

COMMENT S'ELABORE UN CONSENSUS?

OESTROGENES ET RISQUE FRACTURAIRE, LE CAS DE LA CONFERENCE FRANÇAISE SUR LA MEDICALISATION DE LA MENOPAUSE

D. BROCLAIN (1)

Les conférences de consensus, qui visent à la modification de pratiques médicales inappropriées, sont-elles pour autant adaptées au règlement de questions controversées? A partir de l'exemple de la conférence française sur la médicalisation de la ménopause et de la question de la prévention du risque tardif de fractures ostéoporotiques par un traitement hormonal précoce, la formation de conclusions et recommandations consensuelles au sein d'un jury non spécialisé est décrite. Ces conclusions seront vivement contestées par les spécialistes invités : les données scientifiques qui “font preuve” pour les experts ne “font pas preuve” pour le jury. Les raisons de ce désaccord sont analysées. Quelques recommandations sur l'organisation des conférences de consensus portant sur des points controversés sont alors proposées.

Are consensus development conferences, which aim at changing inappropriate medical practices, well designed for managing controversial issues? Based on data collected at the “Menopause's medicalisation” french consensus development conference, the process of elaboration, by a non specialised panel, of conclusions and recommendations on prevention of osteoporotic fractures by early hormonal replacement therapy is described. These conclusions will be contested by the conference's invited specialists : scientific data weighed as “strong evidence” by these experts are not weighed so by the panel. Reasons for this controversy are analysed. Some recommendations about the organisation of consensus development conferences on controversial topics are then proposed.

¹ Médecin et sociologue, rattaché à l'équipe de l'unité 158 de l'INSERM (hôpital Necker, Paris) - 45, rue Emile Desvaux 75019 Paris.

INTRODUCTION

Une part de plus en plus importante du champ des problèmes cliniques rencontrés par les médecins dans leur exercice quotidien peut, aujourd'hui, faire l'objet de recommandations fondées sur des essais thérapeutiques et des études épidémiologiques valides, répondant à des critères méthodologiques exigeants.

La pratique médicale est invitée à se fonder fermement sur une synthèse raisonnée de l'ensemble des données produites par le monde de la recherche et à abandonner sa logique de fonctionnement traditionnelle reposant sur la valeur accordée à l'expérience clinique individuelle, à l'autorité des "patrons" et aux seules déductions logiques effectuées à partir de mécanismes physiopathologiques hypothétiques des maladies (13).

Dans ce nouveau paradigme de la pratique médicale, l'expérience clinique se voit attribuer un nouveau statut, tout en conservant sa valeur indéniable pour le praticien. Au regard des données issues d'une recherche objectivant de plus en plus l'efficacité thérapeutique en termes de vies sauvées, de rapports entre risques et bénéfices, de critères authentiquement cliniques ou d'amélioration de la qualité de vie, l'expérience clinique, qui repose sur des observations non systématiques et un jugement personnel, en est venue à prendre une figure anecdotique, voire fallacieuse dans la mesure où elle s'est avérée conforter trop souvent les praticiens dans des convictions erronées.

Il n'en reste pas moins vrai que le savoir médical est, à l'évidence, loin de se réduire à un ensemble de faits solidement établis. Le plus souvent, les données issues de la littérature apparaissent insuffisantes, quand elles ne sont pas contradictoires, pour guider la décision des praticiens face aux cas individuels. De plus, l'existence d'études publiées fortement probantes n'a globalement qu'une faible capacité à modifier le comportement des professionnels de santé. Les zones obscures, incertaines ou controversées du savoir médical représentent un facteur de grande variabilité dans les pratiques médicales individuelles.

C'est dans ce nouveau contexte cognitif et sous l'aiguillon d'une inflation des dépenses de soins dans la plupart des pays industrialisés que se sont développées de nouvelles modalités de synthèse et de transfert des connaissances médicales, conçues pour minimiser le poids accordé traditionnellement aux expériences individuelles et à l'opinion des experts. Les conférences de consensus en sont probablement l'exemple le plus connu.

Les conférences dites "de consensus" sont apparues en France à partir de 1987 et s'y sont multipliées depuis. Nées aux Etats-Unis dans les années soixante-dix, les "consensus development conferences" représentent un type de congrès médical tout à fait particulier dont la finalité essentielle est la modification de pratiques inappropriées par la diffusion de recommandations destinées aux professionnels de santé (19, 23).

Le modèle originel en a été développé aux USA par l'Office of Medical Application of Research (OMAR) des National Institutes of

Health (26, 27) et a été repris en France par l'Agence Nationale pour le Développement de l'Evaluation Médicale (2,3).

Selon Itzhak Jacoby, l'un des premiers responsables du programme américain de conférences de consensus, “la procédure de recherche de consensus a été élaborée en vue de la résolution de conflits” (9). Leur schéma organisationnel a été emprunté à trois modèles : le procès judiciaire, où les preuves sont présentées devant un jury impartial qui prononce le verdict final après s'être réuni à huis-clos; la réunion d'experts scientifiques qui présentent et discutent entre eux les résultats de leurs travaux; le forum public où tous les acteurs sociaux concernés par le sujet peuvent venir exprimer leur point de vue (8).

PROBLEMATIQUE ET METHODOLOGIE

Toutes les écoles de pensée “rivaless” et tous les courants de pratique sont, en principe, réunis et confrontés dans une conférence de consensus. De plus, le dernier mot sur la synthèse des données scientifiques disponibles ne revient pas, ici, à un groupe d'experts

auto-constitué, comme c'est le cas dans les consensus dits "d'experts", mais à un jury non partisan ².

Le jury rédige ses conclusions à huis-clos, après audition publique des spécialistes invités. Cette audition se déroule en un à deux jours et le jury est encouragé à intervenir dans les débats en posant des questions aux experts

Les données présentées lors d'une conférence de consensus sont complexes et le travail incombant au jury d'analyser et synthétiser les informations disponibles sur une courte période de temps³, se présente comme un exercice difficile, non dépourvu de subjectivité (18). Dans l'élaboration de ses conclusions, le jury reste cependant sensible aux

² Plusieurs modèles de critères ont été utilisés pour choisir les membres d'un jury de conférence de consensus (8). Le modèle du "jury équilibré" consiste à rassembler des experts du thème représentant les différents points de vue existants. Ce modèle a été peu à peu abandonné aux Etats-Unis, car les débats y dégénéraient, en discussions non constructives. Le modèle du "jury neutre", adopté dans les conférences américaines initiées par le NIH, consiste à rassembler un panel pluridisciplinaire d'experts du thème considéré dont aucun n'a publié, avec une opinion tranchée, sur le sujet. Le modèle "anglais" en diffère radicalement puisqu'au Royaume-Uni, les jurys ont été délibérément constitués en majorité, de non experts du thème mais d'expertise reconnue dans d'autres domaines (31). En France, l'Agence Nationale pour le Développement de l'Evaluation Médicale (ANDEM) recommande un modèle mixte : jury pluridisciplinaire, comportant quelques personnes "proches du thème traité" par leurs connaissances ou leur pratique professionnelle et quelques non experts du thème.

³ Dans la plupart des pays organisant des conférences de consensus, les comités d'organisation ne laissent au jury qu'une seule nuit pour rédiger la première version de leurs conclusions. Au matin du troisième jour, ces conclusions provisoires sont mise à l'épreuve d'un débat public animé par les experts, les membres du comité d'organisation et, théoriquement, toute personne présente sur les lieux du débat. Les conclusions définitives, qui seront diffusées aux praticiens, sont également rédigées à huis clos .

arguments scientifiques présentés par les experts (22)⁴. Une question débattue dans le milieu des organisateurs de conférences de consensus (23) est de savoir si le rôle du jury est de conclure sur l'ensemble des données disponibles, même lorsque celles-ci sont imparfaites ou incomplètes, ou s'il doit le faire en fondant son jugement seulement sur l'ensemble restreint des études les plus solides.

