupe traité, donc chez les patients marqueur (+), l'**AMM dans l'indication** étudiée est logiquement **restreinte au sous-groupe** de patients chez lesquels l'efficacité du traitement a été étudiée, donc aux

patients **marqueur (+)**. Une prescription du médicament, pour être conforme à l'AMM, rend donc nécessaire la réalisation du test compagnon, afin de discriminer les patients marqueur (+) et (-). En effet, prescrire le médicament à un patient marqueur (-) consiste alors en une prescription hors AMM. Il faut souligner sur ce point que la restriction de prescription de l'AMM ne doit pas être considérée comme une preuve d'inefficacité du traitement chez les patients marqueur (-) mais qu'elle résulte seulement de l'absence d'évaluation de cette efficacité. Le **principal risque** de ce design est par conséquent d'entraîner, en l'absence de preuves cliniques, une **perte de chance pour** les patients **marqueur (-)** en les privant d'un traitement susceptible de leur apporter un bénéfice, en plus de réaliser un test inutile.

In fine, ce schéma n'est **pas adapté pour démontrer l'utilité clinique** d'un test diagnostique associé à une thérapie stratifiée. Néanmoins, au cas par cas, le niveau de conviction accordé aux éléments précliniques et cliniques précoces lors de l'évaluation peut soutenir dans une certaine mesure cette démonstration. Ce schéma présenterait plutôt essentiellement un intérêt lorsque des données précliniques et/ou cliniques précoces amènent à craindre une toxicité particulière du nouveau traitement chez les patients marqueur (-). Cependant, l'utilité clinique du test ne sera pas démontrée et il peut donc être préférable, lorsque cela est possible, de choisir un design d'« interaction marqueur/traitement » associé à un protocole permettant d'assurer aux patients le niveau de sécurité exigible pour tout essai thérapeutique.

► Schéma d'étude

Le schéma d'étude de type ciblé, également nommé « design ciblé », « *targeted design* » ou bien encore « *enrichment design* », consiste à évaluer le nouveau traitement *versus* le traitement comparateur C uniquement chez des sujets marqueur (+) (cf. Figure 2).

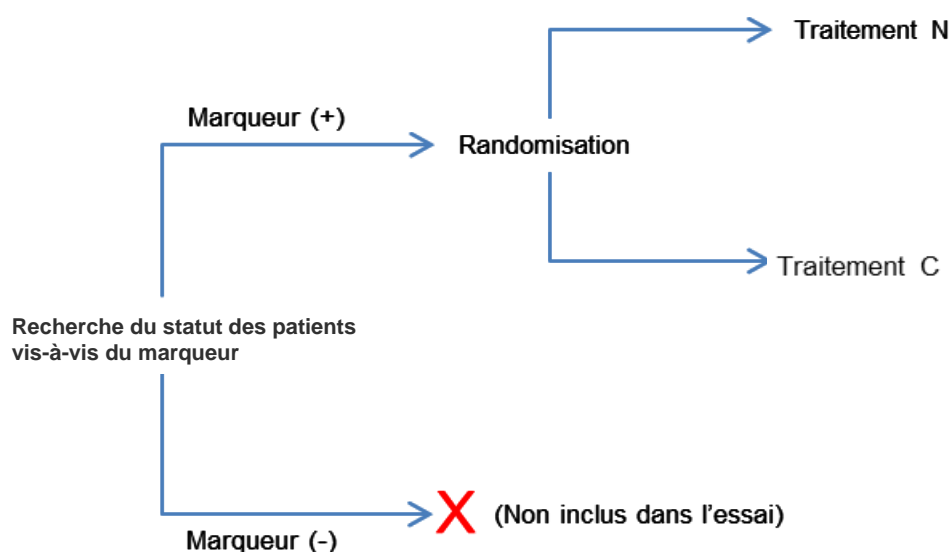


Figure 2 – Schéma d'étude « ciblée »

3.2.2 L'étude de « comparaison de stratégies »

► Caractéristiques de l'approche

Dans le schéma d'étude dit de « comparaison de stratégies », la stratégie d'administration du nouveau traitement basée sur la détermination du marqueur par le test diagnostique est comparée

à une stratégie « standard » où le traitement standard est administré sans détermination préalable du marqueur (cf. Figure 3).

