

Groupe d'experts et auteurs

Olivier CANCEIL, Service hospitalo-universitaire de santé mentale et thérapeutique, Centre hospitalier Saint-Anne, Paris

Jean COTTRAUX, Unité de traitement de l'anxiété, Hôpital Neurologique Pierre Wertheimer, CHU de Lyon

Bruno FALISSARD, Département de santé publique, hôpital Paul Brousse, AP-HP, Villejuif

Martine FLAMENT, Royal Ottawa Hospital Care Group, Ottawa, Canada

Jacques MIERMONT, Unité de consultation familiale, Centre hospitalier Paul Guiraud, Villejuif

Joël SWENDSEN, Laboratoire de psychologie clinique et psychopathologie, Institut universitaire de France, Université Victor Segalen, Bordeaux 2

Mardjane TEHERANI, Service de psychiatrie, Centre Hospitalier Xavier Bichat, Paris

Jean-Michel THURIN, psychiatre, Paris

Ont été auditionnés

Daniel WIDLÖCHER, Université Pierre et Marie Curie, Paris

David SERVAN-SCHREIBER, Psychiatrie clinique, Faculté de médecine, Université de Pittsburgh, États-Unis

Ivy BLACKBURN, Cognitive and behavioural therapies centre, Newcastle upon Tyne, Royaume-Uni

Coordination scientifique et éditoriale

Fabienne BONNIN, attachée scientifique au Centre d'expertise collective de l'Inserm, Faculté de médecine Xavier-Bichat

Catherine CHENU, chargée d'expertise au Centre d'expertise collective de l'Inserm, Faculté de médecine Xavier-Bichat

Jeanne ETIEMBLE, directeur du Centre d'expertise collective de l'Inserm, Faculté de médecine Xavier-Bichat

Catherine POUZAT, attachée scientifique au Centre d'expertise collective de l'Inserm, Faculté de médecine Xavier-Bichat

Assistance bibliographique et technique

Chantal RONDET-GRELLIER, documentaliste au Centre d'expertise collective de l'Inserm, Faculté de médecine Xavier-Bichat

I

Données générales sur l'évaluation de l'efficacité

1

Réflexion sur l'évaluation

Ce document s'intéresse à l'évaluation scientifique de l'efficacité des psychothérapies, mais peut-on évaluer scientifiquement l'efficacité d'une psychothérapie ? En effet, la situation psychothérapeutique met en jeu des êtres humains dans leur singularité subjective. Comment envisager alors de réduire scientifiquement un tel contexte sans en perdre l'essence ? Ces questions ont bien entendu été abordées au cours de cette expertise collective.

On peut considérer en première approximation que la démarche scientifique consiste à tester des hypothèses réfutables au moyen d'expériences reproductibles. Nous trouvons ici trois concepts clés à approfondir dans le cadre particulier de l'évaluation des psychothérapies : hypothèse, expérience reproductible, réfutabilité. Dans le cas présent, la ou les hypothèses à tester sont clairement identifiées : telle modalité de psychothérapie est plus efficace que telle autre modalité thérapeutique. La possibilité de mettre en œuvre des expériences reproductibles permettant de réfuter ces hypothèses est moins évidente. En effet, peut-on réaliser des « expériences » dans un cadre psychothérapeutique sans interférer avec le processus de soin et donc sans biaiser les résultats ? De telles « expériences » devront le plus souvent évaluer le sujet dans sa subjectivité ; peut-on réaliser des mesures subjectives de la même façon que l'on réalise des mesures objectives, et ces mesures ont-elles la même valeur ? Ces « expériences » porteront sur des sujets singuliers dans leur trajectoire de vie et dans leur fonctionnement mental ; comment envisager alors des expériences reproductibles ? Enfin à propos de la notion de réfutabilité, elle fait écho, en recherche clinique, à celle d'inférence statistique au travers notamment des tests statistiques d'hypothèses. Mais peut-on avoir recours à la statistique dans le cadre de l'évaluation de psychothérapies comme on le fait dans le cadre de l'évaluation des médicaments ?

« Expériences » dans un cadre psychothérapeutique

Ce point concerne en fait la recherche thérapeutique en général et pas seulement le domaine des psychothérapies. Ainsi, dans un essai thérapeutique comparant l'efficacité d'un antiulcéreux à un placebo à l'aide d'une méthodologie de « double aveugle » et d'allocation aléatoire du traitement, il y a effectivement une modification du contexte de soin par rapport à la situation clinique habituelle : ni le médecin ni le malade ne savent quel traitement est administré. Rien ne dit alors que la différence d'efficacité constatée en situation expérimentale est véritablement superposable à la différence d'efficacité existant en pratique clinique quotidienne. Certains champs de la clinique sont plus sensibles que d'autres à ce biais potentiel, la psychiatrie en est sûrement un, mais elle n'est pas la seule. Les autorités de santé sont conscientes de ce problème et suggèrent, pour apporter des éléments de réponse à cette question, la réalisation d'études observationnelles de terrain, il est vrai moins rigoureuses et plus difficilement interprétables, mais dont les résultats permettent de corroborer utilement les données issues des essais thérapeutiques classiques.

L'évaluation des psychothérapies relève de la même logique : l'essai thérapeutique comparatif apporte une information capitale du fait de la rigueur et de la transparence de la

méthodologie mise en jeu. Cette information ne se suffit néanmoins pas à elle-même, elle doit toujours être interprétée à la lumière des spécificités de la pratique clinique réelle et pourra éventuellement gagner à être étayée par des études complémentaires, qualitatives ou épidémiologiques.

Mesures objectives, mesures subjectives

Pour évaluer une psychothérapie, il faut le plus souvent évaluer le sujet dans sa subjectivité. Peut-on réaliser des mesures subjectives (tristesse ou intensité d'un délire par exemple) de la même façon que l'on réalise des mesures objectives (tension artérielle ou cholestérolémie par exemple) et ces mesures ont-elles la même valeur ?

Cette question est essentielle. L'expérience montre en effet que, spontanément, le plus grand nombre considère comme une évidence que « la tristesse, ça ne se mesure pas... », ou que « vouloir mesurer la tristesse, c'est perdre *de facto* sa substance, sa complexité ».

Il y a là en réalité un quiproquo concernant la définition du mot mesure. Ce mot, d'usage parfaitement courant, est pourtant très difficile à définir. Si les médecins utilisent implicitement une théorie opérationnelle de la mesure (les résultats des mesures peuvent être manipulés à l'aide de formules et de lois décrivant avec une grande précision les relations unissant les corps physiques), une telle approche est difficile à envisager dans le domaine des mesures subjectives comme la tristesse. En réalité, la mesure dans le domaine de la subjectivité s'entend comme « la représentation numérique d'une caractéristique ». Le sens commun nous invite en effet à constater que l'on peut être « plus » ou « moins » triste. On peut alors envisager de recourir au système symbolique numérique pour représenter cette intuition que nous avons de l'intensité de la tristesse d'un sujet. Le recours à ce système symbolique a ses avantages et ses inconvénients : il permet d'une part de bénéficier des formidables outils que les mathématiciens ont développés au cours des temps pour manipuler les nombres, il nous fait perdre d'autre part un peu de cette intimité que le langage naturel nous procure à partir de l'évocation du mot « tristesse ».

Quand on considère donc que « la tristesse, ça ne se mesure pas... », ce que l'on exprime simplement c'est que le mode de représentation numérique de la tristesse est un système symbolique différent du langage naturel que nous utilisons en permanence. Cette différence n'implique nullement que l'un des systèmes de représentation soit plus ou moins performant que l'autre. Ils ont chacun leur particularité, celle du nombre étant de permettre le recours plus immédiat à une méthodologie d'investigation de type scientifique. Cela ne nous indique pas pour autant si la mesure d'une caractéristique subjective peut être comparée à la mesure d'une caractéristique objective. Là aussi, alors que spontanément il pourrait sembler évident qu'une mesure de tristesse est de nature fondamentalement différente d'une mesure de cholestérolémie, l'examen un peu plus approfondi de cette question laisse en réalité perplexe. Voyons cela à partir de quelques exemples.

À propos des différences entre mesures subjectives et mesures objectives, certains pourraient remarquer qu'un attribut subjectif ne correspond à rien de concret, tout juste à un mot, qui plus est au sens généralement imprécis. Prenons la tristesse. Il est vrai que, curieusement, il est difficile d'en obtenir une définition¹, mais est-ce pour cela que le sens de ce mot est imprécis ? Bien au contraire. La tristesse est un sentiment élémentaire que tout être humain a

1. Bien souvent, des synonymes tels que « chagrin » ou « mélancolie » sont proposés, termes eux-mêmes définis à partir des mots « triste » ou « attristé ». Le dictionnaire « Le Robert » évite cet écueil en suggérant la définition : « état affectif pénible, calme et durable ». On remarquera cependant que le remords, par exemple, obéit à la même définition, or remords et tristesse sont des sentiments clairement distincts. Enfin, on propose parfois : « état affectif associé à la perte d'objet » ; très séduisante, cette approche a néanmoins le désavantage de recourir implicitement à une théorie (psychanalytique), ce qui ne va pas sans poser problème pour une définition.

déjà éprouvé. Il suffit, en fait, que je dise : « je suis triste... » pour que n'importe quelle personne sache exactement ce que je ressens. Pour le moins, on ne peut voir ici aucune imprécision. À l'opposé, le temps est une variable objective que chacun mesure quotidiennement. Mais comment définir le temps ? Philosophes ou physiciens ont en pratique bien des difficultés pour répondre à cette question.

À propos des ressemblances entre mesures subjectives et mesures objectives, on remarquera une évolution parallèle des paradigmes² de mesure dans les disciplines de la subjectivité et dans les sciences les plus « objectives », la physique notamment. En effet, si l'on prend l'exemple de la mesure d'une longueur, l'approche métrologique la plus élémentaire consiste à donner une impression de la grandeur de l'objet à mesurer : c'est « très grand », « assez grand »... Dans le domaine des mesures subjectives, on notera qu'il existe des instruments relevant du même paradigme ; on citera par exemple l'échelle d'impression clinique globale (CGI).

La mesure d'une longueur ne se limite bien sûr pas à des impressions, il existe un étalon, le mètre. Dans le domaine de la psychopathologie, il existe aussi des instruments que l'on peut associer au paradigme de l'étalonnage. Il en est ainsi des items de l'échelle de dépression de Hamilton. Chacun des items de cette échelle est constitué d'une succession de modalités de références parmi lesquelles il faut choisir celle qui semble être la plus proche du patient à évaluer. Par exemple, pour l'item de culpabilité, nous avons : « S'adresse des reproches à lui-même, a l'impression qu'il a causé des préjudices à des gens » ; « Idées de culpabilité ou ruminations sur des erreurs passées ou sur des actions condamnables » ; « La maladie actuelle est une punition. Idées délirantes de culpabilité »... Nous avons bien là une succession de jalons permettant d'étalonner la mesure, un peu à la manière du mètre étalon défini à la Révolution française.

Les physiciens ont cependant abandonné une définition du mètre fondée sur un simple jalon de référence. Ainsi, depuis 1983, le mètre est défini par « la longueur du trajet parcouru dans le vide par la lumière pendant une durée de 1/299 792 458 de seconde ». Cette nouvelle définition marque une rupture épistémologique majeure en métrologie : le processus de mesure fait maintenant partie intégrante d'une théorie au sens fort du terme. En effet, cette définition du mètre implique que la vitesse de la lumière dans le vide est une constante, ce qu'il est indispensable de vérifier expérimentalement. Cette vérification se fait en observant des interférences de franges lumineuses, ce qui fait appel à la théorie ondulatoire de la lumière...

De nombreux instruments de mesure utilisés en psychopathologie relèvent d'un tel paradigme, où l'instrument est étroitement lié à une théorie. C'est notamment le cas des instruments à plusieurs items, dépendant d'une théorie définitoire (celle qui a présidé au choix des items constituant l'instrument).

Pour conclure sur ce point, si nous venons de voir que les différences entre les mesures subjectives et les mesures objectives ne sont pas aussi évidentes que l'on pouvait le penser a priori, nous avons aussi constaté que les instruments de mesure utilisés le plus fréquemment en psychopathologie (les instruments constitués d'une liste d'items) dépendaient fortement d'un système théorique. Or d'une part ces systèmes théoriques ne sont pas toujours bien explicités, mais d'autre part ils peuvent ne pas être congruents avec les théories psychopathologiques sous-jacentes aux psychothérapies évaluées. Ainsi, il n'est pas rare de

2. Dans son sens courant, le terme paradigme désigne un « mot-type » qui est donné comme modèle pour une déclinaison, une conjugaison (le paradigme de la conjugaison des verbes du premier groupe est le verbe « chanter ») ; il est, en ce sens, synonyme de modèle ou d'exemple. En épistémologie, il a été utilisé par Kuhn pour désigner les « accomplissements passés pouvant servir d'exemples et remplacer les règles explicitées, en tant que bases de solution pour les énigmes qui subsistent dans la science normale » ; par exemple, l'expérience dite du labyrinthe en « T » est un paradigme de la théorie du conditionnement opérant de Skinner, l'auscultation est le paradigme de l'examen clinique tel qu'on le pratique depuis Laennec. C'est en ce sens que ce mot est utilisé ici.

voir des thérapies d'inspiration psychanalytique évaluées à l'aide d'instruments développés dans un cadre purement symptomatique ou selon une perspective comportementale. Les résultats de telles études sont donc nécessairement à interpréter avec prudence, même si certains pourraient alors argumenter que l'objectif n'est pas tant d'évaluer la psychothérapie en tant que telle, mais plutôt ses effets (son efficacité). Les instruments développés pour étudier les changements symptomatiques seraient alors adaptés aux examens de toute forme de psychothérapie, à la condition que le contexte de cette comparaison soit bien admis.