Pour accomplir leur tâche, les membres d'un jury de conférence de consensus disposent d'un certain nombre de ressources: l'expérience propre à chacun, les documents écrits (interventions des experts, sélection d'articles scientifiques) mis à leur disposition par le comité d'organisation un mois avant la conférence, les présentations orales des experts et les débats provoqués par les questions posées par le public de la conférence et le jury lui-même.

Quels éléments amènent un jury à accepter ou rejeter comme “preuves suffisantes” les données présentées par les experts?

Comment le jury réagit-il face à des données incertaines?
Comment et dans quelle mesure un consensus est-il atteint en ce cas?

⁴ La qualité d'“expert” est attestée par des publications récentes sur le thème d'intervention. Les présentations des experts, qui sont publiques, doivent se fonder sur une analyse de résultats scientifiques déjà publiés. Si l'expert convié à une conférence de consensus peut avoir une opinion tranchée, il lui est cependant demandé de présenter au jury une information exhaustive et référencée et d'explicitier le degré de certitude attaché à chacune de ses conclusions. De plus, les liens éventuels entre certains experts et des institutions publiques ou privées pour lesquelles les conclusions de la conférence représentent un enjeu symbolique ou matériel doivent, toujours en principe, être connus de tous au nom d'un principe de transparence.

En avril 1991, se tenait à Paris une conférence intitulée "Médicaliser la ménopause?".

Un matériel couvrant l'ensemble des phases de cette conférence de consensus a pu être rassemblé par nos soins

Les interventions des experts, les questions du jury et du public ont été enregistrées et transcrites.

Le jury nous a donné son accord pour être observé dans son fonctionnement et pour que puissent être conservées les versions successives de ses conclusions rédigées à huis-clos : versions initiales, version dite "provisoire" (soumise au débat public du troisième jour), version dite "définitive" (présentée en conférence de presse).

Une fois la conférence de consensus achevée, des entretiens de recherche auprès de membres du comité d'organisation, d'experts ainsi que de tous les membres du jury ont été réalisés .

Pour être en mesure d'analyser la formation, sur fond de controverse, de conclusions présentées comme consensuelles, nous avons choisi de nous concentrer spécifiquement sur les prises de position du jury concernant une question controversée, celle de la prévention du risque d'ostéoporose par la prescription précoce d'un traitement hormonal substitutif et d'un examen complémentaire récent, la densitométrie osseuse.

Après une description succincte du contexte (modalités de mise en place de cette conférence de consensus), nous présenterons les termes du

désaccord sur la prévention du risque fracturaire, en en faisant l'analyse à la lumière de la littérature et des entretiens menés auprès du jury, des experts et des membres du comité d'organisation.

LE CONTEXTE: MISE EN PLACE DE LA CONFERENCE SUR LA MENOPAUSE

Au début de l'année 1989, la Mutualité Française avait confié à la Fondation de l'Avenir pour la Recherche Médicale Appliquée, institution qu'elle avait elle-même mise en place en 1987, la mission de développer une politique d'évaluation en matière de santé. La Fondation s'engagera, entre autres actions, sur un programme d'organisation de conférences de consensus.

Le thème de la médicalisation de la ménopause est proposé à la Mutualité Française en 1990. Les raisons de ce choix ont été explicitées dans un argumentaire attaché au projet de conférence. La ménopause survenant en général autour de 50 ans, environ 300000 femmes atteignent cet âge chaque année en France, avec une espérance de vie de plus de trente ans. Le nombre de françaises ménopausées avoisine neuf millions d'individus. L'argumentaire met en avant que seules “5% des femmes utilisent un traitement hormonal substitutif en France” alors que “les traitements hormonaux paraissent diminuer dans des proportions variables les risques considérés comme secondaires à la déplétion hormonale: risques de maladie coronarienne et d'accident vasculaire cérébral, de fracture et de tassement vertébral ostéoporotique, de dépression, de sécheresse vaginale et d'incontinence urinaire, de rides etc.”.

Ces éléments donnent la mesure des multiples enjeux associés à la médicalisation de la ménopause (4). Enjeux économiques pour les industriels du traitement hormonal substitutif, les fabricants de densitomètres et les financeurs du système de soins, enjeux médico-professionnels, notamment pour les radiologues, les gynécologues et les rhumatologues libéraux, enjeux socio-culturels et de santé publique, enfin, pour l'ensemble des médecins, pour les femmes ménopausées et, plus généralement, pour tous les acteurs concernés.

La Fondation de l'Avenir reçoit mandat d'organiser cette conférence et un comité d'organisation est mis en place. ⁵

Pour assurer la présidence du jury, le comité cherchera une femme, médecin et n'ayant, de surcroît, aucune expertise particulière en matière de ménopause. Une spécialiste en anesthésie-réanimation acceptera de jouer ce rôle, notamment par curiosité pour la procédure. Les experts, quant à eux, seront sélectionnés avant tout en fonction de leur notoriété scientifique, mais parfois aussi en fonction de leur notoriété médiatique⁶. Les douze membres du jury seront choisis par le comité d'organisation sur quelques critères explicites : ne pas être spécialiste de la ménopause,

⁵Aux côtés du consultant de la Fondation de l'Avenir, spécialiste de santé publique, qui siège dans le comité d'organisation en qualité de "coordinateur", s'adjoignent un spécialiste de médecine interne, un gynécologue, un endocrinologue, ainsi que deux femmes, l'une généraliste, l'autre non médecin mais connaissant bien les principes d'organisation des conférences de consensus.

⁶ "On ne peut faire sans ceux qui font l'opinion, quel que soit leur niveau de qualité scientifique", expliquera un membre du comité d'organisation tandis qu'un autre soulignera que "si l'on n'invite pas les personnalités médiatiques à s'exprimer dans le cadre de la conférence, il y a un risque qu'elles puissent se retourner contre ses conclusions".

être capable de s'exprimer et communiquer en groupe, être capable de suivre la teneur des débats scientifiques.

Le jury finalement sélectionné comporte une forte proportion de femmes (9 femmes pour 3 hommes) et une majorité de médecins (8 médecins contre quatre non-médecins) ⁷.

Cinq grandes questions, servant de cadre de travail pour la conférence de consensus et appelant des réponses précises de la part des experts et du jury, seront élaborées par le comité d'organisation :

1- Quels sont les inconvénients et les risques à court et à long terme attribuables à la ménopause, traitée ou non traitée, sur le plan de la santé?

2- Faut-il réserver les traitements de la ménopause à certaines personnes et sur quels éléments? Les pratiques observées en France, comparées à celles des autres pays industrialisés, témoignent-elles d'une position déterminée de la part des professionnels et du public?

3- Les traitements hormonaux présentent-ils des risques à court et à long terme qui devraient les exclure pour certains groupes de femmes?

4- Quels éléments, y compris économiques et sociaux, doivent intervenir pour définir une stratégie préventive ou thérapeutique contre les inconvénients et risques évitables de la ménopause?

5- Quel est l'avenir des thérapeutiques de la ménopause? Quelles mesures publiques ou privées seraient utiles, notamment en

⁷ Le groupe "médecins" du jury comprend cinq spécialistes (une gynécologue libérale, une gérontologue, une épidémiologiste, un interniste et la présidente, spécialiste en anesthésie-réanimation), deux médecins généralistes (un homme et une femme) et un journaliste médical de la presse grand public. Les quatre non médecins ont une activité professionnelle en rapport direct et quotidien avec le monde de la médecine puisqu'il s'agit de deux femmes spécialisées de longue date dans le journalisme médical (presse professionnelle pour l'une, grand public pour l'autre), d'une économiste de la santé et d'une conseillère conjugale travaillant dans un centre hospitalier.

matière de recherche, de réglementation et de communication sociale?