Comme pour le design ciblé, ce schéma ne permet pas de mesurer l'effet du nouveau traitement chez les patients marqueur (-) et ne permet donc pas de démontrer l'utilité clinique du test diagnostique, ni par conséquent le caractère ciblé du nouveau traitement.

En réalité, malgré son appellation, ce schéma ne permet pas non plus de démontrer *stricto sensu* la supériorité d'une stratégie stratifiée « marqueur/traitement », où le choix du traitement est associé au résultat du test compagnon, en comparaison d'une stratégie standard non stratifiée. En effet, obtenir un meilleur résultat avec la stratégie « marqueur/traitement » évaluée par rapport à la stratégie standard non stratifiée peut être observé si le nouveau traitement est plus efficace que le traitement standard chez les patients marqueur (+), mais également si le nouveau traitement est plus efficace que le traitement standard chez l'ensemble des patients, qu'ils soient marqueur (+) ou (-), donc que le test soit cliniquement utile ou non. Par conséquent, une étude utilisant ce schéma et présentant des résultats concluants ne permet de conclure que sur la supériorité de la stratégie globale basée sur l'utilisation du marqueur et du nouveau traitement par rapport à l'utilisation du traitement standard chez tous les patients, mais ne permet pas de savoir si, dans le cadre de la stratégie basée sur le marqueur, la sélection des patients devant recevoir le nouveau traitement en fonction de la présence du marqueur est cliniquement utile.

In fine, ce design ne permettant pas de s'assurer que le traitement est sans bénéfice chez les patients marqueur (-), son principal risque associé est, comme pour le design ciblé, de conclure à tort qu'il ne faut pas traiter ces patients, donc de générer une perte de chance pour ces patients conjointement à la mise en œuvre d'un test inutile.

► Schéma d'étude

Le schéma d'étude correspondant à l'évaluation de stratégies est appelé « design de comparaison de stratégies » ou « *marker-based strategy design* » (3). La stratégie standard de traitement de tous les patients par le traitement standard S est comparée à la stratégie consistant à traiter les patients marqueur (+) avec le nouveau traitement N et les patients marqueur (-) avec le traitement standard S. L'objectif (théorique) est de montrer que la nouvelle stratégie utilisant l'évaluation du marqueur permet d'obtenir de meilleurs résultats que la stratégie standard. Il convient donc de comparer les résultats obtenus avec la stratégie basée sur le marqueur (en regroupant les patients traités par N et par S) avec celle de la stratégie standard.

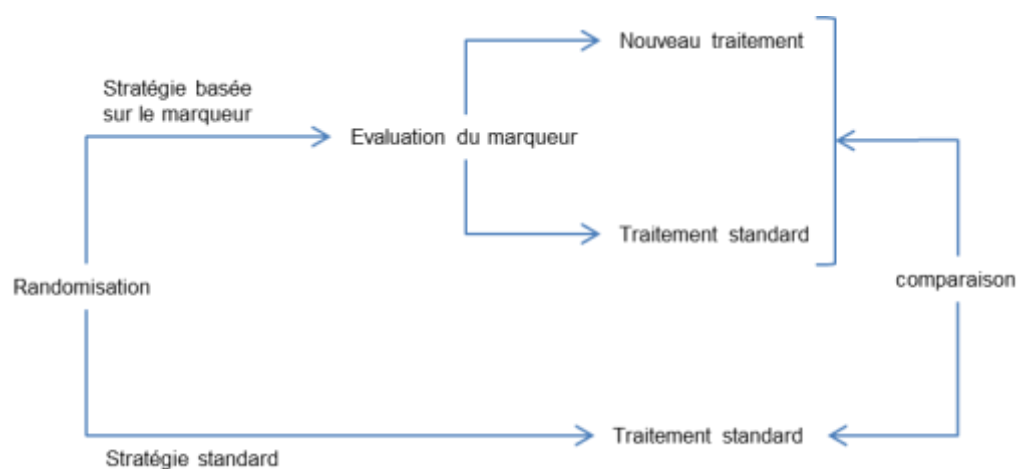


Figure 3 – Schéma d'étude de « comparaison de stratégies »