Reproductibilité statistique

Les études permettant d'évaluer les psychothérapies porteront sur des sujets singuliers dans leur trajectoire de vie et dans leur fonctionnement mental. Comment envisager alors des expériences reproductibles ? Cette question se pose, elle aussi, de la même façon dans les autres domaines de la recherche thérapeutique. En effet, si l'on évalue l'efficacité d'une antibiothérapie dans le traitement de la tuberculose pulmonaire, compte tenu du cycle du bacille de Koch, l'essai devra s'étaler sur une ou plusieurs années. Si un investigateur souhaite reproduire l'essai une fois les résultats publiés, il est possible que l'écologie du germe ait changé et que l'expérience ne soit plus tout à fait la même.

En recherche clinique, la notion de reproductibilité est donc affaiblie par rapport aux sciences expérimentales classiques comme la physique, la chimie ou la biologie.

Enfin, les résultats des thérapeutiques n'étant que rarement superposables d'un patient à l'autre, la reproductibilité des études cliniques ne peut être que statistique. Cela nous amène au dernier point.

Peut-on avoir recours à la statistique de la même façon dans le cadre de l'évaluation de thérapeutiques somatiques et dans le cadre de l'évaluation de psychothérapies ?

La réponse à cette question est de nouveau affirmative. En effet, dans un essai thérapeutique avec allocation aléatoire des traitements, la statistique ne fait que chiffrer la part possiblement attribuable au hasard dans la différence d'efficacité observée entre les traitements comparés. Plus précisément, chaque patient a, sans que l'on puisse généralement le savoir a priori, une aptitude particulière à répondre à un traitement donné. Il est possible que le tirage au sort attribue, par le simple fait du hasard, un plus grand nombre de fois un traitement particulier à des patients au potentiel de guérison le plus élevé. Les tests statistiques évaluent dans quelle mesure ce phénomène aléatoire pourrait aboutir à lui seul à la différence d'efficacité observée entre les traitements. Ce calcul se fait de la même façon que le traitement soit médicamenteux ou qu'il soit issu d'une psychothérapie.

La reproductibilité statistique des résultats issus des essais thérapeutiques en psychiatrie repose en grande partie sur l'adoption de critères diagnostiques comme ceux du DSM-III-R ou du DSM-IV. Ces critères obéissent cependant à une logique interne qui n'est pas sans conséquences sur les interprétations que l'on peut réaliser de ces résultats.

En effet, la démarche du DSM obéit à une recherche légitime de consensus syndromique en psychiatrie. Elle correspond à une induction scientifique, c'est-à-dire au repérage statistique de régularités d'occurrences des signes pathologiques, avec un degré de précision aussi grand que possible quant aux syndromes ainsi différenciés. Dans l'introduction au DSM-IV, il est pourtant mentionné que « le système de classification catégorielle a ses limites », qu'on ne postule pas que chaque trouble mental est une entité circonscrite, aux limites absolues l'isolant des autres troubles mentaux ou de l'absence de trouble mental. De plus, les individus qui partagent le même diagnostic sont susceptibles d'être hétérogènes, même en ce

qui concerne les critères de définition du diagnostic. Enfin, les cas limites sont difficiles à diagnostiquer ou ne peuvent l'être que de manière probabiliste.

Mais il existe alors un problème méthodologique incontournable : si l'induction est la méthode de vérification expérimentale par excellence, celle-ci ne peut, à elle seule, élaborer les hypothèses permettant de confronter d'éventuelles théories concurrentes pour rendre compte des faits observés. De fait, même aux États-Unis, le DSM-IV n'a pas fait disparaître les traités de psychiatrie, qui reposent sur un point de vue nosologique (l'étude des maladies) et non plus seulement syndromique (l'étude des troubles). Le passage de la constatation d'un syndrome à l'établissement d'un diagnostic ne peut relever d'une catégorisation inductive (inclusion du cas observé dans la classe de tous les cas présentant les mêmes caractères), mais d'une hypothèse abductive (recherche risquée de la classe nosologique non connue a priori qui rendra le mieux compte du cas singulier observé). Par exemple, un syndrome de discordance ou d'influence ne signifie pas automatiquement qu'il s'agit d'une maladie schizophrénique (et inversement, une personne souffrant d'une maladie schizophrénique peut ne pas présenter l'un ou l'autre de ces syndromes). Par ailleurs, il se peut que la « classe nosologique » ne soit pas la plus appropriée à l'interprétation des troubles comportementaux et mentaux regroupés en syndromes, d'où les hypothèses (abductives) psychodynamiques, éco-systémiques, humanistes et comportementales-cognitives.

En conclusion, nous venons de voir qu'il n'existe pas de frein conceptuel évident à l'évaluation scientifique de l'efficacité d'une psychothérapie. Il est en effet possible d'envisager de tester l'hypothèse d'efficacité de ces traitements dans le cadre d'expériences reproductibles réfutables. Des nuances doivent cependant être apportées à cette affirmation : tout d'abord la reproductibilité de ces études n'est pas totale, mais ce problème n'est pas propre au domaine des psychothérapies. Ensuite, ces études ne sont interprétables que si les mesures réalisées sont valides. Or, dans le cadre des psychothérapies ces mesures portent sur un phénotype subjectif. Si nous avons vu que de telles mesures ne sont pas conceptuellement très différentes des mesures habituelles, plus objectives, la question de la validité des instruments utilisés se pose, elle, de façon plus aiguë. En effet, les instruments de mesure les plus sophistiqués ne trouvent de sens véritable que quand ils sont insérés dans un champ théorique. Une situation idéale serait que l'instrument retenu dans l'étude soit compatible avec le support théorique utilisé par la ou les psychothérapie(s) évaluée(s). Il faudrait alors disposer d'une théorie du fonctionnement mental permettant de situer les différentes stratégies psychothérapeutiques les unes par rapport aux autres mais, malheureusement, s'il existe un grand nombre de théories parcellaires du fonctionnement mental, une théorie globale est bien loin d'exister aujourd'hui. Une alternative plus modeste (et discutable) est alors de recourir à un critère clinique.

2

Aspects méthodologiques

Paul (1969) a présenté d'une façon lapidaire l'ensemble des questions que soulève l'évaluation des effets thérapeutiques et des processus en jeu dans les psychothérapies : « Quel traitement, par quel thérapeute, est le plus efficace, pour quel sujet, à propos de quel problème, dans quelles circonstances, et comment ? » Ce qui peut donner lieu à sept grandes directions de recherche :

- définir avec précisions les traitements ;
- définir l'efficacité thérapeutique ;
- définir en termes opérationnels les problèmes psychologiques ;
- définir les sujets en termes opérationnels ;
- définir les comportements et les attitudes psychologiques qui sont psychothérapeutiques ;
- définir comment agissent les traitements : quels sont le ou les processus thérapeutiques ?
- définir dans quelles conditions de milieu agissent les traitements.

Cet abord pragmatique de la question pose des problèmes généraux d'ordre épistémologique et méthodologique.

Problèmes épistémologiques

Il s'agit d'examiner tout d'abord le problème de la pertinence d'une évaluation des phénomènes psychothérapeutiques, qui par nature sont subtils, individuels, souvent intersubjectifs et de ce fait se prêtent peu à une quantification sur le modèle des sciences expérimentales, qu'elles soient d'obédience médicale ou psychologique.

On peut se demander, en effet, si la psychothérapie et ses différents modèles ne relèvent pas des sciences idiographiques : autrement dit les sciences qui s'occupent de phénomènes uniques, non répétables et dont on ne peut donner qu'une description individualisée. Les sciences idiographiques sont, par exemple, l'histoire, la géographie et l'archéologie. Et l'on sait l'importance de la métaphore archéologique dans la description par Freud du processus psychanalytique. En revanche, les sciences nomothétiques s'efforcent de définir des lois, d'établir des relations répétables, de tester des hypothèses de plus en plus générales et de proposer des modèles du fonctionnement psychologique et des modèles d'intervention. Le cas individuel est supposé « porter en lui-même la forme entière de l'humaine condition » que l'on peut dégager par induction et repérer dans d'autres individus, au cours d'études statistiques (Cottraux, 1988).

La recherche en psychothérapie oscille en fait entre les pôles nomothétique et idiographique. Freud (1917) a passé une grande partie de sa vie à essayer d'établir des lois générales issues de l'expérience analytique, alors qu'il pensait pourtant que le travail analytique individuel ne pouvait être soumis à l'enquête statistique.

Protocoles de recherche

Les psychologues américains, à la suite de Claude Bernard (1865), ont développé des protocoles expérimentaux de cas individuels pour évaluer les effets et les processus des psychothérapies, en disant que ces protocoles étaient la pierre de touche de l'évaluation, pour ensuite développer des études contrôlées sur un nombre élevé de cas, afin de valider leur approche en fonction de critères méthodologiques universels.

La croyance de chaque psychothérapeute en la valeur du système auquel il adhère, croyance dont le fondement est rarement rationnel, représente un autre problème. Développer un programme de recherche, autrement dit formuler un système psychothérapeutique sous la forme d'un ensemble d'hypothèses testables, consiste à prendre le risque de voir ses croyances s'effondrer, et d'être obligé de modifier sa pratique.

Comme l'a montré Karl Popper (1963), la recherche scientifique est beaucoup moins la quête de la « Vérité » que la mise en place de dispositifs de contrôles susceptibles d'éliminer progressivement les erreurs. C'est « la falsifiabilité » (au sens anglais), autrement dit la « testabilité », qui implique une méthodologie susceptible de démontrer qu'une hypothèse ou une théorie est vraie ou fausse. Cette oscillation de l'individuel au général, et de la « croyance » à la « falsifiabilité », va se retrouver constamment dans les problèmes méthodologiques soulevés par l'évaluation des psychothérapies. Un protocole a pour but d'établir les règles logiques qui sont nécessaires à la sélection et à l'interprétation objective des données recueillies dans le dessein de répondre aux questions posées. La méthodologie consiste, de la manière la plus objective possible, à poser les questions, à élaborer un protocole qui permette de répondre à celles-ci, tout en sélectionnant des mesures fiables et adaptées à la question.

La formulation d'hypothèses de recherche consiste à se poser des questions et à se donner les moyens d'y répondre. Nombre de facteurs peuvent influencer le cours d'une psychothérapie : la nature et le degré du trouble, des événements de vie, l'effet placebo, la méthode thérapeutique utilisée, la relation thérapeutique, une bonne ou mauvaise alliance thérapeutique ou des changements biologiques. Dans les travaux les plus expérimentaux effectués chez l'animal, il existe un effet expérimentateur : l'effet Rosenthal (Rosenthal, 1966 ; Rosenthal et Rosnow, 1975). De plus, le choix des variables étudiées est intrinsèquement lié aux hypothèses et aux résultats obtenus.

Notions d'efficacité et d'utilité clinique

Tout projet de recherche est donc lié à une procédure impliquant à la fois un groupe contrôle et des mesures ce qui le rend réducteur, par nécessité. De nombreuses variables sont prises en compte dans les projets : en particulier les mesures des symptômes et syndromes, de la personnalité, de l'ajustement, du handicap, de la qualité de vie, des processus thérapeutiques, ou de variables indépendantes ayant valeur de prédiction. Les méta-analyses des essais contrôlés se centrent souvent sur la réduction des syndromes et des symptômes les plus représentatifs des troubles étudiés. Les méta-analyses les plus récentes et les mieux faites évaluent les différents aspects du changement et certaines se penchent sur les processus. Cette manière multimodale de considérer le problème du changement est rarement prise en compte. Beaucoup de revues institutionnelles récentes envisagent seulement le problème de l'efficacité, dans la mesure où la définition d'un rapport coût/efficacité favorable est devenue le point nodal dans les politiques de santé. Pourtant, la notion même d'efficacité doit être nuancée.

De nombreux auteurs ont insisté sur le fait qu'il ne suffit pas de modifier les symptômes et les syndromes pour obtenir un résultat satisfaisant pour le patient. Déjà, dans un rapport de

l'OMS (Sartorius et coll., 1993), un certain nombre de critères ont été proposés pour évaluer les résultats des traitements psychiatriques :

- utilité pratique (*effectiveness*) : réduction des symptômes ; amélioration de la capacité à tenir des rôles sociaux et professionnels ; qualité de vie des patients et de leur famille ;
- sécurité ;
- effets secondaires ;
- problèmes éthiques ;
- rapport coût/efficacité comparé à d'autres traitements ;
- possibilité d'être appliqué dans des situations sociales variées et par différentes catégories de professionnels de santé mentale ;
- risque de mésusage (par exemple dépendance à un médicament).

Aujourd'hui, il apparaît de plus en plus qu'à la notion d'efficacité (*efficacy*), qui peut être mise en évidence par les études contrôlées randomisées sur des populations souvent très sévèrement sélectionnées et peu représentatives de la pratique, doit se substituer celle d'utilité clinique (*effectiveness*), c'est-à-dire de l'application pratique du savoir-faire psychothérapique (Clarkin et coll., 1996 ; Gabbard, 1997). Des méta-analyses récentes ont essayé de répondre à cette question.

Shadish et coll. (1997) ont reclassé des études incluses dans 15 méta-analyses des psychothérapies en fonction de leur représentativité des conditions cliniques habituelles et abouti à la conclusion que les résultats sont les mêmes sur un terrain clinique proche des conditions habituelles que dans des projets de recherche portant sur des populations de sujets très sélectionnés.

Méthodes d'évaluation

L'évaluation des psychothérapies a suivi l'évolution de la recherche clinique qui est passée progressivement des études de cas individuels aux protocoles randomisés.