Un mois avant la conférence de consensus, les membres du jury se retrouvent ensemble pour la première fois. Une sélection d'articles scientifiques, une bibliographie et la plupart des interventions d'experts leur sont remises alors par le comité d'organisation. La question de savoir s'il convient de se préparer en étudiant attentivement la documentation ou s'il vaut mieux arriver "en naïf" à la conférence "pour ne pas avoir d'idées préconçues" fera, ce jour là, l'objet d'un débat qui ne pourra être tranché.

DESACCORDS ENTRE EXPERTS ET JURY

Pour la compréhension des arguments échangés, nous rappellerons que l'ostéopénie, diminution physiologique de la densité osseuse, est un phénomène lié au vieillissement, phénomène qui s'accélère au moment de la ménopause. Dans la maladie ostéoporotique, la perte osseuse a atteint un seuil pathologique où surviennent des fractures, soit spontanément, soit pour des traumatismes minimes. Les tassements vertébraux, les fractures du col du fémur et celles du poignet sont les plus fréquentes. L'ostéoporose est donc, par définition, fracturaire. Elle est précédée d'une longue période d'ostéopénie où la réduction de la densité osseuse, même lorsqu'elle est importante, reste asymptomatique. La plupart des fractures ostéoporotiques du col du fémur sont observées après l'âge de soixante-dix ans.

Désaccord sur le pouvoir prédictif de la densitométrie osseuse

Au cours de la session initiale de la conférence réservée aux exposés des experts, le Pr A, spécialiste en endocrinologie osseuse et président du Groupe de Recherche et d'Information sur l'Ostéoporose (GRIO), présentera comme un fait établi la possibilité de prédire le risque individuel de fracture ostéoporotique à long terme à partir d'une seule mesure de densité osseuse réalisée au moment de la ménopause. L'argument en est la convergence de cinq études épidémiologiques prospectives récentes. Une fois les densitomètres standardisés entre eux et une fois adoptée une définition consensuelle de la densité osseuse normale⁸, la mise en oeuvre du dépistage des femmes à risque d'ostéoporose sera, selon cet expert, tout à fait concevable.

Si l'on considère que le traitement hormonal doit être réservé aux femmes à risque d'ostéoporose, poursuivra alors le Pr A, il est nécessaire de faire le dépistage systématique par densitométrie de toutes les femmes ménopausées, 300 000 femmes par an en France, et donner les moyens de ce dépistage et de son remboursement. Ce dépistage peut alternativement être ciblé de manière plus restrictive sur les femmes qui ne peuvent ou ne veulent pas prendre d'hormones, afin de les inciter à se traiter. *“Si on juge le traitement hormonal efficace pour prévenir*

⁸*“Trois options sont possibles pour définir ce seuil de séparation entre normal et pathologique: prendre pour seuil la densité osseuse moyenne de l'adulte jeune, prendre pour seuils les densités osseuses moyennes pour différentes tranches d'âge ou prendre la valeur critique de la densité osseuse au dessous de laquelle le squelette se fracture (seuil fracturaire). Si la première option était choisie, la moitié de la population féminine âgée de plus de soixante ans serait à traiter, ce qui est excessif. Reste à trouver un accord sur l'une ou l'autre des autres options. Cela a été fait pour l'hypertension artérielle et pour le cholestérol, il n'est pas question de ne pas le faire pour la densité osseuse” (Pr A)*

l'ostéoporose, dira le Pr A, rien ne sert de mesurer la masse osseuse".
Poussé par le jury à préciser ses arguments, le Pr A ajoutera que *"dans la majorité des cas,, cette mesure ne doit pas être le moteur de la décision de prescrire un traitement hormonal substitutif"*.

Un chercheur présent dans la salle, épidémiologiste spécialisé dans l'ostéoporose, prendra la parole pour contester les arguments du Pr A :
"Je ne pense pas qu'il existe des études prospectives démontrant qu'il est possible de prédire avec fiabilité un risque de fracture du col du fémur à 70 ans, en faisant une ostéodensitométrie vers 45-50 ans. Dans les études prospectives existantes, la densitométrie est pratiquée chez des femmes bien plus âgées en moyenne. Cette prédiction est peut-être possible à la ménopause, j'en conviens, mais aucune étude prospective sérieuse ne l'a montré et je ne pense pas qu'on puisse l'affirmer".

Le Pr A s'élèvera *"contre cette affirmation péremptoire"*, accusant son contradicteur de méconnaître la bibliographie et répliquera que l'intérêt de la mesure de la masse osseuse chez des femmes de la quarantaine a bien été établi par cinq études prospectives pour prédire un risque de fractures survenant 9 ou 10 ans après.

"Il ne s'agit donc pas de fractures du col du fémur", remarquera alors l'épidémiologiste, qui ajoutera que si certaines études prospectives incluent effectivement au départ des femmes de la trentaine et de la quarantaine, ces études ne précisent pas à quel âge a été pratiquée la densitométrie objectivant une masse osseuse basse.

Le Pr B, spécialiste en méthodologie de la recherche clinique, sera l'un des derniers experts entendus par le jury avant sa première session de travail à huis-clos. Son intervention viendra renforcer les interrogations du jury sur la validité des études présentées sur l'ostéoporose. Après avoir exposé sa perspective sur la médicalisation de la ménopause (*“Prescription de confort, le traitement hormonal substitutif ne doit pas entraîner de conséquences néfastes sur la santé et, si possible, avoir des conséquences bénéfiques”*), le Pr B passera en revue les travaux traitant de l'efficacité préventive du traitement hormonal sur les maladies cardio-vasculaires. Montrant en quoi ces études avaient été réalisées avec une méthodologie inadaptée, il en arrivera à *“l'impossibilité de conclure à un effet bénéfique ou néfaste sur ces pathologies”*. Revenant alors à la question du traitement de la ménopause en général, le Pr B conclura à la nécessité, pour lever l'ensemble des incertitudes existantes, de mettre en oeuvre *“des essais contrôlés contre placebo, de taille et de durée suffisantes”*. Le Pr B achevera son intervention par une sérieuse mise en garde sur les risques méconnus éventuellement liés à la prise d'oestrogènes au long cours par les femmes ménopausées : *“Il conviendra d'entreprendre ces essais contrôlés avant que le traitement hormonal substitutif de la ménopause ne devienne un traitement standard et ne soit à l'origine d'une nouvelle affaire du distilbène...”*

Dans ses conclusions rédigées à huis-clos après l'audition publique des experts (cf. versions 1, 2, 3), le jury considèrera non encore démontrée la capacité de la densitométrie osseuse à dépister les femmes à risque de fractures ostéoporotiques. Il recommandera donc aux praticiens

de ne pas chercher à mesurer la masse osseuse, au moment de la ménopause, dans le but d'en tirer un argument de décision thérapeutique.

Il convient de signaler ici l'existence, dans la littérature (29), d'une controverse sur le bien-fondé de mesurer la valeur prédictive, la sensibilité et la spécificité de la densitométrie osseuse pour apprécier l'intérêt du dépistage du risque fracturaire⁹. Cette controverse n'a été exposée par aucun des experts présents aux membres du jury de la conférence, alors que les enjeux de ce débat méthodologique sont loin d'être négligeables. En témoigne la modélisation, reposant sur des hypothèses "optimistes", reprise par Freemantle et al. (12) : avec un dépistage généralisé par densitométrie osseuse couvrant 70% de la population cible, une sensibilité de l'examen de 37%, une durée de traitement hormonal de dix ans pris par 30% des femmes et en admettant que ce traitement diminue l'incidence des fractures de 50% en conservant cet effet protecteur à long terme, seulement 4% des fractures du col du fémur survenant une année donnée

seraient prévenues par une telle politique de dépistage.

⁹ Outre l'article de Ross, paru deux ans avant la conférence de consensus, on trouvera également les termes de cette controverse dans l'article de Law et al (21) et la correspondance suscitée par cette dernière publication (28).

Version N°1

Enoncés élaborés en huis-clos par un sous-groupe du jury

Le traitement substitutif de la ménopause semble particulièrement indiqué chez les femmes à risque de perte osseuse rapide et d'ostéoporose.