3.2.3 Analyses en sous-groupes *a posteriori*

► Analyses en sous-groupes : des résultats uniquement exploratoires

A ce jour, dans la littérature, il est observable que la majorité des études concluent à l'identification de marqueurs prédictifs à partir de résultats de recherches de marqueurs réalisées *a posteriori*. Ce type de démarche n'est pas méthodologiquement satisfaisant car, comme pour la démonstration de l'efficacité d'un traitement, la preuve de l'utilité clinique d'un test compagnon doit être apportée par une démarche hypothético-déductive.

Ainsi, les analyses en sous-groupes *a posteriori* (dites également *post-hoc* ou rétrospectives) ne sont pas adaptées à la recherche de marqueurs prédictifs en raison de leur nature exploratoire (de type « pêche à la ligne ») et de l'inflation du risque alpha qu'elles génèrent, même si ces marqueurs avaient été listés *a priori*. Cette démarche peut être utile voire nécessaire au repérage de marqueurs candidats d'intérêt. Cependant, les résultats obtenus ne peuvent aboutir qu'à une génération d'hypothèses qui devront être impérativement validées par des études *ad-hoc* (prospectives) de confirmation. A cet effet, le seul schéma permettant la démonstration de l'utilité clinique d'un test est, comme explicité précédemment (cf. section 3.1), le schéma d'interaction marqueur/traitement.

► Recherche d'un marqueur prédictif *a posteriori* après un essai concluant

Cette première situation peut se produire lorsqu'après un essai montrant l'efficacité d'un nouveau traitement sur une population de patients non sélectionnés, une recherche *post-hoc* de déterminants de la réponse peut faire suspecter que le traitement ne serait en réalité efficace que chez certains patients porteurs d'un marqueur défini.

Cette situation présente les problèmes méthodologiques des démarches *post-hoc*, en particulier de multiplicité, avec une inflation du risque bêta de ne pas mettre en évidence à tort l'efficacité du traitement chez les patients marqueur (-), et donc de récuser à tort le traitement de ces patients initialement inclus dans l'essai ayant démontré l'efficacité du traitement sans sélection des patients en fonction de leur statut pour le marqueur. Un essai *ad-hoc* est donc indispensable pour confirmer ces résultats.

Une **problématique éthique** apparaît néanmoins dans ce cas. En effet, par l'essai initial réalisé selon un schéma standard d'étude de thérapeutique non stratifiée et concluant, le nouveau traitement est devenu le traitement de référence dans le domaine concerné. Randomiser à nouveau des patients entre le nouveau traitement et un placebo (ou l'ancien traitement de référence) pose alors un problème éthique. Un **schéma de type « interaction marqueur/traitement »** ne peut donc **pas** être **recommandé**, en particulier à l'égard des patients marqueur (+) qu'il n'est pas raisonnable d'envisager de randomiser à nouveau *versus* le comparateur initial alors que l'étude de confirmation aurait pour objectif de démontrer que le traitement n'est en fait efficace que chez eux. Au final, la recherche d'un marqueur prédictif à la suite d'un essai concluant en population générale aboutit le plus souvent à des résultats exploratoires non confirmés prospectivement donc de niveau de preuve non acceptable pour démontrer l'utilité clinique du test diagnostique permettant d'identifier le marqueur.