Études de cas individuels

L'histoire de la psychothérapie a débuté par des études de cas individuels rapportées par Charcot, Janet, Freud et leurs successeurs immédiats. Les études de cas ont une valeur heuristique irremplaçable. Les psychothérapies modernes doivent toutes quelque chose à « L'automatisme psychologique » de Pierre Janet (1889) et aux « Études sur l'hystérie » de Freud et Breuer (1895). Tous les travaux en psychothérapie, quel qu'en soit le type, commencent par des études de cas individuels, puis des séries, pour se poursuivre par des études contrôlées. Les statistiques sur des séries de cas peuvent évaluer les résultats d'un thérapeute ou d'une institution. Mais ces statistiques globales, si elles étudient certains aspects du processus thérapeutique au cours de recherches corrélationnelles, ne permettent pas de conclure à l'efficacité d'un traitement par rapport à l'évolution spontanée ou par rapport à un autre traitement. Les protocoles de cas individuels cherchent à quantifier une démarche fondée sur des cas uniques.

Protocoles de cas individuels dans la recherche en psychothérapie

Ces protocoles, qui allient démarche quantitative et démarche qualitative, sont issus d'un modèle physiologique destiné à étudier le fonctionnement d'un organisme mis dans différents environnements ou soumis à différent(e)s interventions ou traitements. La démarche quantitative est assurée par des mesures répétées, la démarche qualitative par le fait que le sujet est son propre témoin. Claude Bernard (1865) est le premier auteur à avoir mis l'accent sur cette méthode.

Les protocoles de cas individuels ont été initiés par Shapiro en Angleterre dans les années 1960, puis décrits dans un ouvrage fondamental par Hersen et Barlow (1976) aux États-Unis. Les travaux de Kazdin (1982), Barlow et coll. (1984) et Bellack et Hersen (1984) ont suivi. En langue française, on peut citer l'ouvrage de Ladouceur et Bégin (1980), et des chapitres dans les ouvrages de Cottraux et coll. (1985), et Bouvard et Cottraux (1996).

Les protocoles de cas individuels représentent une manière simple de promouvoir la recherche clinique. Ils consistent à prendre le sujet comme son propre témoin et à l'évaluer au cours de phases tantôt d'intervention tantôt de non-intervention. Ils représentent un moyen rapide de tester une hypothèse sans engager les frais d'une étude contrôlée. Ils ont aussi été utilisés dans la recherche animale sur le conditionnement opérant. À condition d'avoir un grand nombre de points de mesure, une analyse statistique comparative des différentes phases est possible (Hersen et Barlow, 1976).

Protocole A-B quasi expérimental

Il comporte au moins trois phases :

- une ligne de base (ou de référence) qui apprécie le comportement avant tout traitement (phase A). Sa durée varie suivant la nature du problème étudié, ou le caractère éthique de la non-intervention. Une phase de liste d'attente peut être mise en place pour demander au patient de faire des mesures répétées d'un comportement cible ;
- une phase de traitement où est mise en place la procédure thérapeutique (phase B) ;
- une phase de suivi où est évalué le maintien des changements comportementaux obtenus.

La comparaison par inspection des courbes représentant la variable choisie permet d'émettre l'hypothèse que c'est l'intervention thérapeutique qui a été responsable des résultats obtenus.

Pour cette analyse de tendance, il faut au moins trois points par phase pour analyser le protocole (car par trois points passe une courbe et par deux points seulement, une droite).

Protocole expérimental A-B-A-B

Le protocole A-B a été critiqué, il est simplement quasi-expérimental et ne permet pas d'établir un lien causal entre le changement et l'intervention.

Le protocole de retrait A-B-A-B est l'un des protocoles les plus utilisés ; il est considéré comme un véritable protocole expérimental. Il fait succéder par exemple une phase de liste d'attente, une phase d'intervention, suivie d'une phase de non-intervention, puis d'une nouvelle phase d'intervention (protocole A-B-A-B dans lequel A = intervention, B = non-intervention). Des mesures répétées des variables à l'étude permettent de voir s'il y a chez le sujet des changements importants durant les phases d'intervention, ou s'il s'améliore spontanément durant les phases sans intervention. L'hypothèse qui sous-tend ce protocole est que si l'amélioration observée est liée au traitement, le retrait plus ou moins prolongé de celui-ci s'accompagnera d'un arrêt des progrès en cours, ou même d'un retour à l'état initial, ou encore d'une aggravation. Inversement, les progrès reprendront à la réintroduction du traitement. Lorsqu'un traitement (B) mis en place ne donne pas les résultats escomptés, il est

possible d'adjoindre un autre traitement (C) et d'évaluer ainsi l'action combinée des deux traitements (BC) en réalisant un protocole A-B-BC-B ou, au contraire, de supprimer l'un des deux ingrédients d'un « package » thérapeutique pour voir s'il participe effectivement aux bons résultats (protocole A-B-C-A-B-C par exemple).

Protocole B-A-B

Dans ce protocole qui est utile lorsqu'il est nécessaire d'effectuer une intervention, pour des raisons pratiques et éthiques, on intervient, puis on retire l'intervention, pour la réinstaurer ensuite, si le comportement ou problème réapparaît de manière intense.

Protocole de lignes de base multiples à travers les comportements

La méthode utilisée ne fait pas appel à la présentation suivie du retrait du traitement. Elle consiste, après avoir effectué la mesure simultanée de plusieurs comportements lors de lignes de base (A), à introduire, pour chacun des comportements problèmes que l'on mesure, l'intervention thérapeutique (B) à des moments différents. Les changements comportementaux devront apparaître après l'intervention, pour que l'on puisse affirmer qu'il y a eu un effet spécifique du traitement sur chacun des comportements.

Protocole de ligne de base multiple à travers les sujets

Il consiste à effectuer le relevé d'un comportement donné chez plusieurs sujets et à introduire pour chacun d'eux le traitement à des temps différents.

Statistiques des cas à n = 1

Au-delà de la simple analyse de tendance par inspection, trois types de statistiques ont été proposés :

- Comparaison des phases de ligne de base et d'intervention. Tous les logiciels de statistiques actuels peuvent effectuer ces calculs simples. Par exemple A1 + A2 sont comparées à B1 + B2 avec un test t apparié. Il faut au moins dix points de mesure par phase. Quand il s'agit d'un protocole A-B-C, il est nécessaire d'utiliser une ANOVA. En cas de protocole à plusieurs sujets, on peut faire une ANOVA évaluant l'effet sujet et l'effet phase et leurs interactions. Ces statistiques sont discutées car elles ne tiennent pas compte de la corrélation des mesures intra-sujets et elles ne reflètent qu'un effet moyen et ne prennent pas en compte les tendances (la pente).
- Analyses de séries temporelles (*time series analysis*). Elles permettent de comparer les moyennes et les pentes de chaque phase en tenant compte des corrélations. Elles nécessitent un logiciel spécialisé. Il faut au moins dix points de mesure par phase.
- Rn de Revusky : protocoles à lignes de base multiples à travers les sujets. C'est une forme de statistique non paramétrique qui est surtout utile dans les protocoles à travers les sujets (un effectif d'au moins quatre est requis) avec introduction de la phase B de manière aléatoire pour chaque sujet. On évalue les changements de rang des sujets à l'introduction de B. Une somme des rangs est calculée et rapportée à une table de probabilité (Hersen et Barlow, 1976). Le Rn peut être fait « à la main ».

Passage du cas individuel au groupe

Ces protocoles peuvent être pratiqués sur des groupes de sujets en prenant chaque sujet comme son propre témoin et en additionnant les résultats. Il suffit par exemple de mesurer une ligne de base de non-intervention (A) chez tous les sujets et d'effectuer le traitement de manière identique (B). On se trouve alors devant un protocole à mesures répétées A-B. Une

ANOVA à mesures répétées permettra d'évaluer l'évolution du groupe à travers les phases. Un tel protocole a été utilisé de cette façon par Clark et coll. (1985) pour montrer l'efficacité de la thérapie cognitive dans un groupe de 18 patients avec un trouble panique. Ce type d'étude permet en partie de concilier les exigences du cas individuel et de l'étude contrôlée et également de résoudre le problème de la comparabilité des sujets, souvent difficile à établir.

Limites des protocoles de cas individuels

La limite des protocoles de cas individuels reste la généralisation de leurs conclusions à l'ensemble d'une population pathologique. Ils représentent avant tout un moyen simple et peu coûteux pour tester des hypothèses thérapeutiques et s'initier à la recherche clinique en apprenant à sélectionner des mesures, à observer et à enregistrer sur des courbes des événements cliniques évalués jusque-là de façon intuitive. Pour le chercheur confirmé c'est un prélude nécessaire aux études contrôlées, qui permet de rester au niveau de l'individu et de générer des hypothèses qui pourraient ne pas apparaître dans les études de groupes parallèles, en double aveugle et randomisées, effectuées trop rapidement.

Protocoles randomisés dans la recherche en psychothérapie

Le noyau dur de la recherche sur le comportement humain est représenté par les études comparatives, contrôlées (Meyers et Grossen 1974). Cependant, constituer un groupe contrôle pour évaluer une méthode de psychothérapie ne va pas sans problèmes. Il est pratiquement impossible de comparer après randomisation et en double aveugle une psychothérapie active à un « placebo » inerte de psychothérapie sur le modèle des études pharmacologiques. En effet, les phénomènes relationnels et situationnels ainsi que les attentes des thérapeutes et des patients sont des ingrédients actifs de tout système psychothérapeutique.

Placebo en psychothérapie

Plusieurs solutions ont été proposées pour résoudre le problème du placebo en psychothérapie. Le groupe « attention placebo » avec un contact minimum avec un thérapeute qui n'utilise pas les éléments supposés actifs de la thérapie que l'on veut tester permet d'éliminer les effets simples de prise en charge. La liste d'attente durant plusieurs mois avec simple contact téléphonique pose des problèmes éthiques et risque d'aboutir à des sorties d'essai vers une autre thérapie. Le contrôle par une pseudo-thérapie ou une anti-thérapie génère aussi des problèmes éthiques et pratiques. La comparaison d'une thérapie de référence à la thérapie testée entraîne des problèmes d'interprétation, dans la mesure où la nouvelle thérapie risque d'avoir les bénéfices de la nouveauté et une prime au changement (« *band-wagon effect* » des auteurs américains). De même, la comparaison d'une chimiothérapie à une psychothérapie risque d'être biaisée en faveur de la psychothérapie si les patients ont presque tous eu des chimiothérapies inefficaces, sinon ils ne viendraient pas chercher un autre traitement.

Le double aveugle est impossible, sauf en cas de comparaison d'une thérapie à elle-même, associée à un médicament actif ou à un placebo. C'est dire l'importance d'une évaluation indépendante et aveugle aux hypothèses testées. Il n'y a pas de solution idéale, sinon d'évaluer en début de traitement la croyance des patients et des thérapeutes dans le traitement qui a été tiré au sort et d'étudier la corrélation de ces mesures avec les résultats. Le placebo de psychothérapie doit avoir des caractéristiques qui le rendent aussi vraisemblable qu'une thérapie véritable : le placebo doit être crédible.

Variables du thérapeute

Un certain nombre de facteurs liés à l'attitude et au comportement du thérapeute vis-à-vis du patient ont été longtemps considérés comme thérapeutiques. Ainsi, la force de persuasion, la possibilité de créer une atmosphère amicale, l'action anti-démoralisatrice, la chaleur, l'empathie, l'authenticité des sentiments et la considération positive inconditionnelle du patient ont été invoquées. Il faut leur ajouter la compétence, le statut socioprofessionnel, la crédibilité, le décor et la célébrité (Parloff et coll., in Garfield et Bergin, 1978). Ces facteurs, souvent invoqués, ont été peu étudiés de manière empirique. On tend à leur préférer, actuellement, l'étude de l'alliance thérapeutique qui renvoie à l'interaction de deux personnes et non à des qualités personnelles.

Alliance thérapeutique

Le concept d'alliance thérapeutique est dû à Freud (1913) : « compréhension sympathique, affection et amitié sont les véhicules de la psychanalyse ». En thérapie analytique, l'alliance thérapeutique réfère aux aspects les plus rationnels de la relation thérapeutique. Elle est en contraste avec les éléments plus transférentiels et donc irrationnels. L'accent a été mis sur ce processus particulier par Luborsky et coll. (1985) qui ont réalisé une étude comparant la thérapie cognitive associée au conseil, la thérapie analytique associée au conseil et le conseil seul chez des toxicomanes sevrés. La thérapie cognitive et la thérapie analytique étaient égales entre elles et supérieures au conseil seul. Les qualités du thérapeute n'étaient pas corrélées aux résultats. Les prédictors de réussite en thérapie cognitive ou en thérapie analytique étaient l'alliance thérapeutique et la pureté technique : c'est-à-dire suivre consciencieusement le manuel.

Relation de collaboration empirique

Alford et Beck (1997) ont défini la relation thérapeutique en thérapie cognitive comme une relation de collaboration empirique qui serait comparable à celle de deux savants travaillant ensemble sur un problème. Elle sert de fondement à l'apprentissage qui est lié aussi à la relation thérapeutique et aux changements cognitifs du sujet. Cependant, le changement en thérapie est dû aux capacités du sujet et non uniquement à la relation et doit se généraliser dans d'autres contextes que la thérapie. Les auteurs conseillent donc de laisser la responsabilité du changement au sujet sans exagérer l'importance du rôle du thérapeute. La relation thérapeutique ainsi définie est une condition nécessaire mais non suffisante.

Dans une perspective voisine, Cottraux et coll. (1995) ont essayé d'évaluer la relation thérapeutique avec une échelle faite de douze paires d'adjectifs bipolaires qui permettaient d'évaluer sur six points : le thérapeute vu par le patient et le patient vu par le thérapeute. Cette échelle a été utilisée dans un essai randomisé comparant la thérapie cognitivo-comportementale soit avec buspirone, soit avec placebo, en double aveugle dans le trouble panique avec agoraphobie. Les résultats ont montré un effet supérieur de la buspirone sur l'anxiété généralisée mais non sur le trouble panique et l'agoraphobie. Les thérapeutes ont été évalués plus positivement par les patients que les patients par les thérapeutes. Les prescripteurs ont été moins bien évalués que les thérapeutes par les patients. Mais il n'y a pas eu d'effet de la relation thérapeutique, ainsi mesurée, sur la réponse thérapeutique lorsque l'on comparait les répondeurs aux non-répondeurs dans les deux groupes.

