
ÉVALUATION DES THÉRAPEUTIQUES NON MÉDICAMENTEUSES : L'EXEMPLE DU THERMALISME

Bernard Graber-Duvernay

Rédacteur en chef de la Presse thermale et climatique

Les exigences en matière d'évaluation thérapeutique se sont beaucoup accrues au cours des dernières décennies. Elles sont aujourd'hui d'autant mieux appliquées qu'on dispose de bons modèles conceptuels.

Le risque est de prendre parfois ces modèles pour des recettes ou des panacées, d'application presque automatique. Un retour aux sources du raisonnement évaluatif peut se montrer utile, notamment dans un domaine aussi complexe que celui des thérapeutiques non médicamenteuses.

Le médicament

La thérapeutique la plus facile à évaluer est le médicament parce que son administration se prête bien aux procédés expérimentaux et, notamment, aux essais comparatifs en double-aveugle contre placebo qui demeurent la référence première en matière de techniques d'évaluation.

Tout ne se résume pas cependant dans le double-aveugle contre placebo, même pour le médicament. Il faut aussi déterminer l'objectif de l'étude selon qu'est visée la mise en évidence de l'efficacité ou de l'utilité du médicament. Les deux termes peuvent être confondus par le sens commun. Ils expriment cependant des notions distinctes, *efficacité* voulant dire présence d'un effet physiologique favorable alors qu'*utilité* signifie intérêt dans l'arsenal thérapeutique actuel. Lorsqu'une molécule nouvelle a montré chez l'animal de laboratoire des propriétés jugées prometteuses, elle est testée chez l'homme en plusieurs étapes qui impliquent chacune un nombre croissant de sujets puis de patients : la phase I vérifie l'absence de toxicité ; la phase II apprécie l'efficacité probable et teste les meilleures posologie et voie d'administration ; la phase III confirme à plus grande échelle l'efficacité et les risques et situe la molécule nouvelle par rapport aux médicaments existant ; la phase IV décèle d'éventuels effets indésirables rares ou tardifs parfois longtemps après la mise sur le marché. C'est donc à la phase III que se situent les essais thérapeutiques proprement dits avec leur double finalité de détermination de l'efficacité et de l'utilité du médicament.

Daniel Schwartz [1] a dénommé *explicatifs* les essais qui visent à démontrer l'efficacité parce que « plus dirigés vers la recherche fondamentale » et *pragmatiques* ceux qui s'intéressent à l'utilité parce que « plus dirigés vers l'application pratique » (p 34). « L'attitude explicative vise à apporter une réponse sur le plan de la compréhension ou

de la *connaissance* ; elle peut en apporter une par surcroît sur le plan de l'application pratique, mais cette extrapolation n'est aucunement assurée, elle dépend notamment du résultat de l'essai. L'inverse a lieu pour l'attitude pragmatique. » (p.36).

Un essai peut être clairement explicatif ou pragmatique. Il arrive aussi qu'il soit mixte, c'est à dire capable d'être envisagé sous l'un ou l'autre aspect. Il faudra pourtant choisir celui des deux éclairages qu'on souhaite privilégier parce que les deux types d'essais s'opposent presque point par point.

- Par le traitement comparé tout d'abord, la démarche explicative réclamant l'absence d'alternative ou un placebo ou un unique traitement de référence alors que la pragmatique oppose toujours une autre thérapeutique ou une autre stratégie à celle qui intègre le médicament étudié; de plus, les conditions d'administration sont tenues d'être aussi fixes que possible dans l'attitude explicative qui réclame leur égalisation, alors qu'elles peuvent varier dans l'attitude pragmatique qui vise à l'optimisation des conditions, en adaptant, par exemple, la posologie aux réactions du sujet.

- Par le choix des critères de jugement ensuite, qui seront biologiques ou instrumentaux dans la démarche explicative alors que dans la démarche pragmatique ils « expriment le point de vue du malade » (p.35) en matière de longévité et de qualité de vie.

- Par la constitution de l'échantillon qui doit être homogène et constitué de sujets sensibles à l'effet recherché dans l'essai explicatif et refléter la population cible dans le pragmatique.

- Par le mode de comparaison, enfin, qui emploie des hypothèses et des méthodes statistiques différentes, qui considère autrement les sorties d'étude et qui calcule différemment le nombre de sujets nécessaires.

Les deux types d'essais peuvent être conduits simultanément ou successivement. D Schwartz, qui déconseille le mélange des genres, ajoute : « Ainsi est-on obligé le plus souvent de se limiter à un seul des deux problèmes, c'est à dire de trancher le dilemme, certes infiniment délicat : préfère-t-on éclairer la connaissance dans l'immédiat, avec l'espoir d'applications pratiques ultérieures, ou obtenir une application pratique immédiate mais moins bien comprise et moins féconde pour l'avenir ? » (p.37).

Il reste une dernière dimension au delà de la tolérance et de l'efficacité absolue ou comparée d'un médicament. Elle a trait à la vie de ce médicament, c'est à dire à la manière dont il s'insère dans les conduites thérapeutiques des médecins et du public en fonction de son rapport bénéfice/risques mais aussi de facteurs sociaux et historiques qui ne se limitent pas toujours au pouvoir persuasif de la firme qui le fabrique ou le commercialise. C'est l'objet de la pharmacoépidémiologie qui étudie non seulement l'efficacité et le risque mais aussi l'usage des médicaments [2] c'est à dire leur consommation, leur prescription, et leur approche pharmacoéconomique, à quoi il faut ajouter le rôle du médicament comme médiateur entre la société et les individus malades [3].

Thérapeutiques non médicamenteuses

La problématique est comparable. On peut aborder, par exemple, la chirurgie ou la radiothérapie sous les deux aspects : explicatif si l'on souhaite apprécier les conséquences physiologiques d'un geste chirurgical ou la radiosensibilité d'un type de cancer ; pragmatique si l'on souhaite vérifier l'intérêt de l'application de cette chirurgie ou de cette

irradiation à la population de malades et en préciser les modalités.

Une différence importante existe, cependant, avec le médicament, qui est la difficulté de créer des placebos : possible pour la radiothérapie, très difficilement réalisable pour la kinésithérapie ou les approches psychothérapeutiques, et pratiquement inapplicable en chirurgie.

Quand elle n'est pas simulable, une thérapeutique ne peut pas être attribuée à l'insu du patient et les essais qui l'évaluent sont dits « ouverts » et exposés à l'effet placebo. On verra plus loin que cela empêche la démarche explicative qui ne peut tolérer aucune suggestion mais qu'il n'en va pas de même dans l'attitude pragmatique où la procédure du double aveugle, si elle est toujours bonne à prendre, n'est plus impérative.

Les étapes de l'évaluation thérapeutique (Tableau I)

Elles amènent à constater et à mesurer l'effet thérapeutique, puis à s'interroger sur sa relation de cause à effet avec l'agent testé, puis sur sa réalité objective, et enfin sur son utilité.

