

## Commission Évaluation Économique et Santé Publique

9 février 2021

### TRANSCRIPTION DES DÉBATS

#### AVERTISSEMENT

En application des articles L. 1451-1-1 et R. 1451-6 du Code de la santé publique, la HAS réalise un enregistrement des séances de la commission de la transparence (CT), de la Commission d'évaluation des dispositifs médicaux et des technologies de santé (CNEDIMTS) et de la Commission évaluation économique et santé publique (CEESP). Pour en faciliter la communication et la compréhension, la HAS a fait le choix de recourir à une transcription des débats par l'intermédiaire d'une société prestataire.

Cette prestation associe une saisie directe des débats par sténotypie et une transcription assistée par ordinateur ainsi qu'une relecture médicale. L'objet de cette transcription est de permettre de tracer le déroulé des débats dans un souci de transparence et non de fournir une information scientifique validée. En effet, malgré le professionnalisme de cette prestation, il peut persister dans le texte final des incongruités ou des inexactitudes liées à l'usage d'un vocabulaire hautement spécialisé ou à la nature même des échanges verbaux. La HAS n'effectue aucune validation de ces documents.

La HAS rappelle que les seuls documents validés et opposables sont le procès-verbal de la séance et l'avis définitif de la Commission qui sont mis en ligne sur le site de la HAS.

Pour la publication des transcriptions, et dans un but de protection du secret industriel et commercial, certains mots peuvent avoir été occultés. Les occultations éventuelles sont de la responsabilité de l'entreprise exploitant le produit évalué.

Toute reprise d'un ou plusieurs extraits d'une transcription doit être accompagnée d'une mention en précisant la source et respecter la législation sur la publicité.

Les membres des commissions s'expriment à titre personnel dans le cadre de leur mission d'expertise. Les agents de la HAS (chefs de service, adjoints, chefs de projet) représentent l'institution et s'expriment en son nom.

La HAS rappelle que la connaissance des propos tenus en séance par les membres des commissions et les agents de la HAS ne peut en aucun cas justifier des contacts directs de quelque nature que ce soit avec ces personnes, lesquelles sont tenues à une obligation de confidentialité conformément à l'article R. 161-85 du code de la sécurité sociale.

### **Audition ECO-EFFI 420 : KAFTRIO (Vertex Pharmaceutical)**

**Un chef de projet du SEESP.-** Nous allons très rapidement vous donner le contexte avant que l'industriel n'arrive.

**Un chef de projet du SEESP.-** Bonjour tout le monde. Je vous partage les diapositives. Nous vous avons présenté en janvier l'avis de KAFTRIO, avec notre autre chef de projet et les rapporteurs Emmanuelle Fourneyron et Hassan Serrier. KAFTRIO est indiqué dans la mucoviscidose chez les patients âgés de 12 ans et plus, homozygotes pour la F508-del ou hétérozygotes à fonction minimale.

Le laboratoire a déposé son dossier en primoinscription. Il avait revendiqué un SMR important et une ASMR II, avec un prix revendiqué de [REDACTED] euros pour un mois de traitement. À cela, il faut associer le prix de KALYDECO, qui est associé dans l'indication.

En janvier, nous avons voté pour une réserve majeure sur la méthode de valorisation des utilités, et pour sept réserves importantes. L'industriel nous a demandé de lever la réserve majeure et de la déclasser en réserve importante. C'est ce qui fera l'objet de l'audition tout à l'heure. Il nous a également demandé un déclassement d'une réserve importante en réserve mineure pour une réserve sur l'efficience et une réserve sur l'impact budgétaire.

Si nous nous concentrons sur cette réserve majeure, elle portait notamment sur la méthode d'estimation de l'utilité, considérée non conforme. Si nous nous intéressons un peu à ce que le laboratoire nous avait présenté, ils ont tout d'abord présenté ce qu'ils appellent une utilité liée au bénéfice respiratoire. Ils se sont basés sur une étude faite au Royaume-Uni chez 401 patients âgés de 18 ans et plus, avec des scores d'utilité EQ-5D et ils ont appliqué ces scores d'utilité EQ-5D à des catégories de patients.

On voit ici le VEMS supérieur à 70, entre 40 et 70 ou inférieur à 40. Ils ont appliqué des utilités à différents niveaux de sévérité de la maladie, et en considérant que le VEMS, qui correspond au volume expiratoire maximum par seconde, est un caractère purement respiratoire, ils supposent que les valeurs d'utilité EQ-5D ne permettent que de mettre en avant les bénéfices respiratoires du traitement, mais ne considèrent pas la perception de la santé, les vitalités, les interactions sociales, tout ce qui est autre et différent de la fonction respiratoire.

De là, le laboratoire décide de nous présenter un incrément d'utilité qu'ils ont calculé à partir d'un questionnaire spécifique CFQ-R-8D. Ils appellent cet incrément d'utilité comme lié au bénéfice cette fois-ci non respiratoire. Si nous nous intéressons un peu aux différentes dimensions que l'on retrouve dans ce questionnaire spécifique, nous voyons qu'il y a la fonction

physique, la fonction neurologique, l'émotion, la vitalité, la difficulté respiratoire, la douleur abdominale, l'image corporelle et le fait de tousser.

Déjà, nous voyons bien que nous retrouvons des domaines liés à l'aspect respiratoire, donc ce n'est pas tout à fait non lié au domaine respiratoire, et surtout le laboratoire décide d'appliquer un incrément d'utilité calculé à partir de cette échelle spécifique uniquement pour le traitement KAFTRIO, puisqu'il considère que l'effet de traitement ressort significatif seulement sur KAFTRIO versus meilleur soin de support.

Nous nous posons la question du risque de double comptage, car déjà, dans l'échelle EQ-5D, on retrouve beaucoup d'éléments sur la qualité avec tous les domaines que cela comporte, et pas seulement du domaine respiratoire. Voilà ce que le laboratoire a fait.

Pour résumer ce pour quoi nous avons voté pour une réserve majeure, c'était par rapport à :

- cette méthode d'estimation de l'utilité non conforme faite sur une définition d'un incrément d'utilité issu d'un questionnaire de qualité de vie spécifique sans mapping,
- l'application de cet incrément en sus des scores EQ-5D.

L'incrément n'est appliqué qu'aux patients recevant KAFTRIO, et en plus de cela on remarque que le RDCR augmente de 30 % quand on n'applique pas cet incrément.

En plus de ne pas être acceptable en analyse de référence, cette hypothèse est largement en faveur du traitement évalué. Je crois que nous avons à peu près balayé la présentation concernant la réserve majeure. Veux-tu compléter ?

**Un chef de projet du SEESP.-** Non, je pense que pour le moment nous allons nous arrêter là. De toute façon, l'industriel va nous représenter certains éléments.

**M<sup>me</sup> PARIS.-** D'accord, donc nous quittons cette réunion et nous nous rendons sur l'autre lien pour les recevoir.

*(Les représentants de l'industriel rejoignent la séance.)*

**M<sup>me</sup> PARIS.-** Je me présente, je suis Valérie Paris, je préside cette commission et cette réunion. Si vous le voulez bien, je vous laisse vous présenter et nous faire part des observations que vous voulez partager avec nous.

**M. WESTERLOPPE.-** Merci, Madame Paris. Laura Catella, qui est avec nous, va agir sur les diapositives et c'est elle qui va les présenter. Je vais commencer l'introduction tout de suite pendant que les diapositives chargent.