Méthodes de mesure

Il existe de nombreuses échelles d'évaluation des symptômes, des comportements et des processus psychothérapeutiques qui actuellement ont reçu une validation et rendent possible l'étude de problèmes psychopathologiques variés (Guy, 1976 ; Cottraux et coll., 1985 ; Guelfi, 1993 ; Cottraux et Blackburn, 1995 ; Cottraux et coll., 1995 ; Bouvard et Cottraux, 1996 ; Yao et coll., 1996 et 1998 ; Bouvard, 1999 ; APA, 2000 ; Ventureyra et coll., 2002). Il n'est pas possible

de détailler ici toutes les échelles : nous renvoyons donc le lecteur aux références citées ci-dessus qui regroupent les instruments disponibles. Il faut, de plus, souligner que cette liste est loin d'être exhaustive.

Ces échelles doivent être complétées par des questionnaires de personnalité ou des mesures adaptées, en fonction des hypothèses testées.

Les tests comportementaux *in vivo* permettent une mesure directe des performances d'un sujet et peuvent différer notablement des échelles d'évaluation.

Le tableau 2.I présente les principales échelles d'évaluation et questionnaires disponibles en français.

Tableau 2.I : Principales échelles d'évaluation et questionnaires disponibles en français et date de la publication : NB ces échelles sont disponibles en français dans les ouvrages suivants : Cottraux et coll., 1985 ; Guelfi, 1993 ; Cottraux et Blackburn, 1995 ; Cottraux et coll., 1995 ; Bouvard et Cottraux, 1996 ; Bouvard, 1999.

Dépression

Échelle de dépression d'Hamilton, HAM-D ou HRSD (1967) ; Inventaire de dépression de Beck, BDI (1979) ; MADRS, échelle d'évaluation de la dépression de Montgomery et Ashberg (1979) ; Échelle de désespoir (Beck, 1974) ; Échelle d'attitudes dysfonctionnements (Beck et Weissman, 1980) ; Questionnaire des pensées automatiques (Hollon et Kendall, 1985) ; Échelle de sociotropie-autonomie, SAS (Beck, 1983) ; Échelle de ralentissement de Widlöcher (1981) ; Diagramme HARD (Rufin et Ferreri, 1984)

Anxiété globale et anxiété généralisée

Échelle d'appréciation de l'anxiété d'Hamilton, HAM-A (1969) ; Échelle HAD, anxiété et dépression en milieu hospitalier (Zigmond, 1983) ; Échelle de Covi (1984) ; Échelle d'anxiété de Beck, BAI (1988) ; Questionnaire sur les inquiétudes du Penn State (1990)

Phobies

Échelle des peurs III, FSS III (Wolpe et Lang, 1967) ; Questionnaire des peurs (Marks et Matthews, 1969)

Attaque de panique et agoraphobie

Phobie, panique, anxiété généralisée, PPAG (Cottraux, 1985) ; Questionnaire des cognitions agoraphobiques (Chambless, 1984) ; Échelle de sévérité du trouble panique, PDSS (Shear, 1992)

Obsessions et compulsions

Liste des obsessions-compulsions, CAC (Marks, 1977) ; Test comportemental d'évitement (Marks, 1977) ; Liste de pensées obsédantes, LPO (Bouvard, 1980) ; Échelle des quatre rituels cibles (Marks, 1980) ; Échelle de Yale-Brown, Y-BOCS (Goodman, 1989) ; Questionnaire des pensées intrusives et de leurs interprétations, QPII (Yao, 1996) ; Questionnaire sur les croyances obsessionnelles, QCO-87 (*Obsessive compulsive cognitions working group*, 2003) ; Inventaire des interprétations des pensées intrusives (*Obsessive compulsive cognitions working group*, 2003)

Phobie sociale

Échelle d'affirmation de soi de Rathus (1973) ; Échelle d'anxiété sociale de Liebowitz (1987) ; Test des pensées en interaction sociale, TAPIS (Glass, 1982) ; Échelle d'infériorité, EDI (Yao, 1998)

Stress post-traumatique

PCL-S (Wheathers, 1993)

États psychotiques

BPRS, échelle abrégée d'évaluation psychiatrique (Overall, 1962) ; CPRS, échelle d'évaluation psychopathologique générale (Ashberg, 1978) ; Échelle NOSIE, *Nurse observation scale for inpatient evaluation* (Honigfeld, 1969) ; SANS et SAPS, Symptômes positifs et négatifs dans la schizophrénie (Andreasen, 1982)

Handicap

Échelle de handicap de Sheehan

Qualité de vie

Échelle de qualité de vie de Marks (1993, *In* : Cottraux, 1995) ; *Functional status questionnaire*, FSQ (Jette, 1986 ; Martin, 1993)

Personnalité

Inventaire multiphasique de personnalité du Minnesota, version abrégée – minimult – (Hathaway et Mc Kinley, 1991) ; *Toronto alexithymia scale*, TA 20 (Bagby, 1994) ; PAS, *Personality assessment scheme* (Tyrer, 1992) ; SCIDII-PQ (First et coll., 1997) ; PDQ-4 (Hyler, 1994) ; *International personality disorders examination*, IPDE (Loranger, 1991) ; Inventaire de personnalité d'Eysenck, EPI (1968) ; Entrevue structurée diagnostiqueur pour la personnalité

borderline, DIB-R (Zanarini et Gunderson, 1989) ; Inventaire de personnalité de Cloninger, TCI (1993) ; Questionnaire des schémas, SQI et SQII (Young, 1990 et 1994) ; Échelle d'impulsivité de Barrat (1993) ; NEO-PIR, les cinq grands facteurs (Costa et Mc Rae, 1990)

Relation thérapeutique

Questionnaire d'alliance thérapeutique, Haq II (Luborsky et coll., 1996 ; traduction Cottraux, non publiée) ; Échelle de relation thérapeutique, ERT (Cottraux, 1995)

Attentes thérapeutiques

Échelles d'attentes thérapeutiques : thérapeute/patient (Cottraux, 1995)

Instruments d'évaluation du fonctionnement global et de la détresse symptomatique

Mini international neuropsychiatric interview (Sheehan et coll., 1992) ; Santé-maladie psychologique, HSRS (Luborsky et coll., 1962-1991) ; *Global assessment scale*, GAS (Spitzer, 1976) ; Échelle d'évaluation globale du fonctionnement (axe V du DSM-IV) ; Adaptation sociale, SAS (Weissman, 1976) ; Questionnaire de santé (Goldberg, 1978) ; Liste de symptômes, HSCL-90-R (Derogatis, 1973) ; CGI, Impression clinique globale (Guy, ECDEU, 1976)

La vidéo ou les enregistrements au magnétophone permettent d'évaluer aussi bien les patients que les thérapeutes. Malgré les résistances qu'ils suscitent, ils servent à mettre en évidence aussi bien les processus thérapeutiques que les résultats. En différé, deux juges indépendants et « aveugles » peuvent coter les progrès des patients, le contenu de l'interaction psychothérapique et l'adhésion des thérapeutes aux hypothèses et aux techniques thérapeutiques.

Sur le plan des résultats et de leur analyse, deux points sont essentiels et spécifiques à l'évaluation des psychothérapies : une évaluation bien conduite doit avoir des critères et des mesures multiples, de façon à ne pas trop limiter la portée des conclusions ; elle doit aussi analyser en détail les éléments du processus thérapeutique.

Critères de jugement

À côté de la variation des scores d'échelles continues, l'évaluation doit utiliser des critères généraux, discontinus de bons résultats ou « critères de jugement ». Il peut exister un critère dichotomique unique de jugement (succès, échec) ou un critère principal et des critères secondaires.

En effet, des changements statistiquement significatifs d'une échelle sur un groupe peuvent ne refléter que des résultats cliniques médiocres dont la moyenne suffit à rendre significatifs les tests statistiques, si la puissance statistique est élevée du fait du nombre d'inclusions. Inversement, l'absence de changement moyen des scores d'échelles peut, plus rarement, s'accompagner de changements cliniques intéressants pour certains patients ou un sous-groupe de patients.

Donner la « taille d'effet », qui représente la magnitude de l'effet obtenu pour le sujet moyen de l'étude selon qu'il a le traitement ou son comparateur (placebo ou autre traitement), représente un complément nécessaire aux analyses statistiques classiques. On trouvera plus loin, au paragraphe sur la méta-analyse, la définition de la taille d'effet.

Sélection des sujets

Quelles que soient les méthodes utilisées pour le recrutement des patients : annonces, médias, recrutement dans une consultation hospitalière, étudiants ou volontaires payés comme aux États-Unis, elles sélectionnent un échantillon. L'établissement de critères comme ceux du DSM-III puis du DSM-IV (APA, 1994) et leur compatibilité avec les critères de l'ICD-10 (CIM-10, OMS, 1992) ont été vivement critiqués dans notre pays. Pourtant, ils représentent la meilleure approximation provisoire qui permette la sélection relativement homogène de sujets pour une étude.

Critères de qualité des essais thérapeutiques : les règles d'or

Dès 1975, Luborsky et coll. avaient défini un ensemble de critères permettant de juger de l'adéquation de conception d'une étude comparative (tableau 2.II).

Tableau 2.II : Critères d'évaluation d'études comparatives (d'après Luborsky et coll., 1975)

Les patients de chaque groupe ont été répartis de façon aléatoire (ou stratifiée sur les variables pronostiques)
 Ce sont de vrais patients qui sont utilisés, et non pas des acteurs ou des étudiants volontaires
 Les thérapeutes de chaque groupe sont compétents de façon équivalente
 Les traitements sont évalués de façon équivalente par les patients et les thérapeutes de chaque groupe
 Les mesures de résultats prennent en compte les objectifs cibles du traitement
 Les résultats du traitement sont évalués par des mesures indépendantes
 Une information a été obtenue à propos de l'utilisation concurrente d'autres traitements que ceux qui sont évalués, qu'ils soient formels ou informels
 Des échantillons de chacun des traitements comparés sont évalués de façon indépendante pour savoir à quel point les thérapeutes adhèrent à la forme de traitement décrite dans le manuel
 Chacun des traitements comparés a été administré à des niveaux égaux en termes de longueur ou de fréquence
 Chaque traitement a été administré à un niveau raisonnable et approprié à la forme de traitement
 La taille de l'échantillon est raisonnable

Une échelle d'évaluation, construite par Foa et Meadows (1997), comprenant 7 items (tableau 2.III), permet de noter les essais thérapeutiques sur 3 points. Plus la note est haute plus l'essai est de qualité. Maxfield et Hyer (2002) ont proposé une version plus élaborée de cette échelle en ajoutant 3 items supplémentaires (les trois derniers du tableau).

Tableau 2.III : Évaluation des essais thérapeutiques contrôlés (d'après Foa et Meadows, 1997 ; révisée par Maxfield et Hyer, 2002)

Critères	Notation
Symptômes clairement définis	0/0,5/1
Mesures validées	0/0,5/1
Évaluateur indépendant et aveugle	0/0,5/1
Évaluateur entraîné et fiable	0/0,5/1
Traitement présenté dans un manuel	0/0,5/1
Randomisation	0/0,5/1
Fidélité au traitement	0/0,5/1
Pas d'autre traitement concomitant	0/0,5/1
Mesures et entretiens d'évaluation multimodaux	0/0,5/1
Durée du traitement optimale	0/0,5/1

Méta-analyse des essais contrôlés sur les effets des psychothérapies

Le principe de la méta-analyse consiste à considérer que l'ensemble des études sur un sujet donné représente une quantité d'informations en liaison avec l'objectif de la recherche, chaque étude y apportant sa contribution. Il existe deux parties dans chaque étude, l'une commune à toutes les études qui servira à tester l'hypothèse, l'autre spécifique à chaque étude. Cette dernière est liée à l'environnement et aux sujets. En outre, on prend pour hypothèse que l'ensemble des études est un échantillon de toutes les études possibles sur le thème donné.

La méta-analyse a été inventée pour étudier les effets des psychothérapies (Glass, 1976), même si sa pratique s'est ensuite diffusée à l'ensemble de la médecine. Smith et Glass (1977) ont développé une méthode d'analyse de la littérature concernant les études contrôlées sur les effets des psychothérapies. Elle consiste à regrouper les études, coder les résultats, et calculer l'ampleur ou la « taille de l'effet thérapeutique » (*effect size*). Celle-ci correspond, pour un critère donné que l'on étudie en fin de traitement, à la différence de distribution entre le groupe à l'étude et son comparateur (groupe contrôle ou autre traitement).

Le critère est en général le score d'une échelle d'évaluation. La taille d'effet correspond à la valeur moyenne du groupe traité moins la valeur moyenne du groupe contrôle, divisée par l'écart-type du groupe contrôle. Toutes les valeurs qui entrent dans l'équation sont celles de la fin du traitement. La taille d'effet reflète la différence entre les deux groupes en fin de traitement et le gain éventuel du groupe traité par rapport au groupe contrôle. Entre 0,20 et 0,50 une taille d'effet est petite, elle est moyenne entre 0,50 et 0,80, et grande au-delà de 0,80.

Si l'on admet que la distribution est normale dans les deux groupes, on peut estimer le percentile du groupe contrôle auquel se situe le sujet moyen du groupe traité (figure 2.1). Pour une taille d'effet de 0,00, il se situera dans la moyenne du groupe contrôle. Pour une taille d'effet de 1,00, le sujet traité sera plus amélioré que 84 % du groupe contrôle et pour une taille d'effet de 2,00, il sera plus amélioré que 97 % du groupe contrôle. Le tableau 2.IV donne les conversions des tailles d'effet en percentiles (Roth et Fonagy, 1996).