1. *La mesure de l'effet thérapeutique* suppose le choix des variables les plus pertinentes et de l'instrument apte à les mesurer, et de l'échantillon humain sur lequel se fera l'observation.

Les variables désignées comme critères de jugement peuvent être uniques ou multiples et, alors, soit hiérarchisées avec un critère principal, soit regroupées en une combinaison que D Schwartz appelle « résumé exhaustif » dans la démarche pragmatique.

La combinaison de variables en un critère unique est un problème fréquent, rencontré aussi dans le domaine du diagnostic, qui donne lieu à divers procédés dont les plus courants sont l'addition simple, l'addition après pondération, et la segmentation qui tient compte d'opérations logiques plus complexes [4]. Se trouvent ainsi constitués des indices composites, des plus mathématiques (indice de masse corporelle liant le poids et

Tableau I - Les étapes de l'évaluation thérapeutique

Mesure de l'effet	<ul style="list-style-type: none"> • Instruments de mesure pertinents (critères de jugement) • Observatoire assez vaste (échantillon de patients) • Observateur qualifié (compétent et neutre) 	
Relation de cause à effet	<i>Méthodes comparatives</i> (tableau II)	<ul style="list-style-type: none"> • Situation d'intervention • Situation d'observation
	<i>Méthodes non comparatives</i>	<ul style="list-style-type: none"> • répétitions • avis des patients
Objectivité de l'effet	<i>Attitude explicative</i>	Évaluation contre placebo (insu du patient indispensable)
	<i>Attitude pragmatique</i>	Évaluation contre d'autres traitements (absence d'insu accepté)
Utilité de l'effet (service médical rendu)	<ul style="list-style-type: none"> • Rapport bénéfice/risques • Gravité des affections soignées • Place dans l'arsenal thérapeutique de ces affections 	

la taille) aux plus conventionnels (critères diagnostiques de l'Acr). Le « résumé exhaustif » de D Schwartz contient les éléments de la décision binaire (oui ou non) d'utiliser un traitement et doit donc prendre en compte tous les critères de jugement estimés utiles à la décision. Ce sera le plus souvent une combinaison pondérée. « Il est évident que l'établissement de la pondération est une opération très délicate. Mais on ne voit pas qu'elle puisse être évitée au niveau de la décision ; elle est donc absolument logique dans une attitude pragmatique » (p.53), quitte à publier séparément les résultats de chacun des critères, l'utilisateur gardant la liberté « de reconsidérer la question avec une pondération différente de celle qu'ont adoptée les auteurs de l'essai ».

La compétence de l'observateur participe à la qualité de l'instrument de mesure ; elle doit comprendre sa neutralité, c'est à dire sa parfaite indifférence aux conclusions de l'étude ; en pratique, sa participation en aveugle.

L'échantillon humain requis diffère avec l'objectif de l'étude : homogène et fait de sujets réceptifs dans la démarche explicative, « tout venant » représentatif de la *population cible* dans la démarche pragmatique. Il doit être assez important pour que soient dépassées les fluctuations d'échantillonnage de façon à laisser l'effet apparaître.

2. *La recherche des relations de cause à effet*, qui nourrit la curiosité de l'homme dans son environnement depuis la nuit des temps, a été à la fois passive par l'observation et la mémoire d'événements liés, et active par la répétition d'actions et la constatation que les mêmes interventions produisent les mêmes effets.

La technique des *études comparatives*, rendue possible par les progrès de la méthode statistique, est d'apparition récente puisqu'elle remonte au début du XX^{ème} siècle. Elle s'est développée en Angleterre dans les années 1920 [5]. Elle a d'abord été appliquée à l'agronomie qui devait faire face à des situations expérimentales complexes en raison du grand nombre de facteurs susceptibles d'intervenir dans une expérience (sol, exposition, humidité, présence d'insectes...) et de la modicité des effets attendus de beaucoup d'interventions (engrais, sélection de semences...), situations qui exigeaient de grandes surfaces et du temps. La nouvelle technique, en organisant la répartition des surfaces par le tirage au sort de parcelles, a augmenté considérablement la productivité de la recherche agronomique.

La méthode, qui se développa aussi dans l'industrie et l'économie, fut appliquée à la biologie, puis à la psychologie et à la médecine et notamment à la recherche thérapeutique après la seconde guerre mondiale lorsqu'il fut pris conscience que les conditions étaient très voisines de celles de l'agronomie en raison de l'existence de facteurs confondants difficiles à identifier ou à maîtriser et de la fréquente discrétion de l'effet thérapeutique attendu dans la plupart des essais. On a pu alors conclure beaucoup plus vite à partir d'échantillons restreints. La recherche thérapeutique clinique fit des progrès rapides en établissant autant l'inefficacité de traitements consacrés par un usage ancien que l'existence de propriétés favorables pour des thérapeutiques nouvelles.

Cependant, le recours aux essais comparatifs n'est pas indispensable en toutes circonstances. Il est inutile lorsque l'efficacité est si manifeste qu'elle « saute aux yeux » comme l'emploi d'un antibiotique dans une affection microbienne¹ ou le remplacement d'une hanche arthrosique par une prothèse totale ; le simple regroupement d'observations suffit alors. D'autre part, ce recours peut être impossible pour des raisons éthiques

lorsqu'on souhaite tester, par exemple, une molécule nouvelle dans le traitement d'une maladie grave qui dispose déjà d'un médicament efficace et que le temps peut compter.

La comparaison entre deux séries de sujets distincts peut être transférée sur les sujets de la même série alternativement traités et témoins (« sujet pris pour son propre témoin »). On peut rapprocher de cette méthode les démarches non comparatives que sont la méthode *des mesures répétées* [6] qui consiste en l'observation prolongée de patients chez qui l'on observe l'effet d'une intervention thérapeutique sur le cours de leur maladie, ou les enquêtes d'opinion demandant aux patients de comparer eux-mêmes les thérapeutiques qu'ils ont expérimentées.

La méthode comparative n'implique pas que les chercheurs interviennent de bout en bout. Ils peuvent aussi laisser les sujets se déterminer eux-mêmes vis à vis de l'agent étudié. C'est le cas des enquêtes épidémiologiques étiologiques [7] (Tableau II) à la recherche d'une relation entre un agent et une pathologie (tabac et cancer du poumon, graisses animales et athérome) ou, au contraire, de l'effet bénéfique de certaines conduites (allaitement maternel et pathologie du nourrisson).

Enfin, une nouvelle approche dite *naturaliste* [8] se développe dans le monde anglo-saxon. Elle est adaptée à l'étude des phénomènes transdisciplinaires et cherche à faire appel aux modes de raisonnement de la vie courante.