Merci, Madame la Présidente et bonjour Mesdames et Messieurs membres de la commission. Nous vous remercions du temps que vous allez nous accorder aujourd'hui pour cette audition relative à nos spécialités KAFTRIO et KALYDECO. Nous avons sollicité cette audition pour soutenir notre demande de requalification de la réserve majeure mentionnée dans le projet d'avis du 12 janvier dernier, et qui porte à la fois sur la méthode d'estimation de l'utilité, l'utilisation d'un incrément et le choix de l'analyse de référence.

Pour couvrir cette demande, nous sommes venus à trois. Comme j'ai la parole, je me permets de me présenter en tout premier. Je m'appelle Jérémie Westerloppe et je suis en charge, au sein de la filiale française de Vertex, du département accès marché. Je suis venu accompagné de Laura Catella. Je te laisse te présenter.

**M<sup>me</sup> CATELLA.-** Bonjour à tous. Je travaille également au sein du laboratoire Vertex et j'étais en charge de la préparation du dossier d'efficience de KAFTRIO.

**M. WESTERLOPPE.-** Enfin, nous sommes venus accompagnés du Professeur Pierre-Régis Burgel, qui en sa qualité de clinicien ayant une importante expérience en matière de prise en charge de la mucoviscidose, nous rejoins. Professeur Burgel ?

**M. le P<sup>r</sup> BURGEL.-** Bonjour à toutes et à tous. Je suis pneumologue PUPH à l'hôpital Cochin. Je m'occupe de mucoviscidose depuis plus de 18 ans et je suis responsable d'un centre qui suit 600 malades atteints de mucoviscidose.

**M. WESTERLOPPE.-** Merci, Professeur. La problématique à laquelle nous avons été confrontés lors de la construction du modèle d'efficience est qu'au travers de nos diverses expériences, il nous a paru nécessaire de compléter les mesures issues de l'EQ-5D dans la mucoviscidose afin d'être en capacité de mesurer une éventuelle amélioration de la qualité de vie des patients sous traitement.

Pour couvrir ce point, nous vous proposons de suivre le plan suivant, qui sera :

- dans un premier temps d'établir avec vous le diagnostic qui est celui d'un manque de sensibilité d'EQ-5D dans la mucoviscidose ;
- dans un deuxième temps de vous présenter la solution qui aura été développée avec rigueur et méthode ;
- et enfin, d'apprécier l'incertitude qui pourrait découler d'une telle approche.

Pour nous, le premier élément de diagnostic est clair et important. C'est l'absence de cohérence entre les valeurs de qualité de vie issues d'EQ-5D et ce qui est rapporté par les patients ou la

communauté scientifique. Ce constat a été établi sur la base d'études de plus de 1 000 patients souffrant de mucoviscidose et qui ont été intégrés dans nos essais cliniques.

Dans l'étude TRAFFIC et TRANSPORT qui regroupe 997 patients souffrant de mucoviscidose, ce tableau que vous voyez ici avance un niveau d'utilité EQ-5D de l'ordre de 0,91. C'est un niveau très élevé d'utilité, proche d'un état de santé parfait, et comparable à l'utilité de la population générale.

Cette mesure suggère en fait une absence de détérioration de la qualité de vie par rapport à la population générale. Ce n'est pas du tout en accord avec l'impact de la mucoviscidose sur la qualité de vie tel que rapporté par les malades ou la communauté scientifique. Pour aller au-delà des chiffres, je propose au Professeur Burgel de bien vouloir nous faire part de sa connaissance et de son expérience de la maladie et de son impact sur la qualité de vie des patients.

**M. le P<sup>r</sup> BURGEL.**- Très rapidement, le pronostic de la mucoviscidose s'est beaucoup amélioré sur les cinquante dernières années, mais en 2021 cela reste une maladie invalidante. C'est une maladie principalement respiratoire avec la toux, l'encombrement bronchique, l'expectoration, l'essoufflement, les douleurs digestives, les diarrhées.

C'est une maladie qui a un retentissement général. Nous avons des recommandations internationales sur le dépistage de l'anxiété et de la dépression de patients de façon annuelle dans ces populations. C'est une maladie qui nécessite une charge en soins importante, qui a été estimée chez les adultes de l'ordre de deux heures par jour entre les comprimés, les aérosols, la kinésithérapie, etc.

**M. WESTERLOPPE.**- Merci, Professeur Burgel. Nous avons montré en quoi la qualité de vie d'un patient souffrant de mucoviscidose peut être diminuée.

La deuxième étape de notre diagnostic a été d'essayer de comprendre les raisons de l'existence de tels scores si élevés et si proches de la population générale. L'analyse de la distribution des mesures d'utilité de l'EQ-5D nous aide à avancer dans cette compréhension. Sur le graphique de gauche, vous voyez la distribution des scores EQ-5D sur la population des 997 patients dans les essais TRAFFIC et TRANSPORT.

Ce que nous observons est simple, c'est un effet plafond très important. Vous le voyez sur le dernier graphique à droite, qui montre environ 607 malades qui présenteraient une utilité de niveau 1, c'est-à-dire une utilité d'un état de santé parfait. C'est un effet plafond très important, puisque nous avons là 61 % des patients qui seraient dans cet état de santé dit parfait.

Au regard de ce que nous savons, nous pouvons qualifier cette mesure d'anormale, voire d'aberrante. En effet, ce plafond peut expliquer les scores très élevés que l'on retrouve mais

surtout, et au-delà du fait que cela ne représente pas la réalité de la mucoviscidose, cet effet plafond a aussi un impact très important puisqu'il altère la capacité à mesurer une variation future de l'état de santé des patients. À partir du moment où vous avez déjà plus de 60 % des patients qui sont au niveau maximum, il va être dur de mesurer cette variation.

Restent deux questions face à cela. Pourquoi une telle situation ? Comment pouvons-nous la gérer ?

Concernant le « pourquoi », nous nous proposons d'étudier les mesures de qualité de vie avec un outil spécifique validé dans la mucoviscidose chez les patients rapportant un état a priori de santé parfait avec l'EQ-5D. Ce que vous avez ici représenté sur ce graphique, ce sont les résultats de qualité de vie tels qu'appréhendés par le CFQ-R-8D chez les 607 patients de l'essai clinique TRAFFIC et TRANSPORT qui présentent un état de santé dit parfait.

Nous voyons clairement que certains domaines importants de la maladie, comme ceux qui ont été évoqués par le Docteur Burgel, ne sont pas bien mesurés par l'EQ-5D. Nous observons notamment une dégradation importante de la qualité de vie, vous le voyez sur les trois dimensions de gauche. Vous avez la toux, où l'EQ-5D chez ces patients nous dit « état de santé parfait », et où vous voyez que 95 % des patients déclarent une dégradation de leur qualité de vie, idem sur la dimension vitalité, idem sur la dimension émotion. Je ne vais pas vous couvrir l'ensemble des dimensions.

In fine, nous avons un problème, celui de l'absence de cohérence entre le retour des patients ou de la communauté scientifique et les mesures produites par EQ-5D. Comment pouvons-nous le gérer ? C'est la question que nous posons avec pour objectif d'appréhender pleinement l'effet de la mucoviscidose sur la qualité de vie au travers d'une mesure appropriée, et d'être en capacité d'en mesurer les variations éventuelles.

Pour ce faire, je vais laisser la parole à Laura Catella, qui va nous présenter la démarche et la méthode que nous avons développée pour répondre au problème qui nous est posé.