En cas de variable binaire (succès/échec), on utilise l'*odds ratio* (OR) qui permet de comparer les groupes dans les limites d'un intervalle de confiance. La méga-analyse est une variante de la méta-analyse qui prend en compte tous les sujets de toutes les études incluses au lieu de se contenter d'un calcul fondé sur le résultat de chaque étude (DeRubeis et coll., 1999).

Le but de la méta-analyse est essentiellement de résoudre les problèmes des résultats discordants en fournissant une information plus précise sur l'intensité des effets. Elle devrait aussi être utile pour l'identification des répondeurs. En partant de ces prémisses, on peut définir les critères de qualité d'une méta-analyse (Jenicek, 1987 ; Boissel et coll., 1989 ; Cucherat et Cialdella, 1996).

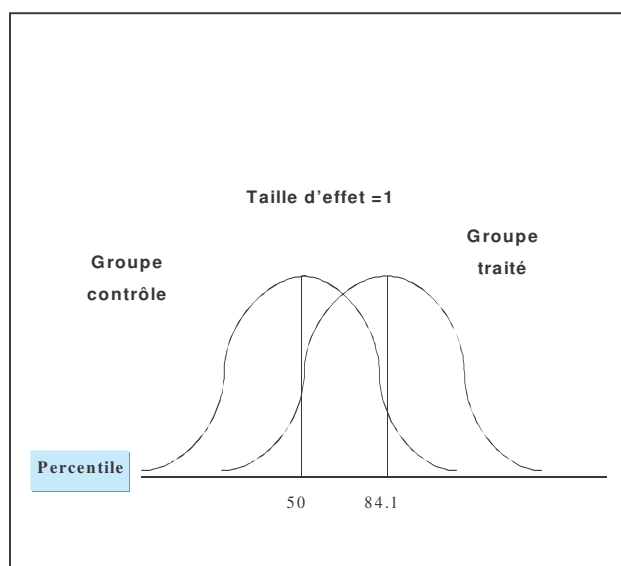


Figure 2.1 : Distribution des scores et taille d'effet

La courbe permet de visualiser les différences pour une taille d'effet égale à un.

Tableau 2.IV : Conversion des tailles d'effet en percentiles (d'après Roth et Fonagy, 1996)

Taille d'effet	Percentile
0,0	50
0,1	54
0,2	57,9
0,3	61,8
0,5	69,1
0,7	75,8
1,0	84,1
1,5	93,3
2,0	97,7

Inclusion de tous les essais de qualité

Ceci implique une recherche extensive des études publiées dans l'ensemble des bases de données et également des études non publiées auprès des auteurs ou des structures de recherche et de soins. Toutes les études du domaine concerné doivent être considérées. Les critères d'inclusion (et d'exclusion) dans une méta-analyse doivent être clairement énoncés. Les études rejetées doivent être indexées à la bibliographie et les raisons de l'exclusion justifiées dans le texte de l'article. La sélection des études doit rester indépendante de leurs résultats, ce qui en principe évite les pseudo-méta-analyses sélectives qui sont des plaidoyers « *pro domo* ».

Idéalement, deux évaluateurs indépendants doivent sélectionner les études et ne conserver que les essais contrôlés randomisés sans biais importants. La qualité des essais doit donc être prise en compte, ce qui ne va pas sans discussion et compromis entre totalité et qualité (Greenwald et Russel, 1991). L'utilisation de règles d'or définissant la qualité des essais thérapeutiques comme celles de Maxfield et Hyer (2002) ou celles appliquées par l'organisation Cochrane, qui sont voisines, peut réduire le bruit de fond créé par des études de basse qualité méthodologique.

Critères de jugement

Il peut s'agir d'un critère binaire succès/échec par exemple : dans ce cas, on rapportera l'*odds ratio* (OR) ou le risque relatif (RR) du groupe traité par rapport au groupe de comparaison. Dans le cas d'un critère quantitatif, ce qui est le plus souvent le cas en psychiatrie, on rapportera la taille d'effet sur le critère de jugement principal : par exemple une échelle d'anxiété ou de dépression. La conclusion tient compte de l'inflation du risque alpha, c'est-à-dire des faux positifs. On donne un seuil de sensibilité plus élevé que $p < 0,05$: $p < 0,01$ est le seuil habituel des méta-analyses.

Utilisation de méthodes statistiques adéquates

Plusieurs méthodes ont été proposées. Au début des méta-analyses des psychothérapies, on utilisait le « d » de Glass calculé selon la formule suivante qui utilise les valeurs obtenues en fin de traitement.

$$d \text{ (taille d'effet)} = \frac{\text{moyenne du groupe traité} - \text{moyenne du groupe de comparaison}}{\text{écart type du groupe de comparaison}}$$

Le groupe traité est le groupe chez lequel on veut étudier l'ampleur de l'effet, le groupe de comparaison peut être un groupe contrôle (placebo, attention placebo, liste d'attente) ou un groupe recevant un traitement déjà validé qui sert de référence. Le « d » a été utilisé dans les premières méta-analyses en particulier celle de Smith et Glass (1977). Certains auteurs ont utilisé une formule voisine, le « d » de Cohen (1988).

$$d \text{ (taille d'effet)} = \frac{\text{moyenne du groupe traité} - \text{moyenne du groupe de comparaison}}{\text{écart type "poolé" des deux groupes}}$$

Il existe de très nombreuses versions de ce « d », sans véritable consensus sur son utilisation.

Comparaison des tailles d'effet

Il faut rapporter non seulement les tailles d'effet mais aussi leurs intervalles de confiance à 95 % et comparer les différentes tailles d'effet pour voir si elles se recoupent ou non.

Homogénéité des études

La méta-analyse prend pour hypothèse que la taille d'effet représente la somme de toutes les études rassemblées en une seule grande étude. Cependant, certaines études peuvent aller dans un sens et d'autres en sens inverse. Par exemple, une partie des études peut montrer que le traitement à l'étude est supérieur au placebo, alors que d'autres études trouvent une égalité avec le placebo. Les moyennes peuvent ne donner qu'un reflet imparfait des résultats. Il faut donc compléter la méta-analyse, même si elle est en faveur de l'efficacité du traitement, par une statistique qui définit si les études sont homogènes ou non, pour, ensuite, rechercher les facteurs d'hétérogénéité. Par exemple, lorsque des études divergent fortement sur un graphique (« *outliers* »), on peut refaire l'analyse sans ces études et étudier en détail leurs modalités. Pour déterminer l'hétérogénéité, on utilise en pratique une version particulière du χ^2 la statistique d'hétérogénéité « Q ». On trouvera le détail de ces calculs dans des travaux spécialisés (Hedges et Olkin, 1985 ; Boissel et coll., 1989 ; Gloaguen et coll., 1998).

Estimation des études non publiées

Une méta-analyse doit couvrir l'ensemble des données publiées. Or certaines études ne sont pas publiées pour des raisons diverses (« *file drawer problem* »). Certaines études présentées en détail dans les congrès et dont il existe des données écrites sous forme de posters ou de résumés ne sont jamais publiées. Dans ce dernier cas, les études peuvent être recherchées. Il est possible de calculer une statistique (*failsafe*) qui tient compte de l'absence potentielle d'un certain nombre d'essais non publiés. Cette statistique calcule le nombre d'études publiées qui réduirait les tailles d'effet moyennes obtenues selon une formule : $k[(dk/dc) - 1]$, « k » représente le nombre d'essais publiés, « dk » moyenne des tailles d'effet, « dc » la valeur selon laquelle l'effet obtenu deviendrait insignifiant : cette valeur est fixée habituellement à 0,05 ; certains ont même suggéré une valeur plus restrictive à 0,20 (Whittal et coll., 1999).

Règles d'or pour évaluer les méta-analyses

On peut, à partir des données de la littérature, créer une échelle mesurant la qualité d'une méta-analyse (tableau 2.V).

Tableau 2.V : Proposition de règles d'or pour évaluer la qualité des méta-analyses

Critères	Notation
----------	----------

Inclusion de tous les essais de qualité sur le sujet	0/0,5/1
Critères de jugement clairement définis	0/0,5/1
Utilisation de méthodes statistiques adéquates	0/0,5/1
Prise en compte de la puissance statistique	0/0,5/1
Comparaison des tailles d'effet	0/0,5/1
Test de l'homogénéité des études	0/0,5/1
Estimation des études non publiées	0/0,5/1

Les sept critères notés de 0 à 1 permettent de donner une estimation globale de la qualité d'une méta-analyse.

Médecine fondée sur les preuves et efficacité des psychothérapies

C'est la possibilité de reproduire un travail expérimental par une équipe de recherche indépendante de l'équipe initiale qui en prouve la validité. Les psychothérapies n'échappent pas à l'analyse de leur efficacité en fonction des preuves de celle-ci. Les principes de l'« *evidence based medicine* », ou médecine fondée sur des preuves, s'y appliquent. La règle est d'analyser la littérature d'une manière hiérarchisée. Dans cette perspective, on étudie dans l'ordre : les méta-analyses ; les études contrôlées ; les études non contrôlées. En l'absence de méta-analyses et d'études contrôlées ou non contrôlées, l'existence d'un consensus fort chez les professionnels concernant les traitements est prise en compte. Plusieurs systèmes qui permettent d'établir des preuves de l'efficacité des psychothérapies ont été proposés.

Niveaux de preuve : Van Tulder et coll. (2002) (*Cochrane review*)

- Niveau 1 : une preuve forte est apportée par des résultats cohérents dans de nombreuses études contrôlées randomisées de haute qualité ;
- Niveau 2 : une preuve moyenne est apportée par des résultats généralement cohérents dans une étude contrôlée de haute qualité et une, ou plusieurs, étude(s) contrôlée(s) de basse qualité ou par des résultats généralement cohérents dans de multiples études contrôlées de basse qualité ;
- Niveau 3 : preuves limitées ou équivoques : seulement une étude contrôlée randomisée, soit de haute ou de basse qualité, ou des résultats contradictoires dans des études contrôlées multiples.
- Niveau 4 : pas de preuve lorsqu'il n'y a aucune étude contrôlée.

Force de la preuve : Eccles et coll. (1998)

Une autre méthode voisine est celle de la « force de la preuve » qui est due à Eccles et coll. (1998) citée et utilisée dans : « *Department of Health. Treatment choice in psychological therapies and counselling. Evidence based practice guideline* » (London, February 2001). Cet ouvrage est le plus récent sur le sujet de l'évaluation des psychothérapies. La force de la recommandation d'une méthode est fondée sur des preuves :

- Force A : directement basée sur la catégorie I ;
- Force B : directement basée sur la catégorie II de preuves ou extrapolée de la catégorie I ;

- Force C : directement basée sur la catégorie III de preuves ou extrapolée de la catégorie II ;
- Force D : directement basée sur la catégorie IV de preuves ou extrapolée de la catégorie III.

Catégories I à IV

- Ia : preuve établie d'après des études méta-analytiques et randomisées ;
- Ib : preuve établie d'après au moins une étude contrôlée ;
- IIa : preuve établie d'après au moins une étude contrôlée sans randomisation ;
- IIb : preuve établie d'après au moins un autre type d'études quasi-expérimentales ;
- III : preuve établie d'après des études descriptives comme les études comparatives, des études corrélationnelles et des études cas-témoins ;
- IV : preuves venant des opinions des experts ou de comités d'experts ou de l'expérience clinique ou d'une autorité respectée ou des deux.

Niveaux de preuve scientifique fourni par la littérature selon l'Anaes

Ces niveaux sont proches des niveaux de preuve que nous avons précédemment envisagés. Les critères de l'Anaes (2001) représentent bien cette opérationnalisation des données disponibles concernant un problème de santé publique précis. Ils permettent un classement facile des données de la recherche aussi bien que de la clinique pour éclairer les décisions. Le tableau 2.VI représente les niveaux de preuve selon l'Anaes.

Tableau 2.VI : Niveau de preuve scientifique et grades des recommandations selon l'Anaes (2001)

Niveau de preuve scientifique fourni par la littérature (études thérapeutiques)	Grade des recommandations
Niveau 1 Essais comparatifs randomisés de forte puissance Méta-analyse d'essais comparatifs randomisés Analyse de décision fondée sur des études bien menées	A : Preuve scientifique établie
Niveau 2 Essais comparatifs randomisés de faible puissance Études comparatives non randomisées bien menées Études de cohorte	B : Présomption scientifique
Niveau 3 Études cas-témoins	C : Faible niveau de preuve
Niveau 4 Études comparatives comportant des biais importants Études rétrospectives Séries de cas	

Critères d'évaluation des psychothérapies chez l'enfant et l'adolescent

À la suite d'une « *task force* » de l'*American psychological association* aux États-Unis et d'un travail du *National health service executive* du Département britannique de la Santé faisant le

point de la recherche sur les psychothérapies pour les troubles mentaux de l'adulte (*American psychological association*, 1995 ; Chambless et coll., 1996 et 1998), une *task force* similaire a été mise en place par l'*American psychological association*, à la fin des années 1990, pour expertiser les données existantes sur l'efficacité des interventions psychosociales dans les troubles mentaux de l'enfant et de l'adolescent. Des critères, adaptés de ceux élaborés par Chambless et coll. (1996) pour l'adulte, ont été proposés pour définir deux niveaux d'évidence empirique ; ils sont présentés dans le tableau 2.VII.