3. *La nature objective* de l'effet observé est l'autre source de préoccupation des chercheurs. Les hésitations sémantiques autour de la notion d'objectivité rendent bien compte du flou du concept : effet physiologique ? effet intrinsèque ? Objectif s'oppose ici à subjectif (et non à fictif) et signifie que l'effet thérapeutique ne doit rien à la suggestion. Comment s'en assurer chez des patients qui apportent chez leurs thérapeutes leur culture, leur histoire, leur imaginaire et toute leur problématique à l'égard de leur maladie et de la manière de la soigner ?

Tableau II - Les situations d'intervention et d'observation des études comparatives

<i>situations d'intervention</i>	<i>situations d'observation</i>
1. recrutement d'un échantillon de patients	<u>Type exposés-non exposés</u>
2. répartition en deux groupes (randomisation > appariement)	1. recrutement d'un échantillon de sujets exposés à une circonstance donnée
⊕ groupe A soumis au traitement A	2. appariement d'un échantillon parmi des sujets non exposés à la circonstance
⊕ groupe B non traité	3. mesure de la conséquence présumée dans les deux échantillons
- ou soumis au traitement B	
- ou soumis à un placebo	<u>Type cohorte prospective</u>
3. mesure de l'effet dans les deux groupes et analyse statistique des résultats	id° l'observation commence avant l'exposition
	<u>Type cas-témoins</u>
	Enquête rétrospective à partir de la conséquence présumée à la recherche du facteur causal

¹ « Il n'est pas besoin d'une méthodologie fine pour reconnaître le premier médicament efficace dans une maladie grave, comme par exemple la streptomycine dans la méningite tuberculeuse. » écrit D Schwartz [1]

On s'y efforce de deux façons. Tout d'abord par l'emploi de critères de jugement d'ordre biologique ou instrumental supposés insensibles à l'inconscient des patients comme à celui des observateurs. Cette attitude peut s'avérer réductrice dans la mesure où elle prive l'observation d'un grand nombre d'informations utiles en provenance des patients ; l'introduction des échelles visuelles analogiques, des indices fonctionnels et des échelles de qualité de vie (*Patient-reported outcomes* ou Pro) a été considérée comme un grand progrès dans les procédures d'évaluation. Par ailleurs l'emploi de critères de jugement objectifs ne lève pas toujours toute ambiguïté. Un travail italien récent [9] met en évidence une élévation des taux d'ACTH, cortisol et béta-endorphines chez des fibromyalgiques à l'issue d'une cure thermale. Cette montée des hormones du bien-être doit-elle être considérée comme un effet physiologique de la thérapeutique thermale ou comme la traduction physiologique d'un effet placebo, ou les deux ? Les mécanismes de l'autosuggestion ne sont pas clairs. Un travail de chercheurs canadiens de l'Université de Colombie britannique apporte un éclairage surprenant. Ces auteurs ont vu se réactiver les circuits physiologiques de production de la dopamine chez des parkinsoniens à la suite d'une injection de sérum physiologique présentée comme un produit nouveau [10] ; les zones activées n'appartiennent pas au domaine cognitif mais aux circuits mêmes de la production du médiateur défaillant, le placebo se révélant un traitement plus physiologique que la dopathérapie substitutive habituelle.

L'autre manière de s'assurer de la nature objective de l'effet thérapeutique observé est de faire en sorte que le patient ne sache pas à quel traitement il est soumis ; c'est l'aveugle du patient, en général accompagné de celui du thérapeute (double aveugle) et, bien entendu, de celui de l'observateur qui devrait être acquis quelque soit la méthode employée.

Il convient ici de revenir à la distinction explicatif/pragmatique.

Dans la démarche explicative qui est le prolongement en clinique humaine de l'étape de la détermination en laboratoire de l'effet physiologique d'un geste thérapeutique, aucune suggestion des sujets ne doit pouvoir se produire sous peine de retirer toute crédibilité à la recherche entreprise.

Dans la démarche pragmatique, le chercheur se trouve en face d'une situation en vraie grandeur qu'il ne contrôle pas mais qu'il va devoir appréhender suffisamment pour pouvoir prédire les conséquences de l'introduction d'un apport thérapeutique. Il serait utopique pour lui de vouloir extirper d'un ensemble hétérogène de facteurs tout ce qui s'apparente à la suggestion. L'effet placebo peut même passer pour un facteur thérapeutique comme un autre à intégrer dans la stratégie testée. D Schwartz est explicite sur ce point (p. 48) : l'effet placebo peut être « soit égalisé dans les deux groupes en rendant le traitement indiscernable (attitude explicative), soit incorporé au traitement étudié (attitude pragmatique) » et, plus loin : « En définitive, l'élimination de l'effet placebo n'est pas toujours obligatoire ; dans certains cas, purement pragmatiques, elle n'est même pas souhaitable. »

4. L'appréciation de l'utilité de l'effet thérapeutique constitue, en fin de compte, l'étape déterminante de l'évaluation. Le rapport bénéfice/risques d'une thérapeutique lui donne un droit d'entrée dans l'arsenal des traitements de la pathologie visée ; le degré d'utilité est donné par le rang dans cet arsenal. Cette valeur est formalisée en France par la notion de *service médical rendu* (Smr). Il revient à un organisme officiel, la commission de la transparence, d'établir le Smr de chaque classe thérapeutique et l'Asmr (amélioration du

Smr) de chaque thérapeutique nouvelle qui conditionnera le niveau de sa prise en charge par l'assurance maladie.

La commission se prononce « après analyse des données disponibles » pour toute thérapeutique ayant obtenu l'autorisation de mise sur le marché (Amm). Elle prend en compte « son efficacité, sa sécurité d'emploi, sa place dans la stratégie thérapeutique, la gravité de l'affection à laquelle elle est destinée, ses conditions réelles d'utilisation et son intérêt en termes de santé publique »¹.

Les essais explicatifs réalisés pour l'obtention de l'Amm du médicament font évidemment partie des données disponibles sur lesquelles se prononce la commission mais son domaine est typiquement, par définition, celui des essais pragmatiques.

Randomisation et comparabilité

Le mot *randomisation* est un anglicisme qui désigne l'action de répartir au hasard. Il est bien traduit par *tirage au sort* (dans son acception ancienne de désignation des conscrits) et s'utilise plus facilement que l'expression *allocation aléatoire* qu'on a voulu introduire.

La finalité de l'essai comparatif est de pouvoir conduire une double observation conjointe sur deux groupes de patients dont l'un seulement est soumis au traitement qu'on souhaite tester. Un essai thérapeutique se déroule dans le temps pendant une période plus ou moins longue ; il y a un avant et un après l'application du traitement ou sa mise en route s'il est continu. Pendant cette période, d'autres facteurs que le traitement vont se manifester, comme l'évolution naturelle de l'affection et diverses influences susceptibles de faire varier les manifestations pathologiques. L'important est d'éviter d'attribuer au traitement testé un effet qui ne provient pas de lui (risque α ou de première espèce) ou de ne pas observer un effet réel par manque de puissance de l'étude (risque β ou de deuxième espèce). C'est à cela que servent les témoins de la série « contrôle » qui vont être soumis à tous les aléas de la série traitée, le traitement en moins, si bien qu'il suffira d'une soustraction pour connaître ce qui revient réellement au traitement. Pour que les témoins puissent remplir leur rôle, certaines conditions doivent être remplies qui relèvent du bon sens et tiennent en un mot : la comparabilité. Les patients des deux séries doivent être interchangeables et le rester pendant la durée de l'étude.