**M<sup>me</sup> CATELLA.-** Merci beaucoup, Jérémie. Tout d'abord, l'objectif de ce travail était vraiment d'apporter une réponse adaptée et correcte à la problématique qui vient de vous être présentée. Pour cela, nous avons fait appel à l'Université de Sheffield pour développer un algorithme de scoring basé sur les préférences et spécifique à la mucoviscidose. Il a été développé à partir de la méthode standard. Il repose sur quatre étapes que je vais vous présenter.

La première est la sélection des dimensions de santé et des questions spécifiques à partir du questionnaire de qualité de vie CFQ-R, qui est validé dans la mucoviscidose, pour développer un système de classification d'états de santé. Dans ce travail, il va s'agir de 32 états de santé qui ont

été dérivés de 8 domaines sélectionnés puis validés par un panel de patients et d'experts cliniques.

La deuxième étape est le classement et la valorisation de ces 32 états de santé avec la méthode usuelle du time trade-off, qui a été réalisée au cours d'entretiens auprès de 345 personnes.

La troisième étape était le développement de l'algorithme de scoring, avec les modèles de régression pour permettre l'attribution d'une utilité à chacun des patients.

Enfin, la quatrième étape est de s'assurer de la validité de l'algorithme, et cela a été fait au travers de la comparaison des performances du CFQ-R-8D aux échelles de mesure génériques EQ-5, mais aussi SF-6D, et cela a été fait à l'aide d'une série d'analyses psychométriques.

C'est l'ensemble de ce travail qui a permis de développer l'échelle de révélation des préférences de façon rigoureuse, et qui nous permet une mesure plus appropriée de l'utilité dans la mucoviscidose. C'est ce que je veux vous montrer dans la prochaine diapositive.

Vous avez ici l'un des résultats de la comparaison des performances. C'est la distribution des scores des deux échelles, EQ-5D et CFQ-R-8D. À droite, nous retrouvons l'effet plafond de l'EQ-5D, qui est égal à 1, tandis qu'avec le CFQ-R-8D nous observons la disparition de cet effet plafond, qui va conduire à une répartition plus représentative de la qualité de vie dans la mucoviscidose. C'est cela qui va rendre possible la mesure d'une éventuelle amélioration de l'utilité, et in fine la construction d'un éventuel incrément d'utilité. Ce sont les résultats que je vous présente dans la prochaine diapositive.

Comment cela a-t-il été réalisé ? Nous avons appliqué le CFQ-R-8D aux résultats de nos études cliniques, et cela nous a permis d'estimer l'utilité pour l'ensemble des patients et de calculer un différentiel entre le traitement expérimental d'une part et le contrôle, qui correspond à l'utilité liée à l'effet de traitement uniquement. Je précise cela car c'est important, ce différentiel lié au traitement a été ajusté sur la variation du VEMS, et l'utilité est donc indépendante de cette variation du VEMS.

En termes de résultat, cela se traduit par un différentiel significatif uniquement chez les patients traités par KAFTRIO. Nous avons aussi réalisé l'analyse chez les patients traités par SYMKEVI et ORKAMBI, cependant nous n'avons pas eu de différence significative dans ces deux essais cliniques.

C'est pourquoi dans notre modèle nous appliquons l'incrément d'utilité uniquement chez les patients traités par KAFTRIO. Cet incrément est de 0,078 et il est ajouté à l'utilité de l'EQ-5D, qui est rapportée dans la littérature dans une population avec une prise en charge standard, et qui

varie de 0,54 à 0,74 selon le niveau de VEMS, ce qui fait que l'incrément lié à l'effet de KAFTRIO représente environ 10 % de la valeur basale d'utilité.

J'aimerais maintenant laisser la parole au Professeur Burgel pour compléter cette différence d'amélioration observée avec KAFTRIO et avec SYMKEVI et ORKAMBI par rapport à son expérience.

**M. le P<sup>r</sup> BURGEL.**- Merci. Je veux juste vous faire part de notre expérience clinique, pas tellement avec SYMKEVI, que nous n'avons pas beaucoup utilisé, mais avec ORKAMBI, qui est disponible depuis 2016 et qui était en soi un progrès par rapport aux traitements symptomatiques de la mucoviscidose, mais un progrès mesuré avec un gain de fonction respiratoire modéré, un gain de poids modéré et une réduction des poussées de la maladie, ce que l'on appelle des exacerbations.

L'histoire avec KAFTRIO est très différente. Nous avons maintenant 400 patients en France traités en ATU par KAFTRIO, et nous voyons non seulement des gains de fonction respiratoire très importants, des gains de poids très importants, mais surtout une modification de l'histoire naturelle de la maladie puisque nous avons des patients insuffisants respiratoires qui ont été sevrés d'oxygène de ventilation et des discussions de transplantation pulmonaire qui ont été interrompues, suspendues, des patients qui étaient déjà sur liste de greffe pulmonaire et qui en ont été retirés. Nous ne sommes donc pas très surpris de ces résultats médicoéconomiques qui rejoignent assez bien ce que nous voyons en pratique clinique.

**M. WESTERLOPPE.**- Merci Professeur Burgel et Merci Laura.

Maintenant que nous avons vu ensemble le diagnostic ainsi que la solution développée, il nous reste à étudier l'impact de cette mesure sur le RDCR. Cette approche plus appropriée pour la mesure de la qualité de vie des malades de la mucoviscidose a un effet sur le RDCR. Ne pas intégrer une mesure juste de la qualité de vie détériore le niveau de RDCR, puisque l'on passe d'un RDCR de 650 000 euros en utilisant l'incrément d'utilité à un RDCR de 845 000 euros sans utiliser cet incrément d'utilité. Il y a donc bien une évolution du RDCR, qui cependant reste de même nature, ou disons du même ordre de grandeur.

Deuxièmement, se pose maintenant la question de la délimitation de l'incertitude sur la mesure. Du fait du caractère plus approprié de la mesure de la qualité de vie qui nous est proposée avec l'incrément, nous avons retenu en analyse de référence la méthode intégrant l'incrément d'utilité, ce qui signifie qu'il existe des analyses de sensibilité pour l'analyse de référence, correspondant à l'incrément d'utilité, mais pas pour le scénario sans l'incrément.

Cependant, l'ensemble des analyses en sensibilité qui ont été effectuées pour explorer l'incertitude dans l'analyse de référence, et qui varient entre -25 % et +30 %, permettent selon

nous d'anticiper l'incertitude inhérente à l'analyse en scénario par transposabilité. C'est ce qui nous amène à dire que nous sommes dans le cadre d'une incertitude délimitée.

En conclusion, nous espérons avoir partagé avec vous le diagnostic qu'utiliser uniquement l'EQ-5D dans la mucoviscidose ne laisse pas la place à la mesure de l'amélioration de la qualité de vie avec de nouveaux traitements, alors qu'il est important que la qualité de vie soit pleinement appréhendée comme rapportée par les patients et que soit prise en compte celle qui est mesurée dans nos essais cliniques.

Enfin, nous espérons aussi vous avoir convaincus du caractère rigoureux de la méthode développée pour trouver une solution au problème rencontré, à savoir une échelle spécifique CFQ-R-8D qui a été développée et validée de manière rigoureuse et qui, malgré les limites méthodologiques en sus de l'EQ-5D, permet une mesure plus appropriée de la qualité de vie des patients souffrant de mucoviscidose ainsi que de son évolution.