Tableau 2.VII : Critères retenus pour l'évaluation des psychothérapies chez l'enfant et l'adolescent définis par l'*American psychological association* (Lonigan et coll., 1998)

Critères pour les interventions « d'efficacité bien établie »

Au moins 2 essais intergroupe, bien conduits, par des investigateurs différents, démontrant l'efficacité d'un traitement de l'une des deux manières : le traitement est supérieur à un placebo ou à un traitement alternatif, ou le traitement est équivalent à un traitement d'efficacité déjà établie, avec une puissance statistique suffisante (n = 30 par groupe)

Ou série importante (par exemple, plus de 9) d'études de cas unique utilisant un bon plan expérimental et comparant l'intervention à un autre traitement

Et manuels de traitement utilisés pour les interventions, de préférence

Et les caractéristiques de l'échantillon d'étude sont clairement spécifiées

Critères pour les interventions « probablement efficaces »

Deux études montrant que l'intervention est plus efficace qu'une condition contrôle de non-traitement (ex. liste d'attente)

Ou 2 études intergroupe conduites selon les mêmes critères que pour les traitements bien établis, mais par le même investigateur

Ou petite série (par exemple, plus de 3) d'études de cas unique utilisant un bon plan expérimental et comparant l'intervention à un autre traitement

Et manuels de traitement utilisés pour les interventions

Et les caractéristiques de l'échantillon d'étude sont clairement spécifiées

Le terme « cas unique » se réfère à un plan expérimental dans lequel le traitement est alternativement appliqué et interrompu, soit à un individu, soit à un groupe d'individus.

BIBLIOGRAPHIE

ALFORD BA, BECK AT. The integrative power of cognitive therapy. The Guilford Press, New York 1997

AMERICAN PSYCHOLOGICAL ASSOCIATION. Task force on psychological intervention guidelines. Template for developing guidelines : Interventions for mental disorders and psychosocial aspects of physical disorders. Washington DC 1995

ANAES (AGENCE NATIONALE D'ACCREDITATION ET D'EVALUATION EN SANTE). Diagnostic et prise en charge en ambulatoire du trouble anxieux généralisé. Recommandations pour la pratique clinique. Paris, 2002. Résumé publié sur Internet : <http://www.anaes.fr>

APA (AMERICAN PSYCHIATRIC ASSOCIATION). Diagnostic statistical manual of mental disorders (Fourth edition) DSM-IV. American Psychiatric Association, Washington DC 1994. Traduction GUELFJ JD, Manuel statistique et diagnostique des troubles mentaux, Masson, Paris 1996

APA (AMERICAN PSYCHIATRIC ASSOCIATION). Handbook of psychiatric measures. Washington DC 2000

BARLOW D, HAYES SC, NELSON RO. The scientist practitioner. Research and accountability in clinical and educational settings. Pergamon, New York 1984

BELLACK AS, HERSEN M. Research methods in clinical psychology. Pergamon, New York 1984

BERNARD C. Introduction à l'étude de la médecine expérimentale. Garnier-Flammarion, Paris (1865) 1980

- BOISSEL, JP, BLANCHARD, J, PANAK E, PEYRIEUX JC, SACKS H. Considerations for the meta-analysis of randomized clinical trials. Summary of a panel discussion. *Control Clin Trials* 1989, **10** : 254-281
- BOUVARD M. Questionnaires et échelles d'évaluation de personnalité. Masson, Paris 1999
- BOUVARD M, COTTRAUX J. Protocoles et échelles d'évaluation en psychiatrie et en psychologie. Masson, Paris 1996, troisième édition 2002
- CHAMBLESS D, SANDERSON W, SHOHAM V, JOHNSON S, POPE K et coll. An update on empirically validated therapies. *Clin Psychol* 1996, **49** : 5-18
- CHAMBLESS DL, BAKER MJ, BAUCOM DH, BEUTLER LE, CALHOUN KS et coll. Update on empirically validated therapies II. *Clin Psychol* 1998, **51** : 3-15
- CLARK D, SALKOVSKIS P, CHALKLEY AJ. Respiratory control as a treatment for panic attacks. *J Behav Ther Exp Psychiatry* 1985, **16** : 23-30
- CLARKIN JF, PILKONIS PA, MAGRUDER KM. Psychotherapy of depression. Implications for the reform of the health care system. *Arch Gen Psychiatry* 1996, **53** : 717-723
- COHEN D. Statistical power analysis for behavioral sciences (2nd edition). Lawrence Erlbaum Associates, Hillsdale NJ 1988
- COTTRAUX J. Problèmes posés par l'évaluation des psychothérapies. *Psychol Med* 1988, **20** : 199-206
- COTTRAUX J, BOUVARD M, LEGERON P. Méthodes et échelles d'évaluation des comportements. Éditions d'application psychotechniques, Issy-les-Moulineaux 1985
- COTTRAUX J, BLACKBURN IM. Thérapies cognitives des troubles de la personnalité. Collection médecine et psychothérapie, Masson, Paris 1995
- COTTRAUX J, NOTE ID, CUNGI C, LEGERON P, HEIM F et coll. A controlled study of cognitive-behavior therapy with buspirone or placebo in panic disorder with agoraphobia. A one year follow-up. *Br J Psychiatry* 1995, **167** : 635-641
- CUCHERAT M, CIALDELLA P. La méta-analyse des essais thérapeutiques : application en psychiatrie. *Encephale* 1996, **XXII** : 378-387
- DERUBEIS RJ, GELFAND LA, TANG TZ, SIMONS AD. Medications versus cognitive behavior therapy for severely depressed outpatients : Meta-analysis of four randomized comparisons. *Am J Psychiatry* 1999, **156** : 1007-1013
- ECCLES M, FREEMANTLE N, MASON J. North of England evidence based guidelines development project : methods of developing guidelines for efficient drug use in primary care. *BMJ* 1998, **316** : 1232-1235
- FOA EB, MEADOWS EA. Psychological treatments for posttraumatic stress disorder : a critical review. *Annu Rev Psychol* 1997, **48** : 449-80
- FREUD S. On beginning the treatment : further recommendations on the technique of psychoanalysis. In : Standard edition of the complete psychological works of Sigmund Freud (vol. 12). STRACHEY J ed, Hogarth Press, London (1913) 1958
- FREUD S. Introductory lectures on psycho-analysis. Standard Edition, vol. 15. Hogarth Press, London (1917) 1961
- FREUD S, BREUER J. Études sur l'hystérie. Presses Universitaires de France, Paris (1895) 1956
- GABBARD GO. The economic impact of psychotherapy : a review. *Am J Psychiatry* 1997, **154** : 147-155
- GARFIELD AE, BERGIN SL. Handbook of psychotherapy and behavior change. Wiley, New York 1978
- GLASS GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976, **5** : 3-9
- GLOAGUEN V, COTTRAUX J, CUCHERAT M, BLACKBURN IM. A meta-analysis of the effects of Cognitive Therapy in depressed patients. *J Affect Disord* 1998, **49** : 59-72

- GREENWALD S, RUSSELL RL. New school for social research. Assessing rationales for inclusiveness in meta-analysis samples. *Psychotherapy Research* 1991, 1 : 17-24
- GUELFY JD (ed). L'évaluation clinique standardisée en psychiatrie (deux volumes). Éditions médicales Pierre Fabre, Castres 1993
- GUY W. ECDEU, Assessment manual for psychopharmacology. NIMH, Rockville MD 1976
- HEDGES LV, OLKIN I. Statistical methods for meta-analysis. Academic Press, New York 1985
- HERSEN M, BARLOW D. Single-case experimental designs. Pergamon Press, New York 1976
- JANET P. L'automatisme psychologique. Alcan, Paris 1889. Rééd. Odile Jacob, Paris 1998
- JENICEK, M. Méta-analyse en médecine. Évaluation et synthèse de l'information clinique et épidémiologique. Edisem & Maloine, Paris 1987 : 99-145
- KAZDIN A. Single-case research designs : Methods for clinical and applied settings. Oxford University Press, New York 1982
- LADOUCEUR R, BEGIN G. Protocoles de recherche en sciences appliquées et fondamentales. Edisem, Maloine, Paris 1980
- LONIGAN CJ, ELBERT JC, JOHNSON SB. Empirically supported psychosocial interventions for children : an overview. *J Clin Child Psychol* 1998, 27 : 138-145
- LUBORSKY L, SINGER B, LUBORSKY L. Comparative studies of psychotherapies : is it true that everyone has won and all must have prizes ? *Arch Gen Psychiatry* 1975, 32 : 995-1008
- LUBORSKY L, MCLELLAN AT, WOODY GE, O'BRIEN CP, AUERBACH A. Therapist success and its determinants. *Arch Gen Psychiatry* 1985, 42 : 602-611
- LUBORSKY L, BARBER JP, SIQUELAND L, JOHNSON S, NAJAVITS LM et coll. The revised helping alliance questionnaire (Haq-II) : psychometric properties. *J Psychother Pract Res* 1996, 5 : 260-271
- MAXFIELD L, HYER L. The relationship between efficacy and methodology in studies investigating EMDR treatment of PTSD. *J Clin Psychol* 2002 ; 58 : 23-41
- MARTIN P, PERETTI S. La qualité de vie : le point de vue du psychiatre. *Psychiatrie* 1999, 16 (Suppl. 2) : 16-27
- MEYERS LS, GROSSENS NE. Behavioral research. Theory, procedure and design. WH Freeman and Company, San Francisco 1974
- OBSESSIVE COMPULSIVE COGNITIONS WORKING GROUP. Psychometric validation of the Obsessive Beliefs Questionnaire and the Interpretation of Intrusions Inventory. Part I. *Behav Res Ther* 2003, 41 : 863-878
- OMS. The ICD classification of mental and behavioral disorders. WHO, Genève 1992
- PAUL GL. Behavior modification research : Design and tactics. In : Behavior therapy : Appraisal and status. FRANCKS CM ed, McGraw Hill, New York 1969
- POPPER K. Conjectures and refutations. The growth of scientific knowledge. Routledge and Kegan Paul, London 1963
- ROSENTHAL RL. Experimenter Effects in Behavioral Research. Appleton Century Crofts, New York 1966
- ROSENTHAL RL, ROSNOW R. Primer of methods for the behavioral sciences. Wiley, New York 1975
- ROTH A, FONAGY P. What works for whom ? A critical review of psychotherapy research. The Guilford Press, New York, London 1996
- SARTORIUS N, DE GIROLAMO G, ANDREWS G, GERMAN A, EISENBERG L (eds). Treatment of mental disorders. A review of effectiveness. WHO, American Psychiatric Press, Washington 1993

- SHADISH RW, MATT GE, NAVARRO AM ET COLL. Evidence that therapy works in clinically representative conditions. *J Consult Clin Psychol* 1997, **65** : 355-365
- SMITH ML, GLASS GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977, **32** : 752-760
- VAN TULDER MW, OSTELO RW, VLAEYEN JW, LINTON SJ, MORLEY SJ, ASSENDELFT WJ. Behavioural treatment for chronic low back pain (Cochrane Review). *In* : The Cochrane Library, Issue 4, 2002
- VENTUREYRA VA, YAO SN, COTTRAUX J, NOTE I, DE MEY-GUILLARD C. The validation of the Posttraumatic Stress Disorder Checklist Scale in posttraumatic stress disorder and non-clinical subjects. *Psychother Psychosom* 2002, **71** : 47-53
- WHITTAL ML, AGRAS WS, GOULD RA. Bulimia nervosa : a meta-analysis of psychosocial and pharmacological treatments. *Behav Ther* 1999, **30** : 117-135
- YAO SN, COTTRAUX J, MARTIN J, BOUVARD M. Etude de validation d'un questionnaire des pensées intrusives et de leurs interprétations (QPII) dans les troubles obsessionnels-compulsifs : importance de l'infériorité. *Revue Francophone de Clinique Comportementale et Cognitive* 1996, **1** : 14-24
- YAO SN, COTTRAUX J, MARTIN R, MOLLARD E, BOUVARD M et coll. Inferiority in social phobics, obsessive-compulsive and non-clinical controls. A controlled study with the inferiority scale. *In* : Behaviour and cognitive therapy today : essays in honour of Hans J. Eysenck. SANAVIO E ed, Pergamon, Elsevier, London 1998

3

Étapes historiques

On peut définir la psychothérapie comme un ensemble de « méthodes psychologiques dont le but est de soulager une souffrance dans un cadre contractuel » (Gerin, 1984). Les méthodes d'évaluation qui, au début des psychothérapies, étaient fondées sur des études de cas plus ou moins nombreux, regroupant des patients semblables ou comparables par des auteurs ayant une grande notoriété, ont progressivement été complétées par des essais contrôlés qui ont connu un essor considérable ces trente dernières années dans tous les secteurs de soins : le registre des essais contrôlés de la « *Cochrane collaboration* » compte en 2003 plus de 37 000 entrées ayant trait aux essais contrôlés. Cette « médecine fondée sur des preuves » a cependant des antécédents lointains que l'on peut retrouver à la fin du dix-huitième siècle.

Antécédent aux travaux actuels

Un antécédent lointain aux travaux actuels de recherche contrôlée sur les psychothérapies peut être retrouvé à l'époque des Lumières.

Évaluation du baquet de Mesmer (1784)

En 1784, le roi Louis XVI désigna une commission pour étudier le magnétisme animal de Mesmer qui permettait des cures spectaculaires des troubles psychologiques et physiques, dus, selon Mesmer, à un blocage des fluides dans le corps. Le magnétisme était transmis par des baquets d'eau aimantée selon une procédure secrète impliquant des morceaux de fer. La commission, composée d'Antoine Lavoisier, Joseph Guillotin et Benjamin Franklin, effectua une étude en double aveugle : les sujets qui croyaient que l'eau était magnétisée avaient les réponses cathartiques attendues au fluide magnétique que l'eau fut magnétisée ou non. Il s'agissait donc d'un simple effet de suggestion. Mesmer affirma haut et fort que la magnétisation de l'eau n'avait pas été conduite dans les règles. Le « mesmerisme » continua à jouir d'un grand succès et survécut à ce camouflet expérimental. En effet, il conservait de nombreux partisans, parmi lesquels Lafayette, la reine Marie-Antoinette et Mozart (McNally, 1999).