1. *La constitution des groupes comparables* peut se faire par appariement : chaque fois qu'un patient est recruté dans un groupe, un autre patient aux mêmes caractéristiques est attribué à l'autre groupe. Ce système, qui peut offrir des commodités stratégiques, a l'inconvénient de ne pas prendre en compte tous les facteurs d'évolution de la maladie dont certains ne sont pas connus.

Le recours au tirage au sort va faire intervenir l'ensemble de ces facteurs tout en dispensant de leur recensement puisque le hasard va répartir équitablement chacun d'eux, connu et inconnu ; en théorie en tout cas : les fantaisies du hasard sont nombreuses et les essais contrôlés n'ont pas de raison d'être épargnés, surtout lorsque les effectifs sont réduits. Divers procédés dits d'équilibrage ont été mis au point pour réduire ces distorsions de façon à assurer l'égalité des effectifs ou celle des principaux

¹ Site Web de l'AFSAPS.

facteurs pronostiques connus. Le tirage au sort est alors segmenté, réajusté tout au long du recrutement, ce qui revient à réintroduire un peu d'appariement.

La randomisation des conditions d'expérience est étroitement liée à la méthode comparative. Son intérêt n'est pas contestable. Il serait vain de vouloir revenir sur le progrès considérable qu'elle a représenté dans la recherche scientifique, mais elle n'est pas obligatoire. Parmi les études qui n'ont pu y avoir recours pour des raisons diverses, beaucoup ont permis des constatations précieuses dont il aurait été regrettable de se priver. La randomisation ne peut pas être considérée comme une fin en soi ; elle demeure un moyen.

2. *Le maintien de la comparabilité* tout au long de l'étude suppose qu'aucun changement ne survienne chez les patients hors effets du traitement (ou du hasard). Le tirage au sort initial permet, en principe, d'éviter ce risque sauf pour ce qui touche au traitement lui-même. D Schwartz donne l'exemple du traitement anticoagulant qui peut induire un changement du mode de vie des patients qui le reçoivent en clair par l'évitement de toutes les circonstances dangereuses de leur vie avec, pour conséquence, l'introduction d'un biais dans l'évaluation des risques. Cet exemple sert à son plaidoyer en faveur de l'aveugle du patient auquel il ajoute celui du médecin, ou double-aveugle, qui complète le maintien de la comparabilité en évitant des différences de comportement médical en fonction du traitement attribué. On peut aussi penser que, dans certaines études ouvertes, les patients ont un préjugé a priori pour l'une des alternatives mises au tirage au sort avec risque d'autosuggestion positive ou négative selon l'option attribuée.

Ce type d'atteinte à la comparabilité pour des raisons de plus grande confiance dans un traitement est moins grave dans l'attitude pragmatique qui prend en compte tous les effets du traitement y compris du domaine psychologique. On comprend ce qu'il peut y avoir de choquant dans cette façon pragmatique de traiter sur le même pied une propriété physiologique et un effet placebo. Elle pourrait même passer pour la négligence d'une des missions du service public qui est la protection contre les abus de confiance. On est certainement plus à l'aise pour entreprendre un essai thérapeutique ouvert quand on a pu prouver au préalable la réalité d'un effet physiologique même faible. A défaut de cette démarche par manque d'essai explicatif pour cause de placebo impossible, il faut se résigner au risque de conclure à l'utilité d'un effet thérapeutique dépourvu de substratum physiologique. Il est bien dans la logique de l'attitude pragmatique, dirigée vers l'application pratique à un moment donné de la connaissance scientifique et exprimant le point de vue du malade, d'envisager en bloc l'impact d'une thérapeutique toutes causes confondues. La moralisation viendra de la suite : un effet placebo isolé ou un effet physiologique faible n'auront qu'un temps.

Le fait de savoir si les assurances maladie doivent prendre en charge la thérapeutique est un autre problème. Il relève des choix des politiques budgétaires dans le domaine de la santé. La France s'est dotée d'un bon outil avec le Smr.

3. *La comparabilité des observations*, ou égalité de mesure, est la dernière exigence des essais comparatifs. Elle est si évidente qu'elle ne mérite de mention que par les difficultés que son application rencontre parfois. Si l'observateur est unique, il doit être compétent, c'est-à-dire doté de bons instruments de mesure correctement utilisés mais aussi parfaitement neutre à l'égard des conclusions de l'étude. Si les observateurs sont multiples, ils doivent recueillir leurs informations de façon identique. Quand les critères

de jugement sont complètement objectifs aux variations de la mesure près (biologie, radiologie, anapath, tracés électriques...) et que l'essai se fait en double aveugle, l'égalité de mesure est présumée parfaite. A l'inverse, dans une étude pragmatique ouverte, où le patient sait quel traitement il utilise et où les critères de jugement sont subjectifs (Eva, indices fonctionnels ou de qualité de vie), c'est de l'objectivité et de la qualité de la mesure que va dépendre la valeur et la crédibilité de l'étude. Les censeurs des protocoles devront être très exigeants sur le choix des critères de jugement, leur validité, leur adéquation aux objectifs, les conditions de leur emploi (les précisions manquent souvent sur les conditions dans lesquelles les patients renseignent leurs questionnaires) et sur l'aveugle de l'examineur lorsque le recueil de certaines données réclame son intervention. Il est manifeste que cet aveugle de l'examineur est bien difficile à garantir lorsque le patient est informé de son traitement. La solution est, sans doute, de renoncer aux examens physiques en ne faisant intervenir que des assistants préposés au recueil d'auto-questionnaires et mis dans l'incapacité d'influencer les réponses de quelque manière que ce soit.

L'évaluation du thermalisme

Le traitement par l'eau des sources thermales, appelé aussi crénothérapie, se distingue des autres thérapeutiques non médicamenteuses par au moins trois caractéristiques : il est multifactoriel (eau, techniques thermales, techniques non thermales et programmes de pédagogie médicale, dépaysement, psychothérapie institutionnelle) ; il n'est simulable à l'insu des patients que pour l'un de ses facteurs, l'eau thermale, qui n'est sans doute pas le plus important ; il suppose une grande disponibilité des utilisateurs aussi bien en matière de temps que de dépenses consenties.

Ces caractéristiques du thermalisme confèrent à son évaluation une spécificité parmi les autres techniques non médicamenteuses concernant le choix des sujets et les modalités de l'évaluation.

Un nombre notable de travaux ont été consacrés à cette évaluation [11,12]. Ils ont cherché soit à minimiser la spécificité thermale pour appliquer au thermalisme les principes généraux de l'évaluation thérapeutique, soit à la prendre en compte en innovant par rapport à l'évaluation du médicament. Nous utiliserons certains de ces travaux pour illustrer les considérations qui vont suivre.