Nous espérons enfin vous avoir montré, grâce aux analyses de sensibilité sur l'analyse de référence, que nous sommes en capacité de délimiter l'incertitude sur le RDCR par transposabilité, nous assurant ainsi que le niveau de RDCR observé reste peu ou prou du même ordre. C'est sur cette base que nous sollicitons auprès de la commission la requalification de la réserve majeure en réserve importante. Nous vous remercions de votre attention.

**M<sup>me</sup> PARIS.-** Merci beaucoup pour cette présentation très claire. Les membres de la commission ont-ils des questions ?

**Un chef de projet du SEESP.-** Bonjour et merci pour votre présentation. J'ai simplement une question pour avoir un peu plus d'explications sur les différentes dimensions que comporte le questionnaire spécifique. N'y aurait-il pas un risque de double comptage de certaines dimensions entre l'EQ-5D et le CFQ-R-8D ?

**M. WESTERLOPPE.-** Je vous remercie de votre question, qui est plus que justifiée. Je vais laisser Laura vous apporter les éléments de réponse.

**M<sup>me</sup> CATELLA.-** Merci. Je suis revenue sur la slide avec les différents domaines du score CFQ-R-8D. Ce sont des domaines issus des 12 domaines du questionnaire CFQ-R, qui lui est validé dans la mucoviscidose et qui est utilisé depuis plusieurs années par la communauté.

Nous avons 7 domaines qui ont été dérivés et 1 domaine qui a été scindé en deux, la toux et les difficultés respiratoires. J'espère que pour les domaines cela répond à la question, sinon je vous laisserai préciser un peu plus la question.

En ce qui concerne le risque de double compte, d'abord on peut dire qu'en effet il ne peut pas être éliminé à 100 %. Toutefois, nous avons développé une méthode qui nous permet d'envisager

que ce risque puisse être vraiment limité, d'abord parce que même si les échelles EQ-5D et CFQ-R-8D sont des échelles de qualité de vie globale, elles ne sont pas sensibles aux mêmes dimensions, comme le montre ce graphique. L'EQ-5D ne semble pas être adapté à la mucoviscidose, et c'est ce que nous espérons vous avoir démontré au cours de la présentation.

À cela s'ajoute le fait que l'incrément d'utilité n'est ajouté que dans le cadre d'une amélioration de la qualité de vie qui est liée au traitement par rapport à une prise en charge standard. Comme je l'avais expliqué, c'est indépendamment du niveau de VEMS alors que l'utilité EQ-5D est utilisée de façon stratifiée par le niveau de VEMS.

Pour répondre un peu plus à la question, je vais préciser comment cet ajustement du VEMS est réalisé. D'abord, les utilités CFQ-R-8D sont estimées pour chaque patient à chaque visite. C'est-à-dire que l'on peut avoir plusieurs mesures par patient. Elles sont estimées sur les données des différents essais pour SYMKEVI, ORKAMBI, KAFTRIO et pour le best supportive care.

Ensuite, nous avons fait des analyses post hoc qui ont été effectuées pour définir l'incrément d'utilité spécifique à ces traitements, et dans les modèles de régression utilisés, qui prennent en compte les mesures répétées pour chaque patient, il y a deux covariables qui sont ajoutées, à savoir le niveau de VEMS à chaque mesure et le traitement utilisé, traitement expérimental ou comparateur.

Le coefficient associé au niveau de VEMS permet de capter l'effet du niveau de VEMS sur l'utilité estimée et le coefficient associé au traitement par modulateur de CFTR permet d'estimer l'impact du modulateur du CFTR sur l'utilité mesurée par le CFQ-R-8D et ajustée sur le niveau de VEMS, c'est-à-dire l'effet de traitement seul, indépendamment du VEMS. C'est ce qui nous amène à considérer que le risque de double compte est limité dans l'analyse. J'espère que cela répond à votre question.

**M<sup>me</sup> PARIS.-** Merci. Hassan ?

**M. SERRIER.-** Je vais continuer un peu dans la même lignée, parce que je ne suis pas sûr de tout comprendre, je suis donc preneur de vos éclaircissements. Il y a deux choses qui me perturbent un peu. La première est ce qui vient d'être demandé sur le risque de double compte. Je ne suis pas sûr qu'il soit totalement pris en compte. Si je comprends bien, il y a quand même une qualité de vie qui est mesurée à un moment donné par l'EQ-5D, à laquelle on ajoute cet incrément qui est mesuré par un second outil.

Vous présentez les choses en disant que l'utilité ici est une utilité sur les aspects non respiratoires. Pour autant, le questionnaire inclut la toux, les difficultés respiratoires et autres. À la limite, même s'il ne les incluait pas, on parle d'utilité donc on peut prendre d'autres exemples, mais dans ce

questionnaire il y a l'émotion et dans le questionnaire EQ-5D, il y a l'anxiété et autres. Pouvez-vous nous en dire un peu plus sur cette distinction respiratoire/non respiratoire et sur ces éléments qui sont quand même dans les deux questionnaires ?

**M. WESTERLOPPE.-** Tout à fait, avec plaisir. Merci de nous permettre d'essayer de clarifier plus le propos. Le premier élément qu'il faut avoir à l'esprit, si vous vous souvenez bien du préambule, c'est qu'encore une fois il y a un diagnostic qui montre que nous avons un effet plafond majeur, puisque vous avez 61 % des gens qui sont a priori dans un état de santé parfait alors qu'ils souffrent de mucoviscidose.

C'est ce constat, ce diagnostic, qui nous a amenés à rechercher la solution que nous proposons pour pouvoir être en capacité de mesurer pleinement l'effet de la mucoviscidose sur la qualité de vie de ces patients et d'être en mesure d'appréhender une variabilité. C'est le premier point, parce que c'est un élément de diagnostic important.

Pour revenir encore plus précisément à votre question sur la dimension respiratoire, il y a peut-être une confusion. Quand nous avons parlé dans le rapport technique d'effet sur la dimension respiratoire, c'est parce que l'EQ-5D est utilisée au travers des utilités qui ont été estimées par Acaster stratifiées en fonction du VEMS, donc du niveau qui représente la fonction respiratoire. À trois stades de la fonction respiratoire sont attachés des niveaux d'utilité.

Ce sont ces niveaux, comme je vous le disais, qui nous montrent qu'il y a peu de variabilité entre les strates de VEMS, notamment entre les patients qui sont entre 70 et 90 et les patients qui sont entre 40 et 70.

Nous ne pouvons pas assurer qu'il y ait 0 % de doubles comptages, mais entre l'alpha et l'oméga il y a quand même une marge. Là où nous pensons que nous maîtrisons en grande partie ce problème du double comptage, c'est qu'au travers de la régression dont vous a parlé Laura, nous sommes bien sur une régression où d'un côté nous avons une mesure où l'on stratifie en fonction de la fonction respiratoire, et une autre partie qui dans l'équation est ajustée à la fonction respiratoire et au fait que l'on soit traité ou non traité par un modulateur de CFTR. C'est cette façon de séparer, cette dichotomie, qui nous permet d'espérer d'éviter ce double comptage.