Néanmoins, des situations peuvent se présenter où un **rational fort** peut apparaître dans le sens d'une probable absence d'efficacité chez les patients non porteurs d'un marqueur, justifiant ainsi éthiquement la conduite d'une étude prospective ayant pour objectif de confirmer (ou non) l'absence présumée d'efficacité du traitement chez ces patients. Dans ce cas, la démarche à suivre est celle d'une **étude prospective de non supériorité**, de type ciblée incluant uniquement des patients marqueur (-), ou bien de type « comparaison de stratégies » avec un bras « stratégie standard » correspondant à l'utilisation du nouveau traitement sans identification du marqueur. Une excellente qualité métrologique d'essai est impérative, le risque majeur étant de conclure à

tort à l'exclusion du traitement de la sous-population des patients marqueur (-). Si l'absence d'intérêt clinique du traitement chez les patients marqueur (-) est confirmée, il est possible de conclure à l'utilité clinique du test, ce dernier apportant, conformément à la définition de l'utilité clinique, une optimisation de la prise en charge des patients, en permettant en particulier d'éviter l'exposition inutile de patients à des effets indésirables potentiels sans bénéfices pour leur santé et de recourir, s'il en existe, à d'autres thérapeutiques potentiellement plus efficaces.

► Recherche d'un marqueur prédictif a posteriori après un essai non concluant

La recherche d'un marqueur prédictif après l'obtention de résultats non concluants dans un essai évaluant un nouveau traitement pose d'importants problèmes méthodologiques. Elle expose en particulier à un **fort risque de découverte artéfactuelle de marqueurs non prédictifs**, pouvant amener à priver à tort du traitement des patients susceptibles d'en recevoir des bénéfices et d'alourdir la prise en charge de tous les patients avec la réalisation d'un test inutile.

En effet, dans le contexte d'un essai non concluant, la recherche d'un marqueur prédictif résulte généralement directement de l'obtention de ce résultat non satisfaisant, qui amène à proposer l'hypothèse d'une efficacité non ubiquitaire mais limitée à un certain type de patients. Cette recherche est alors entreprise par une analyse rétrospective des données de l'essai non initialement prévue. Cependant, ces mêmes données seront ensuite également utilisées comme démonstration de la valeur prédictive de ce marqueur, créant ainsi une situation tautologique. En outre, faisant suite à l'obtention de résultats non concluants, une démarche *post-hoc* de type « pêche à la ligne » est utilisée et génère une multiplicité de comparaisons à l'origine d'une inflation du risque alpha sur l'hypothèse d'efficacité du traitement, revenant à s'octroyer une deuxième chance de conclure, à tort, à cette efficacité.

Par conséquent, comme pour toute analyse en sous-groupes, les résultats obtenus dans cette situation doivent être considérés comme exploratoires et impérativement validés par une étude de **confirmation prospective**. En l'occurrence, la démonstration du caractère prédictif d'un marqueur présumé prédictif par les analyses exploratoires requiert l'utilisation d'un **schéma de type « interaction marqueur/traitement »**.

Il faut souligner que la démonstration isolée de l'efficacité du nouveau traitement chez les patients marqueur (+) avec un autre schéma d'étude ne permettrait pas d'exclure qu'un premier essai conduit avec un nouveau traitement en réalité efficace chez tous les patients ait pu s'avérer « faussement » non concluant en raison d'un manque de puissance, puis qu'un second essai alors réalisé uniquement chez les patients porteurs d'un marqueur identifié dans les analyses en sous-groupes du premier essai ait pu ensuite être concluant uniquement en raison d'une meilleure puissance, que le marqueur soit prédictif ou non.

► Recherche de marqueurs limitée au groupe traité (« *treatment-only analysis* »)

Il n'est pas rare de rencontrer des travaux où la recherche de marqueurs prédictifs **s'apparente à une exploration des facteurs de risque** de l'événement correspondant au critère de jugement dans le groupe recevant le nouveau traitement au sein d'un essai thérapeutique. Cette démarche n'incluant pas les patients du groupe contrôle est appelée « *treatment-only analysis* » et revient en réalité à chercher un facteur de risque d'événement et non à mettre en évidence une modification de l'effet du traitement (19). Elle ne permet donc en aucun cas la mise en évidence de marqueurs prédictifs et expose au risque de fausse identification d'un effet thérapeutique spécifique dans une sous-population de patients marqueur (+).