I - Le problème du recrutement des patients dans les études thermales

L'attitude pragmatique implique que le recrutement ait lieu dans la population cible des utilisateurs de la thérapeutique. De façon générale, ce terme désigne le tout venant des porteurs de la pathologie visée susceptibles de recevoir le traitement.

Dans le thermalisme, il s'ajoute des exigences de disponibilité. On voit ainsi apparaître une catégorie de patients, curistes avérés ou potentiels, chez qui le thermalisme pourra être évalué à côté d'autres chez qui il ne pourra pas l'être. A-t-on le droit de réduire le recrutement aux patients qui manifestent leur disponibilité à suivre une cure thermale ?

Cette attitude serait condamnable en logique explicative qui vise à établir des lois de portée universelle mais elle se justifie en logique pragmatique puisque la population cible est bien celle des curistes potentiels.

Il n'est d'ailleurs pas dit que la population de curistes soit très différente d'une population tout venant. Une enquête chez des curistes rhumatisants a montré une répartition socioprofessionnelle superposable à celle de la statistique Insee de la même année [13] ; en 1995, un sondage Csa a révélé l'attrait du thermalisme dans une grande partie de la population¹ [14].

Se pose ensuite la question des modalités du recrutement.

Certaines études ont réalisé le *recrutement sur place* parmi des patients habitant à proximité de la station thermale [15,16,17] ; d'autres, à distance, chez des patients qui se rapprochaient davantage des curistes habituels [18].

Le recrutement sur place est beaucoup plus commode. Il a néanmoins l'inconvénient de ne pas prendre en compte tous les facteurs thérapeutiques d'une cure (dépaysement, repos) et d'ajouter des contraintes spécifiques (déplacements quotidiens). Il peut aussi receler la possibilité d'un biais par manque de neutralité des patients en faveur du thermalisme, une sorte d'« effet supporteur » traduisant l'attachement des sujets à une activité locale. Comme les utilisateurs venus d'ailleurs n'auront aucune raison de le manifester, on peut le considérer comme propre à l'échantillon de l'essai qui se trouve démuné de ce fait de « validité externe », c'est à dire d'aptitude à voir appliquer à d'autres populations les conclusions de l'étude.

Le recrutement à distance est plus satisfaisant mais il implique des structures plus lourdes et il pose le problème – très lié à la spécificité thermale – de la prise en charge des frais de la cure.

La gratuité des traitements est prévue par la loi Huriet-Sérusclat selon le principe que le coût de l'étude est à la charge des promoteurs [19]. Dans le thermalisme qui utilise des cures de trois semaines, on imagine mal comment l'attribution gratuite d'un séjour à l'hôtel de cette durée pourrait ne pas modifier la neutralité du sujet. Le risque serait grand de faire apparaître un « effet reconnaissance » que n'empêcherait pas le choix d'un comparateur de même gratification si le promoteur est encore thermal.

Cette difficulté du recrutement conduit à préférer les méthodes d'observation aux méthodes d'intervention (Tableau II). D'autres arguments vont aller dans le même sens.

II - Modalités de l'évaluation

Elles découlent de ce qui a été dit plus haut sur les manières d'établir une relation de cause à effet entre la thérapeutique appliquée et l'évolution observée : méthodes d'intervention ou d'observation, à visée explicative ou pragmatique, comparatives ou non, avec ou sans tirage au sort, contre quel comparateur ?

1. Essais explicatifs

Plusieurs travaux, surtout étrangers, ont visé à évaluer l'eau ou les boues thermales comme un médicament : mise au point de placebos crédibles, insu du patient, tirage au

¹ Sondage CSA-Santé Magazine sur 1000 personnes de 18 ans et plus. Février 1995

	oui (tout à fait – assez)	non (pas vraiment - pas du tout)
Croyez-vous aux bienfaits d'une cure thermale ?	68% (37 - 61)	28% (15 - 13)
Souhaitez-vous suivre une cure thermale ?	48% (26 - 22)	48% (21 - 27)

sort. Les essais en rhumatologie sont rapportés dans l'article qui suit celui-ci. Un travail italien de la sphère Orl [20] peut servir d'exemple. Il compare en double aveugle des inhalations d'une eau thermale contre d'autres à l'eau de ville sur le coryza chronique sur des critères objectifs : rhinométrie, temps de transport du mucus, cytologie nasale, analyse chimique et bactériologique du mucus. Les résultats, à la fin du traitement, sont favorables à l'eau thermale.

Ce beau travail, typiquement explicatif, marque bien les limites du genre. Tout d'abord, toutes les eaux thermales ne sont pas simulables ; certaines odeurs, couleurs ou consistances peuvent être difficiles à imiter. Ensuite, l'efficacité de l'eau ou de ses dérivés n'est qu'un des facteurs de la cure. Une étude autrichienne a montré que le nombre et la nature des soins expliquaient moins de 10% de la variance des effets d'une cure thermale chez 153 patients atteints de lombalgie chronique [21]. Enfin, établir la réalité d'un effet intrinsèque n'a d'intérêt que si l'utilité de l'effet a été montrée au préalable : que valent la rhinométrie et la cytologie nasale si les patients n'observent pas de changement ou si un court traitement antihistaminique les aurait soulagés autant à moindre coût ? Le pragmatique doit précéder l'explicatif.

On peut estimer que les essais explicatifs de ce type n'ont plus leur place dans la campagne d'évaluation du thermalisme qui commence. La volonté qu'ils expriment d'opposer aux détracteurs du thermalisme des preuves de la vertu thérapeutique de l'eau thermale appartient au passé. Il s'agit maintenant d'évaluer l'utilité de la thérapeutique thermale telle qu'elle se présente aujourd'hui. Cette évaluation repose sur les essais pragmatiques, même si l'exercice est difficile et qu'il se prête mal à l'emploi des règles de l'évaluation du médicament.

2. Essais pragmatiques comparatifs en situation d'intervention

Le schéma est celui de la première colonne du tableau II. La cure est attribuée, après tirage au sort, à une partie de l'échantillon recruté préalablement tandis que l'autre partie reçoit le comparateur. Les questions portent sur le choix du comparateur et sur la randomisation.

Le comparateur

Faute de placebo capable de simuler une cure thermale complète, la conduite comparée à la cure ne peut être que l'abstention ou une autre stratégie thérapeutique.

L'absence de traitement a été choisie par plusieurs études thermales importantes [15,16,18]. Cela est revenu peu ou prou à comparer une attitude active, le traitement thermal, à une attitude de non intervention avec le risque, en étude ouverte, que les patients lisent le questionnaire de l'étude sous la forme : « Préférez-vous qu'on s'occupe de vous ou qu'on ne vous fasse rien ? » à quoi la réponse est évidente et sans grand rapport avec l'efficacité du traitement. Cela a pu conduire, pour maintenir les témoins dans l'étude, à les gratifier à leur tour d'une cure thermale mais différée jusqu'à la fin de l'observation, faisant ainsi apparaître un biais « liste d'attente » qui n'est que le déplacement de la question précédente vers la constatation : « On ne m'a encore rien fait. » Noter que la technique du placebo a été introduite pour éviter ce biais.