Enfin, comme l'a présenté Laura dans son tableau, c'est un incrément. Cela ne se substitue pas à l'EQ-5D et à la mesure de l'EQ-5D. C'est bien un incrément qui nous permet d'avoir une vision sur un bénéfice et d'être en capacité de mesurer une évolution de la qualité de vie qui est vraiment retranscrite aussi dans les contributions des patients, où vous voyez bien qu'il se passe quelque chose de très important et qui est indépendant du VEMS. Nous sommes donc en capacité de mesurer cet incrément, et cela représente environ 10 % de l'utilité globale.

J'espère que cela éclaire quelque peu le point que vous soulevez.

**M<sup>me</sup> PARIS.-** Merci beaucoup. Y a-t-il d'autres questions des membres de la CEESP ou du service ? Non ?

*(La commission n'exprime aucune observation.)*

**M<sup>me</sup> PARIS.-** Je vous remercie pour votre présentation. Nous allons continuer les délibérations de la CEESP. Je vous souhaite un bon après-midi. Merci beaucoup. Au revoir.

*(Les représentants du laboratoire quittent la séance.)*

**M<sup>me</sup> PARIS.-** Quelles sont les premières réactions ? Je n'ai pas compris leur argumentation sur les doubles comptes. Par contre, j'ai bien compris pourquoi ils ont fait une échelle spécifique. J'ai une question pour vous qui avez l'habitude. Cet attroupement d'utilités à 1 est-il courant ? Ma première réaction, à leur place, aurait probablement été de me demander s'il n'y avait pas eu une erreur dans le traitement des données. Cela paraît même étrange qu'il y ait des utilités faibles, puis plus personne, puis tout d'un coup tout le monde à 1. Est-ce que finalement dans certains stades c'est une toux qui n'est pas si dérangeante ? Si ça se trouve, je dis une horreur pour les malades de mucoviscidose.

**Un chef de projet du SEESP.-** Nous avons rarement les résultats des questionnaires collectés présentés de la sorte. Nous avons toujours les résultats d'utilités moyennes, mais quand on regarde parfois certaines études et certaines publications, on a tendance parfois à observer des pics vers le 1, c'est vrai. Après, il est très difficile de comparer ce qu'ils observent ici à d'autres choses, puisque là ils nous parlent juste de leurs données. Après, il y a probablement d'autres chefs de projet ou membres de la commission qui ont vu ce genre de résultats d'études.

**M<sup>me</sup> PARIS.-** Je ne suis pas du tout frappée par le fait qu'une échelle spécifique soit plus sensible.

**Un chef de projet du SEESP.-** C'est le but.

**M<sup>me</sup> PARIS.-** Par contre, c'est plutôt l'agrégation vers le 1.

**Un chef de projet du SEESP.-** C'était la question du double compte des dimensions. Quand on regarde leur questionnaire spécifique par rapport à l'EQ-5D, on se rend bien compte que sur les trois dimensions supplémentaires, il y en a deux qui concernent la respiration et qui sont directement liées à la mucoviscidose. C'est vrai que l'EQ-5D, qui est un questionnaire générique, reste beaucoup plus général dans ces questions, mais les cinq domaines de l'EQ-5D sont retrouvés dans les huit domaines du questionnaire spécifique.

**M<sup>me</sup> PARIS.-** Lionel ?

**M. PERRIER.-** J'étais assez convaincu par le fait que le questionnaire CFQ-R-8D était, pour cette pathologie et pour ces patients, mieux adapté que l'EQ-5D. Par contre, du coup, mon interrogation porte sur le fait qu'ils aient adapté une méthode mixte, où finalement ils combinent le CFQ-R-8D et l'EQ-5D. Pouvaient-ils avoir une évaluation du RDCR basée uniquement sur l'EQ-5D et une évaluation basée uniquement sur le CFQ-R-8D, et libre à la CEESP de retenir éventuellement soit l'EQ-5D soit le CFQ-R-8D en analyse de référence ? Pourquoi ne l'ont-ils pas fait ?

**Un chef de projet du SEESP.-** Ce n'est pas du tout une chose qu'ils ont présentée ou à laquelle ils ont pensé que de faire une analyse qui n'utilise qu'un questionnaire spécifique. Ils nous le présentent vraiment comme un complément de l'EQ-5D, alors qu'il s'agit aussi d'une valeur d'utilité. Nous sommes donc vraiment sur un ajout. Oui, nous pouvons nous demander pourquoi il n'y a pas une analyse qui n'utilise que le CFQ-R-8D en complément, et peut toujours être proposée en analyse de sensibilité pour montrer l'impact de l'utilisation d'un questionnaire spécifique.

**M<sup>me</sup> PARIS.-** Hassan ?

**M. SERRIER.-** Une fois de plus, j'interviens derrière Lionel, qui a traduit ma pensée. Il faut bien remettre les choses dans le contexte. Je suis tout à fait d'accord avec l'argument lorsque l'on me dit que l'EQ-5D n'est pas sensible pour une pathologie, je peux l'entendre, donc il n'y a pas de problème avec cela. S'ils pensent que ce n'est pas sensible, ok. Je peux entendre qu'un questionnaire spécifique comme celui-là soit plus sensible et à la limite, s'agissant de la méthode qu'ils mettent en place par rapport au questionnaire, il n'y a pas de problème.

C'est effectivement derrière que cela pose problème. Si à la limite ils avaient fait comme l'avait dit Lionel, cela nous aurait posé problème en analyse principale parce que nous nous serions posé des questions en nous disant « ils pouvaient quand même utiliser l'EQ-5D de comparabilité et autres », mais là, en faisant cette espèce de méthode mixte où ils veulent ajouter un incrément qu'ils justifient comme étant non respiratoire versus l'EQ-5D, c'est totalement tordu et c'est aussi ce qui aboutit à cet incrément qui est non significatif du côté du comparateur, et qui n'est pas pris en compte. Même eux l'ont dit, ils essaient de limiter le risque de double compte, mais clairement il n'est pas annihilé. Il est toujours là. À partir du moment où il y a un risque de double compte même faible, cela invalide tout.

Je pense vraiment que là où ils ont été fautifs, c'est principalement là-dessus. Pourquoi n'ont-ils pas fait l'analyse pour l'utiliser en analyse principale à la place de l'EQ-5D ? D'autant que c'est une étude ad hoc. Cela aurait déjà été beaucoup plus propre. Pour moi, il n'y a eu aucune nouvelle

information par rapport à ce que nous savions déjà, et cela reste toujours impossible à utiliser en analyse principale.

**M<sup>me</sup> PARIS.-** Martine ?

**M<sup>me</sup> BUNGNER.-** Oui, merci. Je vais peut-être revenir un petit peu en arrière. Il y a une question que je n'ai pas bien comprise. Je ne voulais pas poser la question là, mais peut-être que vous pouvez me répondre.

Y a-t-il quelque part des hypothèses sur la raison pour laquelle il y a ce plafond ? Je pense que selon ce que l'on peut penser des raisons à ce plafond, on peut peut-être mieux comprendre pourquoi ils essaient d'ajouter les deux. D'une certaine façon, ce serait pour plutôt prendre en compte ce qui est masqué après le plafond, si je puis dire. Pour moi, a priori, cela limiterait les risques de double compte. En tout cas, cela les réduirait considérablement par rapport à ce que vous posez.