Cependant leur détection reste encore problématique. Un antécédent familial d'ostéoporose fracturaire est un bon indicateur.

Il n'existe pas actuellement d'examen ayant une valeur prédictive satisfaisante d'une ostéoporose à risque fracturaire. En particulier, la densitométrie osseuse, outre son coût important, ne présente pas actuellement de fiabilité quant à la standardisation de ses résultats et sa valeur prédictive.(...)

Dans l'état actuel des connaissances, le traitement hormonal substitutif de la ménopause ne peut être prescrit systématiquement aux femmes. En effet, l'efficacité de son rôle préventif à long terme (ostéoporose, accidents cardio-vasculaires) de même que son absence de risque à long terme (cardio-vasculaire, cancérogène) ne sont pas suffisamment démontrés. Dans tous ces domaines, on manque cruellement d'essais cliniques contrôlés.

Version N°2

Enoncés acceptés par le jury, avant discussion publique.

Le traitement substitutif de la ménopause est particulièrement indiqué chez les femmes à risque de perte osseuse rapide et d'ostéoporose fracturaire.

Cependant leur détection est encore discutée. Un antécédent familial d'ostéoporose fracturaire est un bon indicateur.

Il n'existe pas actuellement d'indicateurs cliniques ou paracliniques ayant une valeur prédictive satisfaisante d'une ostéoporose à risque fracturaire. La densitométrie osseuse est en cours de standardisation et des études sont menées pour définir l'intérêt de cet examen dans la prédiction du risque fracturaire. Pour l'instant, cet examen ne peut être recommandé comme argument de décision thérapeutique (...)

L'efficacité du rôle préventif du traitement hormonal à long terme (risques fracturaires de l'ostéoporose, accidents cardio-vasculaires) de même que son absence de risque à long terme (cardio-vasculaire, cancer du sein) ne sont pas encore démontrés. Dans tous ces domaines, on manque cruellement d'essais cliniques contrôlés. (...)

Dans l'état actuel des connaissances, le traitement hormonal substitutif de la ménopause ne peut être préconisé systématiquement à l'ensemble de la population concernée.

Version N°3

Enoncés extraits des conclusions finales du jury, rédigés à huis-clos après le débat public.

Le traitement substitutif de la ménopause est particulièrement indiqué chez les femmes à risque de perte osseuse rapide et d'ostéoporose fracturaire.

Cependant leur détection est encore discutée. Il n'existe pas actuellement d'indicateurs cliniques ou paracliniques ayant une valeur prédictive satisfaisante d'une ostéoporose à risque fracturaire. La densitométrie osseuse est en cours de standardisation et des études sont menées pour définir l'intérêt de cet examen dans la prédiction du risque fracturaire. Pour l'instant, cet examen ne peut être recommandé comme argument de décision thérapeutique (...)

Les oestrogènes administrés à dose suffisante et prolongée évitent la réduction de la masse osseuse liée à la ménopause. Il existe donc une forte présomption clinique de l'efficacité du traitement hormonal à long terme dans la prévention du risque fracturaire de l'ostéoporose. Cette présomption mérite d'être confirmée par des études cliniques rigoureuses. (...)

Dans l'état actuel des connaissances, le traitement hormonal substitutif de la ménopause ne peut être préconisé systématiquement à l'ensemble de la population concernée. (les conclusions du texte complet du jury finissent sur cette phrase

Version N°4

Enoncés diffusés après la conférence par la Fondation de l'Avenir

Le traitement substitutif de la ménopause est particulièrement indiqué chez les femmes à risque de perte osseuse rapide et d'ostéoporose fracturaire.

Cependant leur identification est encore discutée. Il n'existe pas actuellement d'indicateurs cliniques ou paracliniques ayant une valeur prédictive satisfaisante d'une ostéoporose à risque fracturaire. La densitométrie osseuse est en cours de standardisation et des études sont menées pour définir l'intérêt de cet examen dans la prédiction du risque fracturaire. Pour l'instant, cet examen ne peut être recommandé comme argument unique de décision thérapeutique (...)

Les oestrogènes administrés à dose suffisante et prolongée évitent la réduction de la masse osseuse liée à la ménopause. Il existe donc une forte présomption clinique de l'efficacité du traitement hormonal à long terme dans la prévention du risque fracturaire de l'ostéoporose. Cette présomption mérite d'être confirmée par des études cliniques rigoureuses. (...)

Dans l'état actuel des connaissances, le traitement hormonal substitutif de la ménopause peut être proposé et recommandé mais certainement pas imposé à l'ensemble de la population concernée. . (les conclusions du texte complet du jury finissent sur cette phrase)

Désaccord sur l'efficacité anti-fracturaire des oestrogènes

Trois des experts invités, le Pr A, le Pr C, rhumatologue hospitalier du secteur public, et le Pr D, gynécologue hospitalier et consultant pour l'industrie pharmaceutique, présenteront l'efficacité du traitement hormonal substitutif sur le risque fracturaire comme établie par de nombreuses études épidémiologiques concordantes : selon ces spécialistes, les oestrogènes naturels d'origine équine permettent d'obtenir *“une réduction d'au moins 50% des fractures du poignet et du fémur et de près de 90% des fractures vertébrales”*, à condition d'être administrés pendant au moins cinq ans à partir de la ménopause.

Le schéma thérapeutique couramment utilisé en France, associant l'oestradiol (oestrogène naturel d'origine humaine) à un progestatif, ayant, selon le Pr A, *“une efficacité démontrée sur la perte osseuse postménopausique”*, ce dernier considère légitime d'en déduire que leur efficacité anti-fracturaire est *“très probable”*.

Aux yeux de ces experts, la démonstration de l'efficacité préventive du traitement hormonal sur la survenue de fractures repose ainsi sur l'extrapolation aux oestrogènes humains prescrits couramment en France d'effets observés avec les oestrogènes équins utilisés à l'étranger dans les études épidémiologiques citées. De plus, les oestrogènes humains ayant une efficacité sur la perte osseuse post-ménopausique, comme le montrent un ensemble d'études cliniques menées dans le court terme, les experts en déduisent qu'ils préviennent très certainement, à long terme cette fois, la survenue des fractures ostéoporotiques.

Réunis à huis-clos pour rédiger leurs conclusions provisoires, les membres du jury accorderont une importance toute particulière aux biais inhérents à l'ensemble des études présentées par les experts, pour la plupart non contrôlées, rétrospectives et réalisées avec des molécules différentes de celles utilisées couramment en France. *“Allons-nous encourager le traitement systématique alors que l'efficacité et l'absence de risque ne sont pas démontrées?”*, se demanderont-ils. Constatant l'inexistence d'études réellement valides en ce domaine, le jury conclura dans la quasi-unanimité ¹⁰, à l'*“absence d'efficacité anti-fracturaire et d'innocuité a long terme démontrées”* du traitement oestrogénique (versions 1 et 2) en conscience du mécontentement que cette prise de position risquait de provoquer (*“Ça ne va pas faire plaisir aux spécialistes!”*).

Les conclusions provisoires, rédigées jusque tard dans la nuit, seront lues, puis projetées page par page le lendemain matin, la présidente du jury invitant les experts présents à y corriger d'éventuelles “erreurs grossières” et à intervenir pour en améliorer la forme.