3.2.4 Tableau de synthèse

Le tableau ci-dessous présente la synthèse des démonstrations pouvant être apportées par les schémas de développement d'un test diagnostique associé à une thérapie stratifiée, ainsi que les niveaux de preuve permettant de qualifier ou non un test de compagnon et une thérapie de ciblée.

Ces appellations ne préjugent pas de la possibilité d'inscrire un test diagnostique au remboursement dans les circonstances où le test ne serait pas cliniquement utile au sens du présent guide.

Tableau 1. Démonstrations et niveau de preuve associés au schéma de développement choisi pour un test diagnostique associé à une thérapeutique stratifiée au regard des appellations de test compagnon et thérapie ciblée.

Schéma de développement choisi pour le test diagnostique	Démonstration(s) potentiellement apportée(s)	Niveau de preuve considéré pour les appellations de test compagnon et thérapie ciblée
Étude standard avec recherche <i>post-hoc</i> de « marqueurs prédictifs » n'incluant que les patients du groupe traité	Recherche de facteurs de risque du critère de jugement et non de marqueurs prédictifs de la réponse thérapeutique	Pas de conclusion possible
Recherche <i>post-hoc</i> en sous-groupes d'un marqueur prédictif dans une étude sur l'efficacité d'un nouveau traitement en population générale	<ul style="list-style-type: none"> - Efficacité ou non efficacité du traitement dans une population de patients non sélectionnés par un marqueur - Recherche exploratoire d'un marqueur prédictif dont l'utilité clinique du test diagnostique doit être confirmée dans le cadre d'un essai prospectif 	Pas de conclusion possible (En l'absence d'essai prospectif de confirmation)
Étude de « comparaison de stratégies »	<ul style="list-style-type: none"> - Efficacité du nouveau traitement chez les patients marqueur (+), ou chez tous les patients quel que soit leur statut pour le marqueur, supérieure à celle du traitement standard chez tous les patients quel que soit leur statut pour le marqueur - Supériorité de la stratégie basée sur le marqueur identifié par le test diagnostique évalué et l'administration du nouveau traitement aux patients Mq+, par rapport à la stratégie d'utilisation du traitement standard sans détermination du marqueur 	Pas de conclusion possible
Étude « ciblée » chez les patients marqueur (+)	Efficacité du nouveau traitement dans la sous-population de patients identifiés marqueur (+) par le test diagnostique évalué	Niveau de preuve faible à modéré en fonction du poids accordé aux éléments précliniques et cliniques précoces lors de l'évaluation
Étude d' « interaction marqueur/traitement »	<ul style="list-style-type: none"> - Validation de l'utilité clinique du test diagnostique évalué, incluant la démonstration de l'efficacité du nouveau traitement dans la sous-population de patients identifiés marqueur (+) par le test - Appellations de test compagnon et thérapie ciblée 	Niveau de preuve élevé

4. Evaluation économique

L'évaluation médico-économique d'un dossier revendiquant « un test compagnon associé à une thérapie ciblée » ne nécessite pas actuellement l'élaboration d'un guide méthodologique spécifique à ce type d'évaluation. Si la problématique de l'évaluation conjointe d'un test compagnon associé à une thérapie ciblée représente une nouveauté sur les marchés pharmaceutique et biotechnologique, il est largement reconnu que l'évaluation d'un test diagnostique ou d'un test de dépistage ne saurait se limiter à évaluer sa capacité à identifier la présence ou l'absence d'une maladie ou d'un facteur de risque. En effet, cette évaluation doit également apprécier l'utilité clinique du test, autrement dit son impact sur la santé des individus, impact associé notamment à la mise en œuvre d'interventions de santé. Ce cadre d'évaluation représente un standard pour l'évaluation économique d'un test diagnostique ou d'un test de dépistage. Ainsi, pour l'évaluation médico-économique conjointe test/traitement dans le cadre d'une revendication de « test compagnon associé à une thérapie ciblée », **les recommandations de la version actuelle du guide méthodologique de la HAS « Choix méthodologiques pour l'évaluation économique à la HAS » restent applicables. Néanmoins, des mises à jour** portant sur les spécificités des tests diagnostiques dits « compagnons » **sont envisageables**. Elles concernent **notamment** :

- la formulation de la problématique de l'**efficacité conjointe** du test compagnon et de la thérapie ciblée ;
- la typologie des tests diagnostiques (par exemple, fonction du caractère prédictif du marqueur orienté vers l'efficacité ou la toxicité du traitement) ;
- les choix structurants de la modélisation économique : détermination de la population cible, choix des comparateurs ;
- le périmètre des coûts associés aux tests dits « compagnons »⁵.