Lorsqu'un comparateur existe, d'autres biais peuvent se produire. On a vu plus haut que le maintien de la comparabilité ne doit pas souffrir de l'attribution du traitement. Dans

une étude ouverte, le patient reconnaît le traitement qu'il reçoit et sait quel traitement il aurait reçu si le sort en avait décidé autrement. Il peut regretter le choix du sort et exprimer inconsciemment sa déception en sous-estimant les signes d'amélioration spontanée de sa maladie. Le maintien des patients dans l'ignorance du traitement comparé à leur pourrait pallier cet inconvénient mais il est interdit par la loi Huriet qui impose que les patients soient complètement éclairés avant l'entrée dans l'étude, et donc avant le tirage au sort¹.

Tirage au sort

La randomisation est devenue une « valeur culte » en terme d'évaluation ; elle seule confère le top niveau aux essais thérapeutiques dans les grilles de lecture, les revues de la littérature, les échelles de niveau de preuve. Tout chercheur s'efforce de l'inclure dans ses mots-clés².

Or, le tirage au sort ne se pratique qu'en situation d'intervention. Pour que le protocole d'une étude comporte un tirage au sort, il faut donc qu'il s'agisse d'un essai comportant le recrutement en une fois de tout l'échantillon et l'attribution active des thérapeutiques à comparer avec les aléas qu'on vient de voir dans les études ouvertes concernant le recrutement et le maintien de la comparabilité.

Il existe cependant des méthodes plus conformes à la nature de la thérapeutique thermale et donc plus aptes à l'évaluer si l'on veut bien renoncer au tirage au sort. Ce sont les méthodes en situation d'observation. Avant de les envisager, il reste à exposer une autre formule avec randomisation.

Le sujet pris pour son propre témoin

Un échantillon est sélectionné parmi les patients porteurs de la pathologie cible acceptant de suivre une cure thermale. Deux séries sont constituées dans cet échantillon par le tirage au sort de la date de la cure. Certains patients la font au début de l'étude (sujets traités) tandis que les autres sont d'abord témoins avant de devenir traités à leur tour.

La méthode a l'avantage de réduire les problèmes de recrutement et de comparateur tout en augmentant la puissance de l'étude. Elle donne la possibilité d'interroger les patients sur leur préférence entre les deux traitements testés.

En pratique, d'autres problèmes ont dû se poser puisqu'aucune étude n'a été publiée mentionnant ce type de randomisation.

3. Essais pragmatiques comparatifs en situation d'observation

Les modèles figurent dans la seconde colonne du tableau II. La cure thermale est considérée comme un agent auquel les patients se sont exposés librement (comme ils auraient bu ou fumé si l'étude avait porté sur les méfaits de l'alcool ou du tabac). L'effet attendu n'est pas une maladie créée par l'agent mais l'amélioration d'une pathologie préexistante (comme l'allaitement maternel améliore les défenses insuffisantes du nourrisson).

¹ Dans cet ordre d'idée, il est important que toute étude randomisée ouverte comprenne dans son questionnaire initial un item sur le souhait du sujet face à l'alternative thérapeutique proposée.

² Noter, toutefois, que la grille de l'Anaes (Annexe I) attribue le même niveau 2 de preuve aux *Essais comparatifs randomisés de faible puissance* et aux *Etudes comparatives non randomisées bien menées*, et même aux *Etudes de cohorte*.

Un échantillon de curistes porteurs de la pathologie cible est sélectionné et apparié à un autre échantillon de patients non curistes soumis à une autre thérapeutique programmée ou poursuivant leur traitement habituel. Il n'y a pas de tirage au sort ce qui a l'avantage de ne pas mettre les patients en situation d'avoir à regretter leur répartition. Ils peuvent même ne pas être informés de la branche de l'étude à laquelle ils n'appartiennent pas. En fait, il s'agit de deux études d'observation prospectives menées simultanément sur deux groupes distincts.

Une étude parfaitement pragmatique [22] a apparié un groupe de curistes suivant une cure thermale pour une pathologie de l'épaule avec un autre groupe de patients porteurs d'une pathologie identique et de même grade soignés dans un service de rééducation fonctionnelle hospitalier. L'étude a montré une efficacité égale pour un coût moindre dans la station thermale.

Un autre exemple est donné par une étude, rapportée dans ce numéro et [23], d'une station du Massif Central qui a comparé 200 colopathes curistes à 200 colopathes recrutés par téléphone par les médecins prescripteurs au même moment et dans les mêmes régions. Deux questionnaires de qualité de vie ont été soumis à l'entrée dans l'étude et à 6 mois. Par ailleurs, un troisième groupe de 200 sujets non colopathes a été apparié sur l'âge et le sexe dans la population pour vérifier l'étalonnage du questionnaire dans les conditions de l'étude. A 6 mois, les deux groupes se sont améliorés significativement, le groupe de curistes plus que l'autre.

4. Les études non comparatives

Elles ne comportent aucune intervention, consistant seulement à suivre l'évolution d'un groupe de patients soumis à une cure thermale. Appelées études de cohorte, elles sont mal situées dans certaines grilles de niveaux de preuve (Annexe I). Elles sont cependant très riches d'informations lorsqu'elles sont bien conduites (prospectives avec des critères validés et des conditions d'observation rigoureuses) et ne recèlent pas les biais signalés à propos des études comparatives.

Elles peuvent d'ailleurs présenter, dans certains cas, une valeur comparative proche du modèle du sujet pris pour son propre témoin lorsque les patients sont suivis assez longtemps pour offrir à l'observation des périodes sous et hors influence de la thérapeutique ou lorsqu'ils sont interrogés sur leurs préférences c'est à dire quand il leur est demandé de faire la comparaison eux-mêmes.