Lors de la rétro-projection, le Pr D, gynécologue hospitalier et consultant pour l'industrie pharmaceutique, sera le premier à réagir au passage énonçant que l'efficacité antifracturaire du traitement oestrogénique n'était pas encore prouvée (*“L'efficacité préventive sur le risque fracturaire est parfaitement démontrée! (...) Je pense que*

¹⁰ Le “premier jet” (version 1) des conclusions provisoires, rédigé par un sous-groupe des membres du jury, sera soumis à l'ensemble du groupe qui n'en modifiera pas la teneur (version N°2). Un seul membre du jury exprimera son désaccord : *“Je vous trouve extrêmement sévères, mais la faute en revient aux experts. Ils n'ont pas su vous convaincre.”* (gynécologue, jury)

l'ensemble des experts qui ont parlé ici se désolidarisent de telles conclusions!”). Le jury maintiendra publiquement sa position selon laquelle, d'un strict point de vue scientifique, la preuve n'avait pas été apportée. Le Pr D insistera : “*On ne peut pas écrire une phrase comme celle-là, c'est impossible!*”, bientôt appuyé par le Pr C (rhumatologue), visiblement en colère, annonçant son intention de réunir rapidement tous les experts concernés de la conférence pour qu'ils “*apportent leurs commentaires*” à ces conclusions jugées “*surprenantes*”. Le Pr C demandera au jury de trouver une formulation “*plus nuancée*” à propos de l'efficacité préventive du traitement hormonal.

Le texte définitif (version 3), destiné à être présenté en conférence de presse les jours suivants, sera rédigé à huis-clos aussitôt après ce débat public mouvementé.

Au cours de la conférence, deux points distincts ont donc opposé très vivement un certain nombre d'experts au jury. L'un des désaccords porte sur le pouvoir prédictif de la densitométrie osseuse, l'autre sur l'efficacité anti-fracturaire des oestrogènes. Les enjeux médicaux, professionnels, industriels et sociaux en sont particulièrement importants puisque la réponse apportée à chacune de ces questions conditionne à la fois l'extension de l'usage de l'ostéodensitométrie à des fins de dépistage et la prescription d'un traitement prolongé (cinq ans minimum, à vie selon certains spécialistes) à un grand nombre de femmes ménopausées.

Dans un cas comme dans l'autre, le désaccord gravite essentiellement autour de la notion de “*preuve suffisante*”.

DISCUSSION

La notion médicale de “preuve scientifique”.

Le savoir médical se présente comme un territoire composite où des îlots de terre ferme, parsemés de trous, cotoient de vastes zones marécageuses. Nous avons souligné que la pratique médicale avait la possibilité, aujourd'hui, pour des questions cliniques de plus en plus diverses, de s'ancrer sur une synthèse raisonnée et validée des données scientifiques disponibles, autrement dit sur un nombre grandissant d'îlots de terre ferme. En médecine, le développement considérable de la biostatistique, qui a joué un rôle déterminant dans cette mutation, se traduit par l'afaiblissement progressif du poids de la “preuve traditionnelle” que représentent l'expérience clinique, l'opinion des patrons et l'avis informel des experts (13).

Comme le souligne Alain Desrosières (11), la statistique, en tant qu'instrument d'administration de la preuve, porte en elle un triple héritage historique : la certitude mathématique de la démonstration, la certitude physique de l'évidence des sens et la certitude morale du témoignage et de la conjecture ¹¹.

La statistique et le calcul des probabilités occupent en effet de nos jours une place essentielle dans les procédures modernes d'administration scientifique de la preuve, tant dans les sciences de la nature que dans les sciences sociales. Ces outils mathématiques ont été forgés

¹¹“Le mot *probabilité*, qui signifiait au Moyen-Age “avis certifié par l'autorité” est devenu un “degré d'assentiment proportionnel à l'évidence des choses et des témoins”. Puis Leibniz et Nicolas Bernouilli, à la fois mathématiciens et juristes, intègrent ces trois niveaux de certitude en un *continuum*, incluant tous les degrés de l'incrédulité à la conviction complète” (Desrosières).

progressivement pour tenter de maîtriser l'erreur, l'aléa et l'incertitude rencontrés, dans leurs observations et leurs descriptions des phénomènes naturels et sociaux, par les scientifiques de toutes disciplines. La statistique définit conventionnellement les objets qu'elle mesure, précise, par le langage probabiliste, le degré de confiance que l'on peut attacher à ces mesures, dispose d'outils mathématiques permettant d'analyser la nature de la relation entre objets et de mesurer la force des liens. L'affectation de cas singuliers à des classes d'équivalence et la régularité des moyennes s'oppose, dans l'univers statistique, au chaos et à l'imprévisibilité des phénomènes individuels. De plus, les outils analytiques qui y ont été développés (méthodes des moindres carrés, calculs des corrélations, tests de signification, estimations, etc.) présentent des propriétés d'optimalité assez puissantes pour entraîner théoriquement l'accord entre tous et plus particulièrement entre savants. (11).

La recherche clinique ou épidémiologique consiste, dans la plupart des cas, à s'interroger, à propos d'un ensemble bien caractérisé de patients (âge, sexe, type de maladie, etc.), sur la nature de la relation liant une exposition ou une intervention définie (facteur de risque, médicament, chirurgie, test diagnostique, mesure préventive, etc.) aux effets observés dans la population étudiée, mesurés selon des critères cliniquement pertinents (soulagement ou disparition d'un symptôme, complication, mortalité, etc.). Les effets constatés sont-ils clairement imputables à l'intervention considérée ou bien ne le sont-ils qu'avec un modeste degré de certitude? Le poids de la preuve que peuvent représenter les résultats d'une étude dépend à la fois du type et de la qualité de la méthodologie

mise en oeuvre, de la mesure de la force statistique du lien mis en évidence et de la pertinence clinique des variables que les auteurs ont choisi de mesurer.

Désaccords sur la notion de “preuve suffisante”

Pourquoi, dans le cas de cette conférence de consensus, ce qui fait preuve pour les uns ne le fait-il pas pour les autres?

Dans l'optique du paradigme scientifique médical de cette fin de XXème siècle (13) et s'agissant de démontrer l'efficacité anti-fracturaire et l'innocuité à long terme des oestrogènes, les données et la force de la preuve qui permettraient de lever radicalement les incertitudes liées aux imperfections méthodologiques des études disponibles en 1991, ne peuvent être apportées que par la mise en oeuvre d'un essai contrôlé randomisé, débutant vers 45-50 ans et se poursuivant pendant une trentaine d'années (16). Une telle étude est matériellement difficile à mettre en oeuvre.

Pour des raisons de moyens ou pour des raisons éthiques, les questions de santé publique se prêtant mal à l'expérimentation sont en fait nombreuses. En pratique, la communauté médicale en est le plus souvent réduite à devoir extrapoler les connaissances produites dans certaines conditions à d'autres conditions très différentes: extrapolation hautes doses-basses doses, d'une population à une autre, de l'animal à l'homme, d'une condition d'exposition à une autre, etc. Dans cet espace d'incertitudes scientifiques, une place croissante est accordée à l'expert. *"L'expertise, écrit L.Abénhaïm, consiste à dépasser le simple niveau*

des données pour formuler une prise de position scientifique qui peut être la substance d'une décision."(1)

Cet essai n'étant donc pas réalisé et publié au moment de la conférence de consensus, les désaccords sur l'efficacité anti-fracturaire des oestrogènes et sur la valeur de densitométrie osseuse en tant que test de dépistage peuvent être considérés comme des conflits opposant deux attitudes distinctes vis à vis des incertitudes inhérentes aux données scientifiques présentées : ce qui apparaît aux uns comme "fort degré de certitude" représente un "fort degré d'incertitude" pour les autres.

Ainsi, tout se passe comme si le jury s'était refusé à produire des recommandations pour la pratique médicale en l'absence d'études méthodologiquement incontestables.

Pourquoi le jury, en donnant tout son poids au caractère incertain des données publiées, s'est-il comporté en "chercheur" plus qu'en "clinicien", selon la distinction classique du sociologue Eliot Freidson? (15)

"Les experts n'ont pas su nous convaincre"

Les entretiens réalisés auprès des membres du jury montrent toute l'importance des interventions orales des experts et des débats sur l'élaboration du "consensus". A quelques exceptions près, en effet, les articles scientifiques mis à disposition du jury par le comité d'organisation avant la conférence n'ont pu être que survolés, essentiellement par les jurés de formation médicale. Pendant le huis-clos, ces documents ont été surtout utilisés pour vérifier un chiffre.