Je prends juste un exemple pour avoir fait pas mal d'entretiens avec des malades de mucoviscidose. La question de la toux n'est pas simplement, et même pas nécessairement majoritairement liée au problème respiratoire, mais beaucoup également aux autres problèmes d'encombrement d'autres viscères. Du coup, on ne peut pas simplement l'imputer totalement aux problèmes respiratoires. Je trouve que nous allons peut-être un peu vite dans le fait de dire très clairement que ce double compte est massif. Moi, j'aurais plutôt tendance à penser que ce n'est pas si important que cela. Je trouve que c'est dommage, parce qu'il y avait Pierre-Régis Burgel et il aurait pu expliquer cela, [REDACTED]

Est-ce que quelqu'un a une réponse sur le plafond et sur les 40 % qui resteraient ? Je trouve que le graphique que la fille a proposé, qui montre bien ce qu'ils ont essayé de prendre en compte avec un score ad hoc, joue bien sur l'hypothèse qu'il fallait compléter et pas simplement

**M<sup>me</sup> PARIS.-** Nous continuons les interventions, et notre chef de projet essaiera de répondre à Martine si elle le peut. Catherine ?

**M<sup>me</sup> LE GALES.-** Merci.

En les écoutant, il y a quelque chose qui m'a un peu troublée, et je me demandais si c'était dans le dossier. C'est qu'effectivement, ce questionnaire CFQ-R est spécifique, il y a d'ailleurs eu des tentatives de mapping avec l'EQ-5D, etc., mais en fait ils ne prennent pas ce questionnaire. Ils prennent le questionnaire 8. Or, derrière le 8, cela veut dire qu'ils ont sélectionné certaines dimensions parmi les 12 dimensions du questionnaire initial spécifique à la mucoviscidose. C'est un point que je n'avais personnellement pas appréhendé jusqu'à présent.

Je comprends donc qu'ils ont sélectionné 8 dimensions sur un questionnaire spécifique validé, mais du coup, est-ce que cela veut dire que cette espèce de sous-questionnaire a fait l'objet d'une publication pour comprendre quel était l'intérêt ou pas de se limiter à ces 8 dimensions et ce que cela permettait d'appréhender plus finement que ne pouvait le faire le questionnaire avec 12 dimensions ? Effectivement, si on ne comprend pas la démarche initiale de sélection des dimensions, je ne comprends pas bien comment l'on peut justifier qu'il n'y a absolument pas de double compte. C'est mon premier point.

Mon deuxième point est hyper technique, je suis désolée. Ce n'est pas parce que l'on a 4 degrés de différence sur une échelle de 0 à 1 et 4 degrés de différence sur une autre échelle de 0 à 1 que les deux valeurs de 4 valent la même chose. Ce que je veux dire par là, c'est que ce sont des problèmes de métrique. Je ne vois pas comment on peut faire une somme simple de scores cardinaux venant de deux échelles qui sont totalement indépendantes, ou en tout cas qui ont été développées de manière séparée.

En ajoutant le score d'utilité EQ-5D au score d'utilité issu d'une échelle spécifique, je ne comprends pas du tout ce que l'on fait. Ce n'est pas parce que ce sont des chiffres que pour autant on peut additionner les chiffres comme cela. Je suis vraiment désolée, mais j'ai passé suffisamment de temps sur ces échelles d'utilité pour savoir que l'on ne peut pas faire des choses comme cela aussi simplement.

J'ai l'impression qu'il y a des problèmes métrologiques massifs qui ne sont absolument pas appréhendés par cette équipe, et j'aimerais bien savoir si nous avons une publication scientifique dans laquelle nous aurions eu ce genre d'approche hybride validée par des gens ayant regardé si sur le plan métrologique cela fonctionnait ou non.

J'ai l'impression que l'on est en train de donner une magie aux chiffres. Cela reste un phénomène magique mais pas un phénomène sur lequel nous pouvons être complètement confiants. J'ai donc deux questions. Sur ce questionnaire dans lequel on prend 8 dimensions sur les 12, avons-nous une publication expliquant quel est l'intérêt de prendre 8 dimensions parmi les 12 dimensions d'un questionnaire spécifique ? Deuxièmement, est-ce que l'étude dans laquelle ils ajoutent des scores d'utilité issus de leur sous-questionnaire spécifique à des scores d'utilité issus de l'EQ-5D, est validée par des gens qui ont vraiment regardé ce qui était fait derrière sur le plan mathématique ? Pour moi, il y a vraiment de grosses difficultés. C'est tout le problème ces méthodes hybrides, sur lesquelles nous sommes assez circonspects et à juste titre. Avez-vous des informations là-dessus ?

**Un chef de projet du SEESP.-** Merci pour ces éléments, Catherine.

Déjà, sur le premier point, le questionnaire spécifique 8D n'était pas encore publié quand ils nous ont déposé leur dossier, mais il était en cours de publication. Je n'ai plus le détail en tête, mais il y a forcément des explications sur la sélection de ces 8 dimensions, par rapport à toute l'analyse qui est faite pour les transformer en valeur d'utilité.

Ensuite, sur le deuxième point, c'est vrai que c'était la grosse question que nous avions sur l'utilisation de deux résultats différents que l'on additionne tout simplement. À notre connaissance, il n'y a pas de validation de cette méthode qui existerait ou qui serait publiée quelque part. C'est un point qui est simplement décrit dans l'analyse critique et qui est remis dans le libellé de la réserve sur le fait qu'aujourd'hui cette méthode n'est pas validée. S'agissant de la méthode, nous parlons bien de l'utilisation de deux valeurs issues de deux questionnaires différents que l'on additionne simplement.

**M<sup>me</sup> PARIS.-** D'accord. Jacques ?

**M. le D<sup>r</sup> ORVAIN.-** J'ai essayé de comprendre un peu comment ils avaient pu faire cela. Je crois que ce qui rentre en compte aussi, c'est le time trade-off. J'imagine qu'ils ont dû faire une équation de régression où ils prédisent le time trade-off par deux choses, qui sont soit les classes de VEMS ou peut-être carrément le VEMS directement, et leur questionnaire, ce qui leur permet de dire qu'effectivement, dans une certaine mesure, ils ont travaillé sur le résiduel par rapport à ce qui était déjà expliqué dans le VEMS.

C'est ce qui expliquerait qu'il n'y a pas vraiment beaucoup de doubles comptes, sauf sur des considérations un peu compliquées de droites de régression qui ne seraient pas les mêmes selon le VEMS, mais je ne sais pas s'ils ont fait cela. C'est comme cela que j'aurais fait. Il faudrait regarder l'article de Sheffield pour voir s'ils ont fait cela, mais c'est ce qui pourrait avaliser le fait qu'ils n'ont peut-être pas énormément de doubles comptes. Je ne sais pas ce qu'ils ont fait exactement, mais c'est comme cela que je l'aurais fait.

**M<sup>me</sup> LE GALES.-** Est-ce que je peux juste reprendre la parole sur cette histoire d'ajout ?

Pour prendre une image un peu triviale, mon inquiétude est que l'on soit en train d'ajouter des degrés Celsius à des degrés Fahrenheit, pour dire les choses un peu simplement. C'est-à-dire que je raisonne en degrés Celsius, il y en a un qui est à 35 degrés et l'autre qui est à 45 degrés, et je vais chercher 5 degrés Fahrenheit que j'ajoute à mes 35 degrés. C'est ce dont j'ai peur.

**M. le D<sup>r</sup> ORVAIN.-** Oui, mais s'ils ont utilisé le time trade-off derrière, que tu mettes des Celsius ou des Fahrenheit, je peux faire cela dans une équation de régression et il n'y a aucun problème.