Ceci confirme le point de vue de Max Planck (1949) et Roland Kuhn (1962), rapporté par Mahoney (1976). Selon Max Planck et Roland Kuhn, les idées erronées ne disparaissent pas parce qu'on a démontré à la fois leur fausseté et la valeur d'idées nouvelles, mais bien parce que leurs défenseurs sont morts et qu'une nouvelle génération a pris leur place pour instaurer un nouveau paradigme qui, à son tour, devra disparaître après une lutte plus ou moins longue, pour laisser la place à un paradigme meilleur.

Premiers travaux sur les effets des psychothérapies analytiques

Dès 1917 Coriat publiait un travail : « Quelques résultats statistiques concernant le traitement psychanalytique des névroses ». Mais le travail le plus important remonte à 1920, à Berlin,

sous l'impulsion d'Otto Fenichel, qui avait entrepris, malgré le scepticisme de Freud, des travaux systématiques destinés à préciser la technique psychanalytique. La publication la plus citée est celle de Knight (1941), qui effectua une étude rétrospective chez 592 patients traités par psychanalyse dans les instituts de Berlin, Londres et Chicago, ayant eu au moins six mois d'analyse. Knight proposa les critères d'évaluation suivants :

- amélioration symptomatique ;
- productivité accrue ;
- adaptation meilleure, et meilleur plaisir sexuel ;
- capacité à faire face aux conflits psychologiques ordinaires et à les résoudre ;
- amélioration des relations interpersonnelles.

Knight aboutit à la conclusion que 55,9 % des sujets ainsi traités avaient été guéris ou grandement améliorés. Cependant, en incluant les perdus de vue au cours des six premiers mois, il aboutissait à un taux de résultats positifs de 30 % seulement.

Évaluation globale de la littérature (Eysenck, 1952)

Le véritable début des recherches contrôlées est le fait d'Eysenck (1952), qui regroupa 19 études publiées. Eysenck aboutit à la conclusion que 44 % des patients en analyse présentaient une amélioration, alors que le taux d'amélioration spontanée des problèmes névrotiques était de 66 % sur deux ans. En outre, la psychothérapie dite éclectique n'améliorait que 64 % des patients, tandis que 72 % des patients névrotiques hospitalisés ou traités par des généralistes étaient améliorés. Il y avait donc, selon cette statistique, une relation négative entre la psychothérapie et la guérison. La méthodologie rudimentaire d'Eysenck, qui consistait en un regroupement d'études ayant trait à des problèmes disparates et une comparaison à un groupe contrôle arbitraire, a été critiquée par Garfield et Bergin (1978). Ces auteurs considèrent que le taux d'amélioration spontanée serait de 30 % et varierait beaucoup selon la nature des problèmes psychopathologiques étudiés. Quelles que soient ses limites, et ses partis pris, l'étude d'Eysenck eut le mérite d'ouvrir la voie à cinquante années de recherche contrôlée féconde.

Étude contrôlée randomisée (Paul, 1967)

L'étude de Paul (1967) représente une avancée du fait de sa méthodologie rigoureuse. Elle était effectuée sur une population d'étudiants ayant un déficit en compétences sociales et une anxiété sociale diffuse dont l'expression la plus commune était la peur de parler en public. Le protocole randomisé comparait quatre groupes :

- thérapie comportementale (désensibilisation systématique, Wolpe (1961 et 1975) : n = 15 ;
- thérapie psychodynamique : n = 15 ;
- attention placebo sous la forme d'une pseudothérapie n'apportant qu'un soutien relationnel : n = 15 ;
- contrôle : 44 sujets servaient de contrôles non traités.

Le traitement était administré en groupes de trois patients durant cinq séances pendant six semaines par des thérapeutes expérimentés.

L'évaluation était multimodale et indépendante. Elle comprenait une batterie de six mesures évaluant l'anxiété, la personnalité et des tests comportementaux. Ces derniers mesuraient la réponse de stress social *in vivo*, c'est-à-dire l'effet sur le symptôme cible. Enfin, après un suivi de deux ans, les effets étaient évalués par une analyse de variance à plusieurs facteurs, suivie de comparaisons multiples sur chacune des mesures. Cette étude a montré que le groupe soumis à une désensibilisation systématique avait 85 % de répondeurs. Le groupe en thérapie psychodynamique présentait des résultats équivalents au groupe attention placebo (50 % de répondeurs). Ces deux derniers groupes présentaient des résultats supérieurs au groupe contrôle non traité (22 % de répondeurs). Sur le suivi de deux ans, il n'apparut pas de rechute ni de substitution de symptômes dans le groupe qui avait reçu la désensibilisation systématique, d'après les mesures comportementales, d'anxiété et de personnalité.

Dans le groupe attention placebo, l'effet observé, particulièrement important (50 %), permettait d'éliminer le rôle, dans les résultats de la désensibilisation systématique, d'une relation thérapeutique non spécifique. La conclusion des auteurs était que la désensibilisation systématique représentait un traitement spécifique de l'anxiété sociale. Il s'agissait donc de ce qu'on appelle aujourd'hui une étude « pivotale », abordant aussi bien les aspects pragmatiques que les problèmes théoriques. L'étude de Paul (1967) présente également l'intérêt d'avoir fixé des standards pour l'ensemble de la recherche en psychothérapie.

Étude contrôlée randomisée avec des patients ambulatoires (Sloane, 1975)

À la différence de l'étude de Paul (1967), qui avait recruté des étudiants, l'étude de Sloane (1975a et b) portait sur 94 patients ambulatoires qui consultaient dans une clinique universitaire pour des troubles variés. Un tiers présentaient des troubles qui aujourd'hui seraient classés parmi les troubles de la personnalité, les autres souffraient par ordre croissant de plaintes corporelles, de soucis généralisés, d'une estime de soi abaissée, de difficultés relationnelles (surtout avec le sexe opposé), de difficultés interpersonnelles et d'anxiété généralisée.

La randomisation assigna 31 patients à la thérapie comportementale, 30 à la thérapie d'orientation analytique et 33 à une liste d'attente. Les deux groupes actifs reçurent environ 14 séances de psychothérapie sur une durée de quatre mois. L'évaluation comportait une mesure d'amélioration globale, des mesures de trois symptômes cibles, de personnalité et d'adaptation. Les thérapeutes étaient expérimentés et suivaient des lignes directrices écrites qui différençaient clairement les deux formes de psychothérapie. Le groupe contrôle recevait un coup de téléphone par semaine pour s'assurer que rien de grave ne s'était produit et pour annoncer une prise en charge prochaine.

Après quatre mois, la thérapie analytique et la thérapie comportementale montraient une amélioration des symptômes cibles chez les patients supérieure à celle des patients du groupe contrôle, sans différence entre elles. En ce qui concerne l'amélioration globale, 93 % des patients en thérapie comportementale, *versus* 77 % des patients aussi bien en thérapie analytique qu'en liste d'attente étaient améliorés ($p < 0,05$). Les trois groupes étaient améliorés sans différence sur les mesures d'adaptation sociale.

À un an de suivi, les patients ayant eu la thérapie comportementale présentaient plus d'améliorations des symptômes cibles que les patients en liste d'attente. Ceux qui avaient eu la thérapie analytique ne se différençaient pas de ceux de la liste d'attente. Cependant, les traitements additionnels reçus durant un an dans les deux groupes rendaient la comparaison difficile. Un autre fait intéressant est que 14 patients du groupe contrôle (42 %) avaient décliné l'offre de tout traitement après quatre mois, car ils s'étaient spontanément améliorés.

Une des conclusions de l'étude, après le suivi d'un an, était que la thérapie comportementale avait des effets à la fois spécifiques et plus profonds et diffus. Ce phénomène est comparable à la propagation d'une onde : l'effet de diffusion (*ripple effect*) des thérapies comportementales est retrouvé dans de nombreuses études en particulier dans notre pays (Mollard et Cottraux, 1984). À partir de cette période, la recherche sur les thérapies comportementales devait se développer considérablement, pour aborder à peu près tous les troubles psychopathologiques.

Revue générale de la littérature (Luborsky, 1975)

Luborsky et coll. (1975), après une analyse statistique de 40 études, aboutirent à la conclusion de l'efficacité de la psychothérapie en général et l'équivalence des différents types de psychothérapie. Toutes les écoles de psychothérapie « avaient gagné et méritaient un prix », verdict de l'oiseau Dodo dans « Alice au pays des merveilles ». En effet, dans cette revue, les résultats obtenus avec les différentes formes de thérapie étaient supérieurs à ceux obtenus dans le groupe contrôle. En fait, la méthodologie de Luborsky apparaît assez rudimentaire : elle consistait à établir des « boîtes de scores » pour inspecter les fréquences de résultats. Par exemple, pour la comparaison entre psychothérapies en général et contrôle, on pouvait observer la configuration suivante de la boîte de score :

- thérapie à l'étude meilleure que le comparateur : 20 ;
- thérapie à l'étude et comparateur égaux : 13 ;
- comparateur meilleur que la thérapie à l'étude : 0.

La revue de Luborsky incluait des études randomisées de recherches portant sur de « vrais » patients, recevant de « vrais » traitements (« *bona fide patients and treatments* ») qui étaient codés par deux évaluateurs indépendants, ce qui pour l'époque représentait un progrès. Mais aucun calcul n'est rapporté et les conclusions sont un simple commentaire sur ces fréquences. On peut aussi observer que la revue portait sur un mélange de divers troubles et de diverses formes de psychothérapie.

Luborsky, cependant, aboutit à quelques conclusions spécifiques : la thérapie comportementale apparaissait meilleure pour les phobies spécifiques que les autres méthodes, la pharmacothérapie était meilleure que la psychothérapie en général, et la combinaison pharmacothérapie plus psychothérapie était supérieure à la psychothérapie seule. La tentative de synthèse effectuée par Luborsky avait le mérite d'ouvrir la voie à une méthode plus sophistiquée : la méta-analyse.

Méta-analyse des effets psychothérapeutiques (Glass, 1976 ; Smith et Glass, 1977)

L'évaluation des effets et des résultats des psychothérapies s'est ensuite développée de manière exponentielle, ce qui obligea les évaluateurs à inventer une méthode statistique pour effectuer une analyse quantitative de l'ensemble des études disponibles : la méta analyse (Glass, 1976 ; Smith et Glass, 1977 ; Smith et coll., 1980). Devant la nécessité de rendre compte des effets des psychothérapies aux instances officielles et aux compagnies d'assurances, de nombreux travaux ont essayé d'évaluer, aussi bien globalement que par trouble psychopathologique, les résultats de la littérature. On peut observer, qu'à partir de cette époque, qui suit le premier choc pétrolier (1973), c'est le « tiers payant » qui représentera le demandeur social le plus pressant d'évaluation des psychothérapies. Un autre facteur est

l'éducation du public par les médias et le fait que les patients, en général, demandent de plus en plus de comptes aux soignants. La méta-analyse, ainsi que les classifications des maladies psychiatriques comme le DSM et l'ICD, a pour objectif de répondre à cette demande sociale. Cette époque a été marquée par des méta-analyses globales qui ont été fortement critiquées

Le travail initial le plus important reste celui du groupe de Smith et Glass (Smith et Glass, 1977 ; Smith et coll., 1980) qui, après avoir regroupé 475 études portant sur 25 000 sujets, a calculé la taille d'effet (*effect size*). Ils aboutirent à la conclusion que pour l'ensemble de ces études, elle était de 0,75 ; autrement dit, chez des personnes ayant eu une psychothérapie, l'amélioration moyenne était de 75 % supérieure à celle observée chez des sujets du groupe témoin. Ils regroupèrent les différents types de thérapies en deux grandes sous-classes :

- désensibilisation systématique, implosion, modification du comportement et Gestalt-thérapie ;
- thérapie psychanalytique, thérapie non directive de Carl Rogers, psychanalyse d'Adler, thérapie cognitive, analyse transactionnelle et thérapie éclectique.

Smith et Glass aboutirent à la conclusion qu'il n'y avait pas de différence entre ces deux sous-classes.

Critiques de la méta-analyse de Smith et Glass

Une des nombreuses critiques adressées à Smith et Glass était que le regroupement d'études disparates mettant sur le même plan des travaux effectués avec des étudiants volontaires et de véritables patients faisait que leurs conclusions ne reflétaient pas la réalité de la pratique clinique courante (Rachman et Wilson, 1980). De plus, certaines études importantes avaient été négligées et le regroupement des thérapies cognitives avec les thérapies psychanalytiques plutôt qu'avec la thérapie comportementale apparaissait comme une erreur importante. En fait, si l'on représente les tailles d'effet méthode thérapeutique par méthode thérapeutique, on s'aperçoit, comme le montre le tableau 3.I, que trois méthodes cognitivo-comportementales présentent les meilleurs résultats. Et que la seule méthode à avoir une taille d'effet importante, c'est-à-dire supérieure à 0,80, est la désensibilisation systématique.

Tableau 3.I : Résultats de la méta-analyse de Smith et Glass par méthode thérapeutique (d'après Rachman et Wilson, 1980)

Méthode	Taille d'effet
Désensibilisation systématique (TCC)	0,91
Thérapie cognitive (Ellis) (TCC)	0,77
Modification du comportement (TCC)	0,76
Thérapie adlérienne	0,71
Exposition en imagination (TCC)	0,64
Thérapie rogerienne (Thérapie humaniste)	0,63
Thérapie psychanalytique	0,59
Analyse transactionnelle (Thérapie humaniste)	0,58
Gestalt-thérapie (Thérapie humaniste)	0,26

Prioleau et coll. (1983), reprenant 32 études contre placebo de l'étude de Smith et Glass (1977), tout en obtenant la même taille d'effet que Smith et Glass, ont abouti à la conclusion qu'il n'y avait pas de différence entre un regroupement de diverses formes de psychothérapie et placebo. Les études étaient principalement des études ayant trait à la thérapie psychanalytique et aux thérapies humanistes, principalement rogerienne (82 %) ;

seulement 6 études sur 32 (18 %) portaient sur des thérapies cognitives ou comportementales.