La HAS considère que la **démonstration de l'utilité clinique du test diagnostique est un pré-requis à l'évaluation de l'efficacité conjointe d'un test compagnon et de la thérapie ciblée associée** (cf. Tableau 1). En d'autres termes, une évaluation économique labellisée « Evaluation conjointe d'un test compagnon associé à une thérapie ciblée » correspond à une étude médico-économique fondée sur des données d'un niveau de preuve suffisant pour démontrer l'utilité clinique.

⁵ L'analyse conjointe d'un test compagnon associé à une thérapie ciblée implique la prise en compte de l'ensemble des coûts de réalisation des tests (quel qu'en soit le résultat).

Conclusion

Le concept de **médecine stratifiée** consiste en une approche thérapeutique encore récente s'appuyant sur des couples « **médicament/test diagnostique** ». Au sein de cette approche, le test diagnostique a pour objectif théorique de permettre de **sélectionner**, en fonction de leur statut pour un marqueur présumé prédictif, uniquement les patients chez lesquels le traitement est susceptible d'apporter un bénéfice et de limiter ainsi l'utilisation de ce traitement à ces seuls patients.

Les concepts de **thérapie ciblée** et **test compagnon** s'inscrivent dans le cadre de la médecine stratifiée mais sont beaucoup plus étroits et donc plus exigeants en termes de démonstrations requises. Ainsi, pour permettre de revendiquer qu'une thérapeutique stratifiée peut être qualifiée de ciblée ou qu'un test diagnostique peut être nommé test compagnon, la démonstration doit être apportée qu'en plus d'être efficace chez les patients porteurs du marqueur, le traitement ne l'est pas chez les patients ne le portant pas. Ces démonstrations sont par ailleurs les fondements de la validation de **l'utilité clinique du test**, c'est-à-dire de son aptitude à améliorer le devenir clinique des patients et apporter une valeur ajoutée en termes d'optimisation de la prise en charge. En l'occurrence, l'utilisation du médicament et la réalisation du test étant dans ce contexte indissociables, l'utilité clinique du test ne peut être évaluée que conjointement à celle de l'efficacité du traitement au sein d'un même essai. Seul le **design de type « interaction marqueur/traitement »** permet alors d'apporter conjointement ces démonstrations. Les autres schémas d'étude, où la détermination d'effet du nouveau traitement est limitée à la sous-population de patients marqueur (+), ne permettent de conclure qu'à l'efficacité (ou non) du traitement dans cette sous-population de patients, mais en aucun cas à l'utilité clinique du test.

Au regard de la HAS, lorsqu'un test diagnostique permet la sélection des patients afin de répondre aux **exigences formulées au sein de l'AMM** d'un traitement, l'évaluation pourra, en fonction des démonstrations apportées, aboutir à **deux types de conclusions** générales :

- si la **démonstration de l'utilité clinique du test n'a pas été apportée** :
 - le test sera considéré comme un **test diagnostique « classique »** permettant d'identifier la sous-population de patients dont la sélection est notifiée au sein de l'AMM. Bien que sans utilité clinique démontrée, la réalisation de ce test n'en sera pas moins indispensable d'un point de vue réglementaire afin de permettre l'utilisation du traitement dans le cadre de son AMM,
 - l'appellation « test compagnon » ne pourra pas être revendiquée. La thérapie ne pourra pas être considérée comme ciblée. En effet, la preuve de la valeur prédictive du marqueur n'étant pas apportée, la possibilité d'une efficacité chez tous les patients ne pourra pas être exclue,
- si la **démonstration d'utilité clinique du test est apportée** :
 - le test sera considéré comme un test diagnostique permettant d'identifier les patients dont la sélection est notifiée au sein de l'AMM,
 - le test pourra de plus être qualifié de « **test compagnon** » et la **thérapie** considérée comme **ciblée**.