**M<sup>me</sup> LE GALES.-** Cela ne suffit pas, parce qu'il y a des problèmes de bornage et des problèmes d'intervalle, etc. Cela ne suffit pas du tout. Le zéro n'est pas obligatoirement au même endroit.

Le 1 n'est pas obligatoirement au même endroit. Il faudrait quand même que nous ayons le document pour être sûrs qu'ils n'ont pas ajouté du Celsius à du Fahrenheit.

**M. le D<sup>r</sup> ORVAIN.-** Après ce n'est pas grave si tu mets des coefficients de régression qui sont différents, selon que tu utilises du Fahrenheit ou du Celsius. Ce qui peut avoir des conséquences plus compliquées, ce sont des droites de régression partielles avec des intercepts différents selon le niveau de VEMS. C'est peut-être là que tu peux avoir du double compte, mais là nous entrons vraiment dans des équations de régression un peu complexes. Sinon, mathématiquement, cela tient la route, maintenant je ne sais pas ce qu'ils ont fait.

**M. SERRIER.-** Si je peux entrer dans la discussion, Catherine tu as pris l'exemple des degrés Fahrenheit et des degrés Celsius. Si l'on prend des mètres et des pouces, peu importe, si l'on a simplement deux outils de mesure qui ont juste une mesure différente, alors effectivement en mesurant les deux et en calculant un incrément d'un côté, on peut traduire.

Là, la difficulté est que ce que l'on mesure même est différent. C'est-à-dire que dans un cas, potentiellement, on va prendre en compte l'utilité liée à l'émotion, puis un peu à la respiration, puis à cela, etc. On arrive à un certain niveau, et on veut ajouter une dimension, mais de l'autre côté on ne sait pas non plus où est la frontière donc on ajoute encore autre chose avec encore autre chose avec encore autre chose, avec des chevauchements au milieu.

Comme essaie de l'expliquer Catherine depuis tout à l'heure, nous ne sommes même pas sûrs que le total soit comparable, ou en tout cas proportionnellement, et en plus nous avons des dimensions qui se chevauchent. Pour passer de l'un à l'autre, et même eux le présentent comme cela, c'est censé être quelque chose de différent mais en réalité cela se chevauche un peu.

**M. le D<sup>r</sup> ORVAIN.-** S'il y a une équation de régression sur le time trade-off, tu expliques ton time trade-off.

**M. SERRIER.-** C'est ce qu'ils ont fait. Ils ont fait une équation de régression sur le time trade-off pour avoir les scores sur le questionnaire. Sur cela, il n'y a pas de problème. Ok, ils ont leur score sur le questionnaire, sauf qu'ils se sont basés sur des dimensions différentes et plus ou moins nombreuses, donc on n'arrive pas forcément au même score.

**M. le D<sup>r</sup> ORVAIN.-** J'ai cru entendre quand même qu'ils disaient que le VEMS entrait dans l'équation de régression. C'est pour cela.

**M. SERRIER.-** Bien sûr, mais quelque part cela vient encore ajouter au flou. Concrètement, que corrige le VEMS dans toutes ces dimensions ? Est-ce qu'il y a une étude qui a concrètement montré concrètement ce que vient corriger le VEMS ? Est-ce que le VEMS vient corriger la toux ? Nous ne le savons pas non plus.

Je suis assez d'accord avec Catherine. À avoir tout mélangé comme cela, finalement nous ne savons pas ce que cela donne. Imaginez deux secondes que ce coefficient, même sur l'autre questionnaire, va de 0 à 1. Ils n'ont pas utilisé le « 1 » correspondant dans le nouveau questionnaire. Cela veut dire que tous les patients remplissent bien les fonctions qui ont été définies et les caractéristiques qui ont été définies. Dans l'EQ-5D, ce sont des fonctions totalement différentes.

Vraiment, l'argument de Catherine me parle beaucoup. C'est le problème d'ajouter des choses qui ne sont pas sur des mêmes échelles et qui ne sont pas sur des mêmes bases. Si l'on ajoute à cela le fait qu'il y a un peu de chevauchements, de doubles comptes dans l'utilité, cela donne quelque chose d'ininterprétable.

En plus, du fait de jouer sur la différence pour ajouter un incrément, ils observent un incrément qui n'est pas significatif dans le bras normal et ils ne le prennent pas en compte, ce qui ajoute encore un élément en faveur de leur produit. Je trouve donc que tout cela est difficile à interpréter.

**M<sup>me</sup> LE GALES.-** Ce que je trouve dommage, c'est qu'ils n'aient pas fait une analyse avec leur propre niveau d'utilité et qui soit indépendante de l'EQ-5D, parce que cela aurait vraiment été extrêmement intéressant.

J'entends l'argument qui consiste à dire que si l'on a un questionnaire de qualité de vie spécifique bien fait avec des mesures d'utilité on va réussir à avoir des informations que l'on ne captera pas avec une échelle générique. C'est le principe même.

C'est le fait de vouloir tout mélanger qui pose problème. Il me semble que s'ils avaient fait une analyse de référence avec l'EQ-5D et qu'ensuite ils nous avaient fait une analyse complète à côté avec leur échelle, nous aurions pu dire des choses et nous nous serions vraiment servi de leur travail.

À vouloir tout mélanger avec comme argument final celui de dire qu'en fin de compte il n'y a pas vraiment de différence que l'on ait un QALY à 600 000 ou un QALY à 800 000, je me dis que manifestement, on a des problèmes d'échelle de valeurs.

**M<sup>me</sup> PARIS.-** Oui, c'est vrai que le « même ordre de grandeur » à 30 % était une remarque intéressante.

Avez-vous eu le temps de prendre connaissance avant ou pas vraiment du document que mentionne notre chef de projet dans le fil de discussion ? Est-ce de nature à changer la façon dont vous regardez les choses ? Est-ce qu'il faut que nous nous donnions un peu de temps ? Est-ce qu'il

faut que nous nous décidions aujourd'hui ? Je sais que normalement nous sommes censés voter aujourd'hui.

**Un chef de projet du SEESP.-** Ce sont des choses que nous avons à disposition au moment de l'analyse critique. C'est simplement que la publication en tant que telle n'était pas encore sortie, mais ils nous ont apporté tout le niveau de détail qu'ils pouvaient apporter dans leur rapport technique.

**M<sup>me</sup> PARIS.-** D'accord. Sylvain ?

**Un chef de projet du SEESP.-** J'ai juste une petite remarque. Nous sommes d'accord qu'ils utilisent l'EQ-5D à 3 niveaux de question ?

**Un chef de projet du SEESP.-** Oui.

**Un chef de projet du SEESP.-** Il a été remplacé par le 5L, parce qu'on sait qu'il a cet effet plafond, qui biaise l'ensemble des résultats. Du coup, pourquoi n'utilisent-ils pas le 5L ? C'est peut-être une question un peu naïve.

**Un chef de projet du SEESP.-** Dans les données dont ils disposaient et qu'ils ont utilisées dans leur modèle, ils ont pris une étude anglaise qui avait des valeurs avec l'EQ-5D 3L.

**Un chef de projet du SEESP.-** Du coup, quitte à refaire un indicateur, est-ce que le plus simple n'aurait pas été de prendre le 5L pour pallier cet effet plafond plutôt que d'aller développer un nouvel outil ? Je trouve un peu étrange d'aller développer un outil alors qu'on a toute une littérature sur cet effet plafond du 3L. C'est juste un petit commentaire.