A une exception près¹², l'ensemble des membres du jury dira “ne pas avoir été convaincu” par les experts¹³.

Le consensus, proche de l'unanimité, atteint par le jury en matière de prévention du risque ostéoporotique tient à des éléments d'une autre nature que le seul débat scientifique. Les convictions des experts invités à intervenir sur ce thème, perçues comme telles, se sont de fait traduites par une réaction de méfiance dans le jury, renforcée par des questions laissées sans réponse, un ton jugé “trop affirmatif”, une place insuffisante laissée au doute ou au “non-savoir”, des réactions de colère ou de révolte considérées suspectes pour des scientifiques. L'évaluation en cours de l'ostéodensitométrie par l'Agence Nationale pour le Développement de l'Evaluation Médicale (ANDEM) a, de plus, représenté un facteur de doute supplémentaire pour certains membres du jury.

Le jury, en tant que groupe indépendant et extérieur aux “communautés” des rhumatologues, radiologues et gynécologues, s'est

¹²12 “(...) *J'ai été un peu fascinée par ce que disaient les experts qui ont parlé de l'ostéoporose. Je trouvais que c'était bien (...) Sans doute étais-je trop dans mes convictions de spécialiste....Sur l'ostéodensitométrie, je n'étais pas d'accord avec les autres membres du jury. Les experts n'ont pas su être assez convaincants...*” (gynécologue, jury)

¹³13 “*Avant cette conférence, je croyais sincèrement que, dans une certaine mesure, le traitement oestrogénique permettait d'éviter l'ostéoporose. J'ai découvert avec stupéfaction que peut-être l'évite-t-il, peut-être ne l'évite-t-il pas, mais que, pour le moment, on n'a aucune preuve de cette chose ni de son contraire.*” (Présidente du jury) ou encore “*La littérature n'a pas réussi à me convaincre. Le fait que certains des experts avaient une attitude agressive et, me semble-t-il, assez intéressée dans le développement de la technique ou du traitement de l'ostéoporose, a déterminé également nos conclusions. (...) Nous sommes arrivés là aussi en l'absence de réponses claires aux questions que nous avons pu poser aux experts. Ils n'ont pas réussi à nous convaincre...*” (épidémiologiste, jury).

comporté comme s'il incarnait l'intérêt collectif et devait prendre position contre les intérêts privés des industriels et de certains segments professionnels. La position adoptée par le jury sur la densitométrie peut, en effet, être interprétée comme un souci de ne pas légitimer le développement d'une nouvelle technique coûteuse dont l'intérêt, en termes de santé publique, n'est pas encore clair. Quant à la position sur l'efficacité anti-fracturaire des oestrogènes, elle traduit une attitude prudente : dans le doute, surtout ne pas faire courir de risque, essentiellement celui de cancer du sein, à la population des femmes ménopausées. Le "primum non nocere" adopté par le jury va à contre-courant de la tendance médicale, industrielle et médiatique qui consiste à dramatiser les effets pathologiques attribués à la ménopause et leur coût social, minimisant les risques et surestimant les bénéfices du traitement hormonal substitutif (4, 10).

Le fait que ce jury de douze membres était composé de neuf femmes a pu jouer ici dans le sens d'un surcroît de prudence. Mais les entretiens montrent une réelle diversité d'attitudes vis à vis du traitement hormonal, en général nuancées, d'enthousiasme pour certaines, de réticence pour d'autres. Comme le souligne l'un de ses membres (gérontologue), *"La plupart des femmes du jury ayant la quarantaine, il leur était impossible, en tant que femmes confrontées à leur propre ménopause, de se maintenir dans une position totalement objective."*

Chacune a abordé le sujet avec un surplus de motivation mais aussi avec sa problématique personnelle".¹⁴

Un "consensus" négocié et forcé

Les énoncés initiaux du jury sur la prévention de l'ostéoporose ont été significativement remaniés à la suite des vives réactions d'un certain nombre d'experts lors débat public organisé autour de ses conclusions "nocturnes".

En effet, l'"*absence d'efficacité démontrée*" (version 2) devient "*forte présomption clinique de l'efficacité du traitement hormonal dans la prévention du risque fracturaire, présomption méritant d'être confirmée par des études cliniques rigoureuses*" (version 3).

Ce renversement de position révèle le caractère négocié du texte final et témoigne d'une dynamique conflictuelle entre les spécialistes invités à intervenir dans la conférence de consensus et les non spécialistes composant le jury.

De son point de vue, le jury, poussant les experts à expliciter la force de la preuve scientifique sous-tendant leurs présentations, déclare n'avoir rencontré que des convictions ou des arguments d'autorité.

¹⁴ La Présidente du jury n'exercera aucune pression, stricto sensu, sur le jury, mais laissera transparaître, tout le long de la conférence, sa vision personnelle sur la médicalisation de la ménopause (la surveillance médicale spécifique des femmes à cet âge de la vie est prioritaire, à ses yeux, sur la prescription d'un traitement hormonal). Ce comportement sera interprété par plusieurs experts comme un "manque d'objectivité flagrant".

De leur côté, les experts reconnus du domaine n'ont pas l'habitude d'avoir à apporter la preuve de ce qu'ils avancent devant des non spécialistes qui, d'ordinaire, les croient “sur parole”.

Comme le souligne un membre du comité d'organisation, “*la démarche particulière des conférences de consensus n'a pas été acceptée en profondeur par tous les participants. Il n'est pas facile, et parfois impossible, pour un spécialiste d'accorder sa confiance à un jury pluridisciplinaire dont aucun membre n'est expert reconnu de la question traitée*” (généraliste, comité d'organisation).

Il est intéressant de noter ici que ce rapport de forces entre experts et non experts se prolongera au delà de la présentation par le jury de ses “conclusions définitives” à la presse : la version diffusée par la Fondation de l'Avenir quelques mois plus tard présentera, ce qui n'est d'ailleurs pas conforme aux règles du jeu, un certain nombre de modifications substantielles supplémentaires opérées sous la pression des spécialistes¹⁵. Dans les semaines suivant la conférence de consensus, la

¹⁵A la fin de la conférence de consensus, inquiet par l'état de fatigue manifeste du jury, le coordinateur s'engagera à faire parvenir les conclusions aux experts avant leur publication définitive. Dans les deux mois qui suivront la conférence de presse, il recevra ainsi une vingtaine de lettres proposant des modifications, certaines en provenance du jury. Ces rectifications seront soumises, par courrier, à l'approbation de l'ensemble des membres du jury. La version 4 comporte ainsi deux modifications, l'une touchant à la densitométrie, l'autre à la formulation de la conclusion du “consensus”. Plusieurs experts ayant considéré le passage concernant la densitométrie osseuse comme “*une condamnation de la technique*”, l'énoncé “*cet examen ne peut être recommandé comme argument de décision thérapeutique*” (version 3), deviendra “*cet examen ne peut être recommandé comme argument unique de décision thérapeutique*” (version 4). Par ailleurs, certains trouvant la formulation de la conclusion du “consensus” trop négative, l'énoncé, “*le traitement hormonal substitutif de la ménopause ne peut être préconisé systématiquement à l'ensemble de la population concernée*” (version 3) sera remplacé par “*le traitement hormonal substitutif de la ménopause peut être proposé et recommandé mais certainement pas imposé à l'ensemble de la population concernée*” (version 4).

Société Française de Rhumatologie, considérant le jury incompetent et ses conclusions scandaleuses envisagera une “réaction officielle” et l'Association Française pour l'Etude de la Ménopause (AFEM), pour des motifs similaires, projetera d'organiser, entre experts cette fois, une “contre conférence de consensus”.