Méthode des mesures répétées

La méthode, défendue par deux auteurs américains [6] est exposée dans l'annexe II. Quand on interroge des patients à intervalles réguliers sur leur affection à l'aide, par exemple, d'un indice fonctionnel ou d'une échelle de qualité de vie, on projette une sorte de grille temporelle sur leur maladie qui en rend l'évolution aussi lisible que sur un graphique. On crée ainsi les conditions idéales d'une observation pour une intervention thérapeutique qu'on pourra renouveler comme on répète une expérience de physiologie une fois le montage expérimental effectué. Une telle enquête a été réalisée sur l'effet du traitement thermal de l'arthrose de la hanche ou du genou [25] en interrogeant les patients à plusieurs reprises au téléphone à l'aide de l'indice de Lequesne. Ce travail a eu les honneurs d'un débat méthodologique [26]. Il lui a été reproché de ne pas démontrer « l'efficacité *intrinsèque* de la cure thermale » pour les raisons que, d'une part,

l'échantillon n'a pas été prélevé au hasard dans une population de malades mais parmi des curistes qui peuvent ne pas être neutres vis à vis du traitement testé et, d'autre part, de ne pas détecter l'effet placebo. Le problème de la représentativité de l'échantillon thermal a été abordé plus haut à propos de la population cible. Il se pose dans toutes les études thermales (et même dans toutes les études, tout sujet pressenti étant libre de refuser de participer à l'essai d'une thérapeutique qui ne lui convient pas). Le second reproche – ne pas distinguer l'effet placebo – revient à refuser toute évaluation qui n'utiliserait pas le double aveugle. On a vu que ce reproche ne se justifiait pas dans l'attitude pragmatique qui, dans les études ouvertes, prend en compte l'effet placebo avec les autres facteurs de l'action thérapeutique et qui est bien celle qui inspire cette étude.

Rien ne s'oppose, d'ailleurs, à ce que ces observatoires des pathologies soient utilisés pour des comparaisons, soit avec d'autres interventions thérapeutiques, soit avec l'évolution naturelle de la maladie en l'absence d'intervention. Il se pourrait alors que se trouve ainsi réalisé, sans double aveugle ni randomisation, le meilleur modèle pour l'évaluation de la thérapeutique thermale.

Les enquêtes de préférence thérapeutique

Une enquête dans une station rhumatologique a consisté à demander à 379 curistes de noter 9 thérapeutiques courantes de leur pathologie, dont la cure thermale, sur une échelle verbale à 5 classes [24]. La cure vient en tête avec 80% de réponses " bon " ou " excellent ". Les limites ne tiennent pas au caractère subjectif de ces opinions (qui ont la même valeur que les réponses aux Eva ou aux échelles fonctionnelles) mais aux conditions de l'enquête menée sur le lieu de la cure, en pleine période d'investissement psychologique des patients.

D'autres approches existent pour appréhender le phénomène thermal qui portent sur les populations de curistes, leur répartition et leur attente [24], et sur les répondants à la crénothérapie [27].

En conclusion

Ce travail s'est efforcé de dissiper certaines confusions qui règnent parfois autour du concept d'évaluation thérapeutique en revenant à la distinction fondamentale entre les attitudes explicatives et pragmatiques définies et développées par Daniel Schwartz. Certaines de ces confusions peuvent être entretenues par la notion de *niveaux de preuve* utilisée dans les grilles de lecture et par les experts, qui fait une place telle aux conditions de l'essai explicatif que les essais pragmatiques s'en trouvent lourdement handicapés. Les niveaux de preuve ne se rapportent, en réalité, qu'à l'efficacité et il est normal qu'ils s'établissent sur les critères de l'essai explicatif. Un travail pragmatique ne devrait pas être jugé à cette aune mais à la mesure de l'apport du traitement testé à l'arsenal thérapeutique de l'affection considérée.

Faute de quoi, les promoteurs d'études, dont l'objectif est la valorisation ou la justification de la thérapeutique qu'ils défendent, chercheront à faire passer leurs évaluations sous les fourches caudines de la méthode explicative, quitte à leur imposer des distorsions susceptibles d'ôter toute signification réelle à leurs résultats.

Le thermalisme en a donné plusieurs exemples dont certains ont été exposés ici. Faisant

partie des thérapeutiques traditionnelles qui ont gardé droit de cité, dont le domaine d'application évolue avec les progrès des autres thérapeutiques, il doit constamment montrer dans quelles indications il continue à rendre service. Il n'a plus à prouver une quelconque « activité intrinsèque » : faute de placebo, il n'y parviendrait pas et cela n'aurait plus beaucoup de sens si longtemps après son apparition sur le marché. De toute façon, cela ne le dispenserait pas de la démarche pragmatique, la plus apte à établir la première qualité qu'on lui demande de prouver : son utilité.

Résumé

Le thermalisme, comme la plupart des thérapeutiques non médicamenteuses, ne peut pas être simulé par un placebo, sauf pour l'un de ses facteurs : l'eau thermale. Il doit donc être évalué de façon ouverte pour l'essentiel de son action. Cela ne constitue pas une gêne pour les études pragmatiques destinées à montrer l'utilité de la thérapeutique thermale dans ses principales indications, d'autant plus que l'utilité est la principale qualité à démontrer pour une thérapeutique qui est présente « sur le marché » depuis longtemps. Les études pragmatiques ouvertes s'accommodent mieux des méthodes d'observation dérivées de celles de l'épidémiologie étiologique que des techniques expérimentales réclamées par les essais explicatifs. Leur qualité repose sur le choix des critères de jugement, la neutralité de l'observateur, et les conditions de l'observation. Les diverses étapes de la démarche évaluative sont passées en revue et servent de base à la présentation critique de certains des principaux travaux thermaux. Deux méthodes sont présentées comme plus aptes que d'autres à évaluer la thérapeutique thermale : l'une est dérivée de l'épidémiologie étiologique, l'autre de la méthode des mesures répétées.

Références

1. Schwartz D, Flamant R, Lellouch J. *L'essai thérapeutique chez l'homme*. 2ème édition. Paris, Flammarion Médecine-Sciences, 1981.
2. Begaud B. - *Dictionnaire de Pharmacoépidémiologie*. 3ème édition. Bordeaux, Arme-Pharmacovigilance éditions, 1998.
3. Lebrun Th. La nécessaire prise en compte de l'utilité sociale du médicament dans l'évaluation du service médical rendu. *Quot Méd* 2003;n° 7369 (24 juillet).
4. Dougados M. *La Mesure. Initiation du clinicien au langage statistique et méthodologique*. Paris, Expansion scientifique française, 1996.
5. Schwartz D. *Méthodes statistiques à l'usage des médecins et des biologistes*. 4ème édition. Paris, Flammarion Médecine-Sciences, 1993.
6. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research* pp. 37-43. Boston. Houghton Mifflin. 1963.
7. Rumeau-Rouquette Cl, Blondel B, Kaminski M, Bréart G. *Épidémiologie Méthodes et pratique*. Paris, Flammarion Médecine-Sciences, 1993.
8. House Ernest. *Evaluating with Validity*.
9. Bellometti S, Galzigna L Function of the hypothalamic adrenal axis in patients with fibromyalgia syndrome undergoing mud-pack treatment. *Int J Clin Pharm Res* 1999;19(1):27-33.
10. Fuente-Fernandez (de la) Raul, Thoma J, Rossi R V, Schulzer M, Calne D B, Stoessl A J. Expectation and Dopamine Release : Mechanism of the Placebo effect in Parkinson Disease. *Science* August 2001;293:1164-66.