**M. SERRIER.-** C'est juste que leur argument pour justifier cet effet plafond était que l'une des dimensions n'était pas prise en compte. C'est pour cela qu'ils cherchaient plutôt un outil qui permettait de capter cette dimension. Je pense que c'est pour cela qu'ils n'ont pas voulu aller vers le 5L.

**M<sup>me</sup> PARIS.-** Bien. Lionel ?

**M. PERRIER.-** Je pense que la méthode mixte n'est pas recevable. C'est mon avis personnel. Dans la conclusion, nous avons noté les deux possibilités de valeur, 845 000 et 853 000. Si l'industriel avait produit ces résultats uniquement sur le CFQ-R-8D, et à condition que ce questionnaire ait été validé dans la littérature, nous aurions pu ajouter éventuellement ces résultats dans la conclusion et dire que globalement, la réalité se situe entre l'EQ-5D et le CFQ-R-8D. Si cela n'a pas été fait dans le cadre du dossier, nous ne pouvons bien sûr pas insérer cette information dans la conclusion, donc je ne vois pas trop quelle est l'amélioration possible au niveau de la conclusion.

**M. le D<sup>r</sup> ORVAIN.-** Je suis en train de regarder les publications des gens qu'ils invoquent, parce qu'il n'y a pas forcément un article direct, mais il y a des personnes, Rowen et Brazier, qui ont l'air d'être des gens assez conséquents sur les scores d'utilité. Ils ont beaucoup publié.

**M<sup>me</sup> LE GALES.-** Nous n'avons pas de problème avec Brazier, c'est sûr.

**M. le D<sup>r</sup> ORVAIN.-** Après, je n'arrive pas à retrouver l'article qui expliquerait spécifiquement ce qu'ils ont fait.

**M. SERRIER.-** Une chose qui doit être claire est que la méthode qu'ils ont appliquée pour passer de leur questionnaire spécifique à des scores d'utilité est tout à fait bonne. Ce n'est pas cela qui est remis en cause. C'est toute la difficulté. Justement, s'ils avaient mis en place cette méthode et qu'ils avaient utilisé ce questionnaire pour l'utilité globale dans l'étude, pourquoi pas ? Le problème est qu'ils ont ensuite utilisé ce bon travail pour ne calculer qu'un incrément d'utilité et l'ajouter à une autre mesure de l'utilité.

Comme le dit Lionel, s'ils avaient fait ce travail jusqu'au bout, c'est-à-dire en l'utilisant globalement, nous aurions pu utiliser l'information d'une manière ou d'une autre et cela aurait été intéressant.

**M<sup>me</sup> PARIS.-** J'aurais besoin de revoir la conclusion. J'étais en train de la chercher dans mes fichiers. Pourquoi est-ce qu'on n'utilise pas le scénario où il n'y a pas la prise en compte de cet incrément ? Cela change franchement la valeur du RDCR, mais est-ce que c'est ce que nous avons fait ? J'étais en train de chercher la conclusion.

**Un chef de projet du SEESP.-** Oui. En fait, nous avons rejeté leur analyse de référence et nous présentons dans la conclusion les résultats de l'analyse en scénario qui n'utilise pas cet incrément. Nous avons mis une réserve majeure sur la méthode de valorisation des utilités en analyse de référence, mais nous présentons des résultats qui n'utilisent pas cette méthode et qui sont donc estimés acceptables.

**M. le D<sup>r</sup> ORVAIN.-** On trouve quand même des articles qui expliquent comment ils font ces méthodes de régression.

**M<sup>me</sup> LE GALES.-** Je pense que le problème n'est pas du tout la façon dont ils ont construit les scores d'utilité associés. On se doute assez bien de la façon dont ils l'ont fait. Ce qu'a mis notre chef de projet dans le fil de discussion est suffisant pour comprendre dans les grandes lignes la méthode qu'ils ont utilisée pour construire leur score d'utilité. Ils sont partis d'un nombre réduit d'états de santé décrits avec leur questionnaire spécifique, sur lequel ils ont fait faire des TTO à des patients, et ensuite ils ont utilisé des méthodes de régression pour générer l'intégralité des scores d'utilité dont ils avaient besoin. Nous n'avons pas de problème par rapport à cela. Ce n'est

pas le sujet et cela a été fait avec des gens dont, pour l'instant, nous pouvons penser que la réputation donne grande confiance sur la qualité méthodologique de cette génération des scores d'utilité sur l'échelle spécifique. Ce n'est pas le problème.

Le problème que nous avons est de savoir pourquoi l'on ajoute les deux scores ensemble et si le travail méthodologique nécessaire a été fait et a été restitué dans le dossier de manière suffisamment complète pour que nous ayons un très bon niveau de certitude nous permettant de penser que l'on peut ajouter des scores d'utilité sans aucune difficulté, c'est-à-dire que l'on peut faire  $A + B$  sans problème. De ce que je comprends, nous n'avons pas cela. C'est cela, le problème. Ce n'est pas l'usage des régressions logistiques pour générer les scores d'utilité du CFQ-R-8D. Je crois que nous avons éclairé le propos, ce n'est pas le problème.

**M. le D<sup>r</sup> ORVAIN.-** Je suis complètement d'accord avec toi. Ce que je dis juste, c'est qu'ils ont beaucoup publié sur cette question.

**M<sup>me</sup> LE GALES.-** Je crois que là-dessus, on est bon.

**M. le D<sup>r</sup> ORVAIN.-** Ils ont y compris publié sur le fait de savoir comment faire du mapping d'une échelle vers une autre, donc visiblement ils ont beaucoup travaillé sur cette question d'incrément d'utilité. Maintenant, je suis d'accord avec toi, je ne sais pas comment ils ont fait.

**M<sup>me</sup> PARIS.-** Sur ces aspects, le dossier serait donc incomplet pour pouvoir accepter, même si c'était correct, d'après vous ? Nous n'avons pas assez d'information, même en admettant que ce soit correct. D'ailleurs, vous disiez « il y a un risque de ».

**Un chef de projet du SEESP.-** C'est ce que disent très bien Catherine et Hassan. Nous n'avons aucune notion de la validité de cette addition de scores issus de questionnaires différents, bien que chaque questionnaire — que ce soit l'EQ-5D ou le questionnaire spécifique — soit individuellement tout à fait valable.

**M<sup>me</sup> PARIS.-** Il va falloir que nous concluons. J'imagine que vous avez des engagements vous aussi, donc il va falloir que nous concluons.

J'ai l'impression qu'il n'y a quand même pas un consensus pour lever cette réserve majeure. Il y a peut-être des voix dissonantes, mais peut-être que nous pouvons passer au vote. Je propose donc a priori que nous maintenions cette réserve majeure, essentiellement parce que nous ne connaissons pas toute la méthode utilisée. Nous manquons d'informations dans le dossier pour comprendre ce qu'il se passe dans la méthode utilisée, donc je propose que nous votions en gardant cette réserve. Y a-t-il d'autres choses que nous devons voir ?

**Un chef de projet du SEESP.-** Après, le reste était dans les documents qui ont été envoyés à la commission. C'était principalement la levée de la réserve majeure.

**M<sup>me</sup> PARIS.**- D'accord. Si personne ne s'y oppose, je propose que nous votions. Carine, je te laisse procéder.

*(Il est procédé au vote par appel nominatif.)*

Résultat du vote :

Favorable : 12 voix.

Abstention : 1 voix.

*(L'avis est adopté à l'unanimité.)*