Revenons à nouveau aux raisons possibles du renversement de position du jury au cours de la conférence même. Composé d'une majorité de non spécialistes de la ménopause, ses membres se sont donc retrouvés dans une situation inconfortable face aux experts du thème. Comme l'exprime l'un des jurés, *“nous avons eu raison de dire que l'efficacité anti-fracturaire des oestrogènes n'était pas démontrée. Si nous avons atténué cette conclusion, c'est que nous nous sommes laissés impressionner par les experts. (...) Il fallait quand même arriver à un texte qui donne le moins de prise possible à la polémique (...) De toutes façons, dans un débat contradictoire entre "hyper-sachants" et "peu-sachants", les peu-sachants se retrouvent toujours en position de faiblesse (...) Si on avait pris le temps de bien réfléchir, si on s'était donné deux jours de plus, on aurait pu mieux bétonner nos arguments...”* (médecin et journaliste, jury)

Aux Etats-Unis, les membres du jury des conférences initiées par l'OMAR sont experts du thème, préparent la trame du consensus avant la conférence, commencent à rédiger le jour de la première séance publique et ont l'après-midi, la soirée et la nuit du second jour pour rédiger les conclusions provisoires (9).

Laisser à un jury non spécialisé le temps d'une nuit pour rédiger la première version de ses conclusions a laissé planer un doute, tant chez les experts qu'au sein du jury¹⁶, sur la qualité du texte rédigé dans de telles conditions. Le risque d'un tel marathon nocturne est également de conduire à un "consensus" forcé et précipité.

Le degré de préparation du jury en amont et le temps octroyé à la rédaction des conclusions définitives en aval peuvent contrebalancer les effets d'une session nocturne unique sur la qualité des productions du jury.

CONCLUSIONS

"Lorsque nous parlons de controverse, écrit Bruno Latour, nous supposons déjà qu'il y a des adversaires bien identifiés; un lieu où leurs arguments ont pu se rencontrer; un ordre du jour commun afin de décider les points à discuter; une série d'épreuves jugées déterminantes; des juges acceptables par tous afin de décider qui a gagné ou qui a vaincu..." (20). Comparées aux controverses telles que les connaissent des sciences "dures" comme la physique ou la chimie, les controverses du domaine médical semblent ne pouvoir être qu'exceptionnellement closes au nom de la seule science. Un exemple, fréquent en médecine, en est donné par les controverses autour de

¹⁶ "Après un gros effort de synthèse à faire très vite, je crois que ni les uns, ni les autres, n'étions en état de critiquer les propositions des autres. On était trop fatigués" (généraliste, jury) ou encore "Il n'y avait pas de temps, donc pas de possibilité d'échanger vraiment. Nous avons abouti à un texte comportant des points de vue hétérogènes. Comme un article écrit par quatre personnes différentes avec des bouts de phrase de l'un ou de l'autre, et puis ça ne tient pas vraiment debout. Ce qui est tout autre chose qu'un travail d'élaboration commun avec du temps, de la réflexion, des échanges." (gérontologue, jury).

l'établissement de seuils séparant le normal du pathologique. Cette catégorie de débats ne peut déboucher, selon Tristram Engelhardt Jr et Caplan, sur un mode de résolution rationnel et définitif, fondé sur l'appel aux faits et au raisonnement rigoureux sur ces faits, car *“il n'est pas possible de faire appel à un observateur moralement neutre ou à un groupe d'observateurs désintéressés en vue d'obtenir une réponse. Une réponse particulière reflète un jugement de valeur particulier”* (32).

Du point de vue de leurs organisateurs, les conférences de consensus offrent un cadre propice au règlement de controverses, alors que leur finalité immédiate n'est pas la clôture de controverses scientifiques mais la production de recommandations destinées aux praticiens. L'opinion selon laquelle les conférences de consensus permettent le règlement de controverses a déjà été contestée par plusieurs travaux sociologiques (33; 24; 35; 6).

Si l'un des objectifs possibles d'une conférence de consensus est d'amener le jury à prendre position dans une controverse pour mettre les parties adverses en accord autour d'un texte, des représentants des diverses “écoles” devraient être présents dans le comité d'organisation, les experts présentant les divers points de vue devraient impérativement être confrontés autour de la même question pour argumenter leurs positions en public et face au jury.

Sur un point controversé, l'exigence de présentations d'experts solidement argumentées différenciant clairement conviction, présomption et preuve, mérite d'être renforcée. Un temps suffisant devrait être donné

aux experts, tant pour leurs exposés que pour les débats, de façon à laisser au jury la possibilité d'assimiler réellement les termes de la controverse.

Les questions controversées pourraient faire, en conséquence, l'objet de sessions spéciales dans le programme des conférences de consensus.

De plus, le jury devrait être choisi selon des critères qui lui permettent d'être reconnu, en tant qu'entité pluridisciplinaire, comme "compétent" par les parties adverses. Le temps et les moyens nécessaires à l'élaboration d'une prise de position solide au plan scientifique doivent lui être donnés. Si la "compétence" du jury est une notion qui reste quelque peu indéfinie, la capacité d'analyse critique de l'information médicale devrait en constituer un élément essentiel.

Si l'ensemble de ces conditions n'est pas rempli, le "consensus" éventuellement atteint risque fort, nous semble-t-il, de n'être que celui du jury et l'un des effets prévisibles de la conférence de consensus sera... l'exacerbation stérile de la controverse.

La question de savoir si le jury, en fonction des questions scientifiques considérées, doit recevoir comme consigne de s'appuyer sur les données disponibles, même si elles sont imparfaites, ou bien de ne se référer strictement qu'aux études méthodologiquement incontestables, nous semble finalement être une fausse question.

Si l'objectif d'une conférence de consensus est avant tout d'apporter aux praticiens des moyens pour prendre des décisions dans des situations

où les données de la science sont incertaines, le jury ne devrait pas avoir à choisir entre des données qui seraient “incontestables” et des données qui ne le seraient pas, mais à présenter les données disponibles selon leur caractère de quasi-certitude, de présomption ou de simple conviction.

Pour adapter sa pratique aux données de la littérature, le praticien a besoin de se faire une idée relativement précise de la solidité de ces données. Faute de temps et de formation, il ne peut faire seul la recherche documentaire, l'analyse et la synthèse des publications originales lui permettant de décider quel crédit accorder à une information importante pour son exercice quotidien. L'explicitation des niveaux de preuve sur lesquels reposent les conclusions de conférences de consensus et, plus largement, les recommandation de pratique clinique répond à un besoin. Le Groupe de travail canadien sur l'examen de santé périodique (Canadian Task Force on the Periodic Health Examination) a été historiquement le premier à s'être efforcé de rendre accessible aux praticiens, dès 1979, la présentation des preuves à l'appui

de ses recommandations (7) ¹⁷. L'US Preventive Services Task Force, organisme fédéral américain chargé d'évaluer les fondements scientifiques des pratiques préventives, adoptera, dès sa création en 1984, la classification graduée élaborée par la Canadian Task Force on the Periodic Health Examination (34). En 1986, le canadien David Sackett, innovant au delà du strict domaine de la prévention primaire et du dépistage couvert par les deux groupes nord-américains, sera le premier à proposer une classification originale des recommandations selon le niveau de preuve applicable, cette fois, aux pratiques thérapeutiques (30).