Dans tous les cas, les performances diagnostiques du test évalué doivent assurer la fiabilité des résultats des études menées. Sa mise en œuvre doit prévoir des contrôles de qualité et respecter l'ensemble des normes de qualité en vigueur dans le domaine concerné.

Références

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 B, 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. Ferte 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

Annexe 1. Liste des tableaux

Tableau 1. Démonstrations et niveau de preuve associés au schéma de développement choisi pour un test diagnostique associé à une thérapeutique stratifiée au regard des appellations de test compagnon et thérapie ciblée.20

Annexe 2. Liste des figures

Figure 1 – Schéma d'étude d' « interaction marqueur/traitement »	15
Figure 2 – Schéma d'étude « ciblée »	16
Figure 3 – Schéma d'étude de « comparaison de stratégies »	18

Fiche descriptive

Intitulé	Descriptif
Méthode de travail	Guide méthodologique HAS
Date de mise en ligne	Avril 2014
Date d'édition	Uniquement disponible sous format électronique sur www.has-sante.fr
Objectif(s)	<p>Exposer une synthèse pragmatique des fondements théoriques et méthodologiques de l'évaluation par la HAS d'un test diagnostique permettant d'identifier un marqueur présumé prédictif associé à une thérapie stratifiée.</p> <p>Définir et exposer les preuves requises par la HAS pour la reconnaissance ou non des appellations « test compagnon » et « thérapie ciblée ».</p> <p>Définir et apporter des outils méthodologiques pour la démonstration de l'utilité clinique des tests diagnostiques concernés par l'évaluation.</p> <p>Informers du niveau de preuve accordé par la HAS aux schémas d'études des tests diagnostiques concernés par l'évaluation.</p>
Demandeur	Autosaisine HAS
Promoteur	Haute Autorité de santé (HAS), service évaluation des actes professionnels (SEAP)
Pilotage du projet	<p>Coordination SEAP : Carole GIRAUD, chef de projet, SEAP (chef de service : Michèle MORIN-SURROCA, adjoint au chef de service : Marc GUERRIER)</p> <p>Secrétariat : Christine MAYOL, assistante, SEAP</p> <p>Coordination externe : Michel CUCHERAT, consultant</p>
Recherche documentaire	<p>Carole GIRAUD, chef de projet, SEAP</p> <p>Michel CUCHERAT, consultant externe</p>
Auteurs de l'argumentaire	<p>Carole GIRAUD, chef de projet, SEAP, sous la responsabilité de Marc GUERRIER, adjoint au chef de service, SEAP</p> <p>Clémence THEBAUT, chef de projet, service évaluation économique et santé publique (SEESP) et Salah GHABRI, chef de projet, SEESP, sous la responsabilité de Catherine RUMEAU-PICHON, chef de service, SEESP</p> <p>Michel CUCHERAT, Consultant</p>
Relecture - Validation	<p>Version intermédiaire du document soumise pour avis (octobre - novembre 2013) :</p> <ul style="list-style-type: none"> - En interne : Commission d'évaluation économique et santé publique (CEESP), Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNEDiMTS), Commission de la Transparence - En externe : INCa, ANSM, DGOS, DGS, DSS, SFPT, SNITEM, LEEM, SIDIV, CNAM, UNCAM, CEPS <p>Validation par le Collège de la HAS : février 2014</p>
Autres formats	Pas d'autre format que le format électronique disponible sur www.has-sante.fr
Documents d'accompagnement	Guide méthodologique annexe intitulé « Test compagnon associé à une thérapie ciblée : Annexe scientifique » disponible sur www.has-sante.fr



Toutes les publications de l'HAS sont téléchargeables sur
www.has-sante.fr