11. Queneau P, Boulangé M, Françon A et al. *Médecine thermale - Faits et preuves*. Masson Coll Abrégés, Paris 2000.
12. Queneau P, Françon A, Graber-Duvernay B. Réflexions méthodologiques sur 20 essais cliniques randomisés en crénothérapie rhumatologique. *Thérapie* 2001;56 : 675-84.
13. Graber-Duvernay B, Berthier N, Berthier F. Enquête psycho-sociologique auprès de 900 curistes des Thermes nationaux d'Aix-les-Bains. *Press Therm Climat* 1994;131(2):72-80.
14. *Santé-Magazine* avril 1995;232:126-28.
15. Guillemin F, Constant F, Collin JF, Boulangé M. Short and long-term effect of spa therapy in chronic low back pain. *Br J Rheumatol* 1994 ;33:148-151.
16. Constant F, Collin JF, Guillemin F, Boulangé M. Effectiveness of spa therapy in chronic low back pain : a randomized clinical trial. *J Rheumatol* 1995;22:1315-20.
17. Graber-Duvernay B, Françon A, Forestier R. Efficacité du berthollet d'Aix-les-Bains sur les manifestations fonctionnelles de l'arthrose des mains - Essai thérapeutique contrôlé. *Rhumatologie* 1997;49(4):141-156.
18. Nguyen M, Revel M, Dougados M. Prolonged effects of 3 week therapy in a spa resort on lumbar spine, knee, and hip osteoarthritis : follow-up after 6 months. A randomized controlled trial. *Br J Rheumatol* 1997;36:77-81.
19. Dangoumau J. L'argent dans les essais cliniques. Le point de vue de la Cnamts. *Lett Pharmac* 2001;15(6):116-17.
20. Marullo T, Abramo A. [Effects of one cycle of inhalation crenotherapy with radioactive fluoridated oligomineral]. [Article in Italian] *Acta Otorhinolaryngol Ital* 2000 Aug;20(4 Suppl 63):1-13.
21. Strauss-Blasche G, Ekmekcioglu C, Vacariu G, Melchart H, Fialka-Moser V, Marktl W. Contribution of individual spa therapies in the treatment of chronic pain. *Clin J Pain*. 2002 Sep-Oct;18(5):302-9.
22. Guillaumin B, Engel M, Montastruc P. Pathologies capsulo-ligamentaires de l'épaule. Evaluation clinique de la crénothérapie à Amélie-les-Bains. *Presse Therm Climat* 1994;131(4):209-214.
23. Chareyras JB, Porte M, Bommelaer G. Enquête sur la qualité de vie du colopathe. Résultats bruts à 1 mois et à 6 mois après une cure thermale à Châtel-Guyon ; comparaison avec une population de non curistes. *Presse Therm Climat* 1998;135(3):151-154.
24. Graber-Duvernay B, Emin JF, Blattier F, Duplan B. Opinion de patients sur quelques thérapeutiques courantes en rhumatologie. *Rhumatologie* 1996;48(2):43-47.
25. Forestier R. Amplitude et suivi de l'effet de deux cures thermales successives sur la gonarthrose et la coxarthrose. Enquête prospective ouverte chez 51 curistes recrutés consécutivement. *Rev Rhum [Ed Fr]* 2000;67:427-36.
26. Bouvenot G, Ambrosi P. Evaluation de la crénothérapie en rhumatologie. *Rev Rhum [Ed Fr]* 2000;67:408-10.
27. Graber-Duvernay B, Berthier N, Berthier F. Enquête répondeurs - non répondeurs en crénothérapie. Recherche des facteurs prédictifs de l'efficacité d'une cure thermale en rhumatologie. *Press Therm Climat* 1997;134(3):186-92.



ANNEXE 1

Les niveaux de preuve

Selon Sackett

Niveau 1	Plusieurs essais randomisés avec $\alpha < 0,05$ et $1-\beta > 0,9$ ou méta-analyse valide
Niveau 2	Un seul essai randomisé avec $\alpha < 0,05$ et $1-\beta > 0,9$ ou plusieurs essais randomisés avec $\alpha < 0,05$ et $1-\beta < 0,9$
Niveau 3	Etudes contrôlées non randomisées Etudes cas-témoins Etudes de cohortes Séries de cas Consensus d'experts (accord professionnel fort)

Grille Anaes

Tableau des niveaux de preuve et grades de recommandations donnés par l'Anaes dans le document « Diagnostic, prise en charge et suivi des malades atteints de lombalgie chronique » (Service des recommandations et références professionnelles, 20 septembre 2000).

Niveaux de preuve scientifique	Grade des recommandations
Niveau 1 ❖ Essais comparatifs randomisés de forte puissance ❖ Méta-analyse d'essais comparatifs randomisés ❖ Analyse de décision basée sur des études bien menées	A Preuve scientifique établie
Niveau 2 ❖ Essais comparatifs randomisés de faible puissance ❖ Etudes comparatives non randomisées bien menées ❖ Etudes de cohorte	B Présomption scientifique
Niveau 3 ❖ Etudes cas-témoins	C
Niveau 4 ❖ Etudes comparatives comportant des biais importants ❖ Etudes rétrospectives ❖ Séries de cas	Faible niveau de preuve

A partir du JAMA ; 1995

ANNEXE 2

The TIME-SERIES EXPERIMENT

(D'après Campbell Donald T, and Stanley Julian C. *Experimental and quasi-experimental designs for research*. Boston. Houghton Mifflin. 1963. pp. 37-43)

Les séries temporelles de mesures permettent d'apprécier l'effet d'un changement expérimental beaucoup mieux que les seules études avant/après.

La figure ci-contre représente par des points chacune des mesures répétées à intervalles réguliers. Le trait vertical en pointillé désigne l'intervention expérimentale. Le double trait entre les deux points à cheval sur le trait pointillé de chacune des 5 séries indique la variation observée dans les deux mesures qui encadrent l'intervention (avant/après). La variation est identique sur les 5 séries mais on voit bien que sa signification n'est pas la même. La relation de cause à effet est très probable dans les séries A, B et C et très improbable dans les deux autres.

La principale faiblesse du test est la possibilité du biais dit 'historique' qui est l'intervention simultanée d'un facteur différent de l'événement plus ou moins lié à lui. Son évitement nécessite une très bonne délimitation de l'événement expérimenté.

Les autres biais sont peu probables, ou absents pour ceux qui tiennent à

l'existence d'un groupe contrôle : 'maturation' qui est dû à la modification spontanée de l'échantillon au cours de l'étude ; 'testing' qui suppose une réaction différente à l'examen lorsqu'on le répète ; 'instrumentation' qui survient lorsqu'il se produit un changement dans la méthode de mesure (changement d'appareil, changement d'examineur) ; 'régression' vers la moyenne quand les sujets sélectionnés ont, au départ, des scores extrêmes par rapport à la population globale ; 'sélection' lorsque les sujets ne sont pas comparables pour les principales variables pronostiques ; 'mortalité' qui concerne les perdus de vue ; 'interactions' diverses entre les sujets et l'étude.