Les organisateurs de conférences de consensus et les promoteurs de recommandations de pratique clinique ont aujourd'hui à leur disposition de multiples grilles d'analyse critique de la qualité des publications scientifiques (5,17) ainsi que plusieurs classifications hiérarchisant les niveaux de preuve et la force des recommandations. Cependant, rares

¹⁷ Chacune des études disponibles portant sur le dépistage examiné par le groupe (palpation des seins, frottis cervical, pression artérielle, etc.) a été classée dans l'une ou l'autre de quatre catégories méthodologiques numérotées de I à IV, du niveau de preuve le plus élevé (I : "essai contrôlé randomisé bien conduit") au niveau de preuve le plus bas (IV : opinions d'experts fondées sur l'expérience clinique, études descriptives ou consensus informels d'experts). Dans un second temps, les recommandations ont été graduées par le groupe de travail selon la force des preuves disponibles en faveur ou en défaveur de l'inclusion du dépistage considéré dans un examen de santé périodique : l'existence d'au moins une étude de niveau I, selon qu'elle est favorable ou défavorable à la pratique du dépistage périodique, conduit à une recommandation forte, notée respectivement A ou E; l'existence d'études de niveau II (études convergentes de cohorte ou cas-témoin valides, études comparatives temporelles ou géographiques, etc.) conduit à une recommandation de force moyenne, notée B ou D, selon le sens de leurs conclusions; l'inexistence d'études ou des données de niveau IV conduisent à une recommandation de force C, reposant sur un niveau faible de preuves ne permettant pas de conclure sur l'inclusion ou l'exclusion du dépistage dans un examen de santé périodique (7).

sont les organismes ou les sociétés savantes, surtout en France, qui ont fait un usage pertinent et performant de ces outils pourtant indispensables à la production de recommandations de qualité qui soient utilisables par les praticiens dans leurs décisions quotidiennes. Leur maniement par des jurys de conférences de consensus parfaitement formés à la lecture critique des publications médicales mériterait d'être testée dans notre pays. L'estimation d'un niveau de preuve, souligne Anne Fagot-Largeault, ne peut résulter de l'application mécanique d'un raisonnement déductif : les liens de causalité sont toujours jugés sur un faisceau composite de preuves fragmentaires dont la valeur résulte d'une appréciation humaine (14). Les jurys de conférences de consensus n'ont-ils pas justement pour rôle essentiel de réaliser une telle appréciation dans l'interdisciplinarité, l'indépendance intellectuelle et la transparence méthodologique?

Ce travail a bénéficié du soutien de l'Agence Nationale pour le Développement de l'Evaluation Médicale ainsi que de l'INSERM U158. Les citations et la description du déroulement de cette conférence de consensus telles qu'elles apparaissent dans l'article ont été validées par l'ensemble des membres du jury et des experts cités.

BIBLIOGRAPHIE

- 1. Abenhaïm L. La décision, l'exemple du risque technologique. In Gilles Brücker et Didier Fassin ed(s). Santé Publique. Paris : Marketing, 1989 : 824-39.**
- 2. Andem. Les conférences de consensus. Bases méthodologiques pour leur réalisation en France. Rapport du Service des Etudes, Agence Nationale pour le Développement de l'Evaluation Médicale : Paris, 1990 : 15 pages.**
- 3. Andem. Guide pratique pour la réalisation d'une conférence de consensus. Rapport du Service des Etudes, Agence Nationale pour le Développement de l'Evaluation Médicale : Paris, 1992 : 47 pages.**

4. Bell SE . Changing ideas: the medicalisation of menopause. Soc Sci Med, 1987; 24 : 535-42.
5. Bergmann JF, Chassany O, Segrestaa JM et al. Examen critique des grilles de lecture des essais thérapeutiques. Thérapie 1994; 50 : 181-4.
6. Broclain D. Cholestérol et risque coronarien: le choix des stratégies de prévention n'est-il que scientifique? In P.Aïach, N.Bon et J.P.Deschamps ed(s) Comportements et Santé, questions pour la prévention. Nancy : Presses Universitaires de Nancy, 1992 : 130-39.
7. Canadian Task Force on the Periodic Health Examination. The periodic health examination. Can Med Assoc J. 1979; 121 : 1193-1254.
8. Chrzanowski R. Programme de développement de consensus aux National Institutes of Health, USA. Méd et Hyg 1988; 46 : 933-37.
9. Council of health care technology. Consensus development at the NIH, improving the program. Institute of Medecine, Washington : National Academy Press, 1990, 81 p.
10. Delanoë D. Représentations savantes et profanes de la ménopause. Modifications d'un phénomène de nature par des processus de culture. (Mémoire pour le DEA en Anthropologie Sociale et Ethnologie). Ecole des Hautes Etudes en Sciences Sociales, Paris : septembre 1992, 132 p.
11. Desrosières A. La politique des grands nombres, histoire de la raison statistique. Editions La Découverte, Paris 1993, 437 p.
12. Effective Health Care. Screening for osteoporosis to prevent fractures. Effective Health Care 1992; 1, School of Public Health, University of Leeds, England.
13. Evidence-based medicine group. Evidence-based medicine. JAMA 1992; 268 : 2420-25.
14. Fagot-Largeault A, Les causes de la mort. Histoire naturelle et facteurs de risque. Librairie Philosophique J.Vrin, Paris 1989, 420 pages.
15. Freidson E. La profession médicale. Payot, Paris, 1984, 369 p.
16. Goldman L, Tosteson A. Uncertainty about postmenopausal estrogen (editorial). N Engl J of Med 1991; 325 : 800-2.

17. Guyatt GH. User's guides to the medical literature (editorial). *JAMA* 1993; 270 : 2096-7.
18. Jacoby I. Evidence and consensus (editorial). *JAMA* 1988; 259: 3039
19. Jacoby I. Sponsorship and role of consensus development programs within national health care systems. In Goodman C and Baratz SR ed(s) *Improving consensus development for health technology assessment: an international perspective*. Washington DC : National Academy Press, 1990 : 163 p.
20. Latour B. Pasteur et Pouchet: hétérogénéité de l'histoire des sciences. In M.SERRES ed. *Éléments d'histoire des sciences*. Paris : Bordas, 1989 : 423-445.
21. Law MR, Wald NJ, Meade TW. Strategies for prevention of osteoporosis and hip fracture. *BMJ* 1991; 303 : 453-59.
22. Lomas J, Anderson G, Erkin M et al. The role of evidence in the consensus process. *JAMA* 1988; 259 : 3001-5.
23. Lomas J. Words without action? The production, dissemination and impact of consensus recommendations. *Annu Rev Publ Health* 1991; 12 : 41-65.
24. Markle GE, Chubin DE. Consensus development in biomedicine: the liver transplant controversy. *The Milbank Quarterly* 1987; 65 : 1-24.
25. McGlynn EA, Kosecoff J, Brook RH. Format and conduct of consensus development conferences. Multination comparison. *Intl J of Techn Assess in Health Care* 1990; 6, 450-69.
26. Mullan F et Jacoby I. The town meeting for technology. The maturation of consensus conferences. *JAMA* 1985; 254 : 1068-72.
27. Perry S. Consensus development, an historical note (editorial). *Intl J of Techn Assess in Health Care* 1988; 4 : 482-4.
28. Preventing osteoporosis (Corespondence). *BMJ* 1991; 303 : 920-22.
29. Ross PD, Heilbrun LK, Wasnich RD et al. Methodological issues in evaluating risk factors for osteoporotic fractures. *J Bone Min Res* 1989; 4: 649-56.

30. Sackett DL "Rules of evidence and clinical recommendations on the use of antithrombotic agents", *Chest* 1986; 86 (suppl) : 2S-3S.
31. Stocking B, Jennett B, Spiby J. *Criteria for change, the history and impact of consensus development conferences in the UK*. King's Fund Center, Londres 1991.
32. Trisram Engelhardt Jr H et Caplan AL. *Patterns of controversy and closure: the interplay of knowledge, values and political forces*. In Trisram Engelhardt Jr H et Caplan AL eds *Scientific controversies*. Cambridge (MA) : Cambridge University Press, 1987 : 1-23.
33. Vinokur A, Burnstein E, Sechrest L et al. *Group decision making by experts: field study of panels evaluating medical technologies*. *J Person and Soc Psychol* 1985; 49 : 70-84.
34. Woolf SH, "Practice guidelines, a new reality in medicine. II. Methods of developing guidelines.", *Arch Intern Med* 1992; 152 : 946-52.
35. Wortman P, Vinokur A, Sechrest L. *Do consensus conferences work? A process evaluation of the NIH consensus development program*. *J Health Politics, Policy and Law* 1988; 13 : 469-498.