Le fait que dans la revue de Smith et Glass il n'y avait que 22 % de « vrais » patients a entraîné d'autres travaux.

Andrews et Harvey (1981) ont repris, dans l'étude de Smith et Glass, 81 études ayant trait aux problèmes névrotiques de 2 202 vrais patients et ont classé les thérapies en trois catégories : comportementale (regroupée avec la thérapie cognitivo-comportementale), psychanalytique (regroupée avec la thérapie cognitive et la Gestalt-thérapie) et humaniste (thérapie non directive de Carl Rogers regroupée avec le conseil). Ils ont abouti à la conclusion qu'il y avait des différences entre les trois catégories. La thérapie comportementale, avec une taille d'effet de 0,97, était supérieure au placebo, à la thérapie analytique brève (taille d'effet de 0,74 par rapport au placebo) et à la thérapie humaniste (taille d'effet de 0,35 par rapport au placebo). La thérapie humaniste ne différait pas significativement du placebo (étude australienne).

Limites des méta-analyses globales : problèmes méthodologiques

La principale pierre d'achoppement des études méta-analytiques est le regroupement d'études concernant des pathologies ou des problèmes psychologiques variés et aussi les biais d'inclusion tenant à l'allégeance idéologique des auteurs. L'analyse de l'étude méta-analytique de Wampold et coll. (1997) permet de comprendre les conclusions abusives qui peuvent être tirées de méta-analyses qui, à la fois, sont globales et présentent des biais importants d'inclusion. Wampold et coll. (1997) comparaient des thérapies dite « de bonne foi » à des thérapies représentant de simples contrôles sans véritable intention de traiter. Selon ces auteurs, une thérapie de bonne foi est une thérapie dans laquelle le thérapeute, ayant un bon niveau de maîtrise et un manuel, développe une relation et adapte le traitement au patient. Il existait cependant une importante limitation à cette étude : elle n'incluait pas la totalité des données, contrairement aux règles habituelles de la méta-analyse. Seulement six journaux publiés aux États-Unis entre 1970-1995 étaient pris en compte. Les résultats exprimés en taille d'effet sont présentés dans le tableau 3.II.

Tableau 3. II : Résultats après regroupement de thérapies (d'après Wampold et coll., 1997)

Comparaisons	Taille d'effet
Thérapie <i>versus</i> pas de traitement	0,82
Thérapie <i>versus</i> placebo	0,48
Placebo <i>versus</i> pas de traitement	0,42
Différences entre thérapies « de bonne foi »	0,00-0,21

Mais on peut observer que sur les 114 études incluses, 79 (69 %) comparaient des formes différentes de thérapie cognitive et comportementale entre elles (Crits-Christoph, 1997). De plus, il est habituel de trouver une équivalence dans les essais contrôlés randomisés de ces méthodes qui se recoupent en bien des points (Cottraux et coll., 2001). En outre, seulement quatre impliquaient la thérapie analytique, aucune la thérapie familiale, trois incluaient des enfants et adolescents. Wampold et coll. soulignent d'ailleurs eux-mêmes qu'ils n'envisagent pas les catégories diagnostiques une par une, et ne peuvent donc conclure au sujet de l'efficacité de chaque type de thérapie pour chaque problème spécifique. Mais les auteurs ne devraient pas non plus parler de « psychothérapie » en général, puisque plus des deux tiers des études entrées dans la méta-analyse appartiennent au même courant psychothérapique, les thérapies cognitivo-comportementales, et représentent de ce fait la source de variation la plus importante.

On voit ici comment on peut faire dire à la méta-analyse ce qu'on souhaite entendre. Aussi, la recherche s'est maintenant centrée sur l'étude de traitements spécifiques appliqués à des troubles aussi définis que possible et les méta-analyses modernes s'efforcent d'inclure tous les essais de qualité, y compris ceux qui ont été publiés dans des revues obscures, présentés dans des congrès ou encore oubliés par leurs auteurs dans les tiroirs.

Historique des travaux sur l'évaluation des thérapies familiales

Concernant l'évaluation des thérapies familiales, les dix premières études d'évaluation ont été publiées entre 1972 et 1983, recensant 178 articles, dont 83 non contrôlés. Ces premières études n'ont pas abouti à un consensus clair concernant l'efficacité des thérapies familiales. Mais les revues de la littérature soulignent les défauts méthodologiques des études (caractères inadéquats des groupes contrôles, non fiabilité des mesures, absence de délai suffisant avant l'appréciation du suivi, aspect mal défini des fondements théoriques permettant les comparaisons). Au cours des années 1980, plusieurs méta-analyses et revues (Gurman et coll., 1986 ; Hahlweg et Markman, 1988 ; Hazelrigg et coll., 1987 ; Shadish et coll., 1993) fournissent plusieurs arguments pour conclure à l'efficacité des thérapies de couple et de famille mais sans apporter des résultats très précis. La méta-analyse de Shadish et coll. (1993) marque une étape dans l'évolution des travaux. Les auteurs proposent en effet des indications plus circonstanciées, des comparaisons entre différentes orientations théoriques en thérapie familiale et une analyse de la signification clinique des résultats en fonctions des orientations et des problèmes traités. Ils soulignent également l'importance de la prise en compte des méthodologies des études pour l'appréciation des résultats. Un autre aspect qui vient progressivement à l'ordre du jour dans les études est l'étude du maintien des résultats positifs sur le long terme. Enfin, l'effet des thérapies familiales est apprécié non seulement sur le trouble du patient concerné mais aussi sur le fonctionnement global de la famille. Pour les troubles de l'enfant, les thérapies familiales occupent une place de choix. Les travaux les plus conséquents se sont intéressés aux techniques psycho-éducatives qui mobilisent la famille et qui ont obtenu des résultats significatifs pour certains troubles (autisme en particulier). Les travaux les plus récents sont évoqués dans le chapitre sur l'évaluation des thérapies familiales

Historique des travaux d'évaluation de thérapies chez l'enfant

La première recherche publiée sur l'évaluation des psychothérapies chez l'enfant est probablement celle d'Evelyn Alpern en 1946, comparant deux approches thérapeutiques courantes : l'une dans laquelle l'enfant et ses parents sont vus et traités ensemble, l'autre dans laquelle l'enfant et ses parents sont vus et traités séparément. Les résultats de l'étude étaient exprimés en termes d'impression du clinicien sur les progrès de l'enfant, sans considération de catégories diagnostiques, ni randomisation aux deux conditions de traitement, ni évaluation standardisée pré- et post-traitement.

Dix ans plus tard, la revue de Levitt (1957) sur l'évaluation des psychothérapies chez l'enfant, contemporaine de celle d'Eysenck (1952) chez l'adulte, aboutissait, comme cette dernière, à la même conclusion pessimiste : les formes traditionnelles de psychothérapie ne sont pas plus efficaces que l'absence de traitement. Dans les années 1950 et 1960, la première génération d'études sur l'efficacité des psychothérapies posait la question de leur capacité à

induire une modification de la personnalité, sans spécification ni différenciation des formes de psychothérapie utilisées, ni des problèmes cliniques à traiter. L'essor de la thérapie cognitivo-comportementale (TCC) dans les années 1960 et 1970 a par la suite fortement contribué à promouvoir une approche plus scientifique, tant de la pratique clinique que de la recherche thérapeutique. La question « Est-ce que la psychothérapie marche ? » est alors remplacée par « Quelles procédures spécifiques sont les plus efficaces pour traiter tel ou tel problème clinique ? ». C'est le début des études avec allocation randomisée à l'un des différents traitements posés en comparaison. Depuis les années 1980, la troisième génération de recherches en psychothérapie utilise le modèle des essais cliniques contrôlés issu de la pharmacothérapie, les diagnostics DSM et des manuels décrivant précisément les traitements utilisés.

En conclusion, l'ère des comparaisons globales de méthodes psychothérapeutiques, s'appliquant à des troubles psychopathologiques variés regroupés et impliquant des volontaires, est révolue. La stratégie actuelle est de comparer pour chaque trouble selon l'ICD ou le DSM, des types de psychothérapie bien définis de manière opérationnelle entre elles, au placebo ainsi qu'à des traitements médicamenteux. Vu le nombre des essais thérapeutiques contrôlés portant sur des troubles précis, il est relativement facile d'étudier les thérapies cognitives et comportementales et la thérapie interpersonnelle. Il s'avère plus difficile d'étudier les thérapies familiales non cognitivo-comportementales, les thérapies psychanalytiques, l'hypnose ou les thérapies humanistes (Gestalt thérapie, analyse transactionnelle, thérapie rogerienne) bien qu'il existe des essais contrôlés pour ces différentes méthodes dans différentes indications. Il est de toute façon illusoire de vouloir étudier scientifiquement toutes les formes de psychothérapie étant donnée leur prolifération : 200 à 400 selon les diverses estimations.

BIBLIOGRAPHIE

- ALPERN E. Short clinical services for children in a child guidance clinic. *Am J Orthopsychiatry* 1946, **26**
- ANDREWS G, HARVEY R. Does psychotherapy benefit neurotic patients ? A reanalysis of the Smith, Glass and Miller data. *Arch Gen Psychiatry* 1981, **38** : 1203-1208
- CORIAT IH. Some statistical results of the psychoanalytic treatment of the psychoneuroses. *Psychoanal Rev* 1917, **4** : 209-216
- COTTRAUX J, NOTE ID, CUNGI C, LEGERON P, HEIM F et coll. A controlled study of cognitive-behavior therapy with buspirone or placebo in panic disorder with agoraphobia. A one year follow-up. *Br J Psychiatry* 1995, **167** : 635-641
- COTTRAUX J, NOTE I, YAO SN, LAFONT S, NOTE B et coll. A randomized controlled trial of cognitive therapy versus intensive behaviour therapy in obsessive compulsive disorder. *Psychother Psychosom* 2001, **70** : 288-297
- CRITS-CHRISTOPH P. Limitations of the dodo bird verdict and the role of clinical trials in psychotherapy research : comment on Wampold et al., 1997. *Psychol Bull* 1997, **122** : 216-220
- EYSENCK HJ. The effects of psychotherapy : An evaluation. *J Consult Psychol* 1952, **16** : 319-324
- GARFIELD AE, BERGIN SL. Handbook of psychotherapy and behavior change. Wiley, New York 1978
- GERIN P. L'évaluation des psychothérapies. Presses universitaires de France, Paris 1984
- GLASS GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976, **5** : 3-9

- GURMAN AS., KINSKERN DP, PINSOFF WM. Research on marital and family therapies. In GARFIELD SL, BERGIN AE. eds, Handbook of psychotherapy and behavior changes, 3rd edition, Wiley, New York, 1986 : 565-624
- HAHLWEG K, MARKMAN HJ. Effectiveness of behavioral marital therapy: Empirical status of behavioral techniques in preventing and alleviating marital distress. *J Consult Clin Psychol* 1988, **56** : 440-447
- HAZELRIGG MD, COOPER HM, BORDUIN CM. Evaluating the effectiveness of family therapies : An integrative review and analysis. *Psychological Bulletin* 1987, **101** : 428-442
- KNIGHT RP. Evaluation of the results of the psychoanalytic therapy. *Am J Psychiatry* 1941, **96** : 434-446
- KUHN R. The structure of scientific revolutions. University of Chicago Press, Chicago 1962
- LEVITT EE. The results of psychotherapy with children : an evaluation. *J Consult Psychol* 1957, **21** : 189-196
- LUBORSKY L, SINGER B, LUBORSKY L. Comparative studies of psychotherapies : is it true that everyone has won and all must have prizes ? *Arch Gen Psychiatry* 1975, **32** : 995-1008
- MAHONEY MJ. Scientist as a subject. The psychological imperative. Ballinger Publishing Company, Cambridge Massachussets 1976
- MCNALLY RJ. EMDR and Mesmerism : A comparative historical analysis. *J Anxiety Disord* 1999, **13** : 225-236
- MOLLARD E, COTTRAUX J. Changements positifs de l'inventaire multiphasique de personnalité du Minnesota (MMPI) chez 70 sujets traités par thérapie comportementale. *Psychol Med* 1984, **16** : 2385-2394
- PAUL G. Insight versus desensitization in psychotherapy two years after termination. *J Consult Psychol* 1967, **31** : 333-348
- PLANCK M. Scientific autobiography and other papers. Philosophical Library, New York 1949
- PRIOLEAU L, MURDOCH M, BRODY N. Analysis of psychotherapy versus placebo studies. *Behav Brain Sci* 1983, **6** : 275-285
- RACHMAN S, WILSON GT. The effects of psychological therapy, Pergamon, London, 1980
- SHADISH WR, MONTGOMERY LM, WILSON P, WILSON MR, BRIGHT I, OKWUMABUA T. *J Consult Clin Psychol* 1993, **61** : 992-1002
- SLOANE RB, STAPLES FR, CRISTOL A, YORKSTON NJ, WHIPPLE K. Short-term analytically oriented psychotherapy versus behavior therapy. *Am J Psychiatry* 1975a, **132** : 373-377
- SLOANE RB, STAPLES FR, CRISTOL A, YORKSTON NJ, WHIPPLE K. Psychotherapy versus behavior therapy. Harvard University Press, Cambridge, Massachusetts 1975b
- SMITH ML, GLASS GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977, **32** : 752-760
- SMITH ML, GLASS GV, MILLER TI. The benefits of psychotherapy. John Hopkins University Press, Baltimore 1980
- WAMPOLD BE, MONDIN GW, MOODY M, STICH F, BENSON K, AHN HN. A meta-analysis of outcome studies comparing bona fide psychotherapies : empirically, "All must have prizes". *Psychol Bull* 1997, **122** : 203-215
- WOLPE J. The systematic desensitization treatment of the neuroses. *J Nerv Ment Dis* 1961, **132** : 189-203
- WOLPE J. The practice of Behavior Therapy. Stanford University Press, 1969. In : La pratique de la thérapie comportementale. Traduction J. Rognant, Masson, Paris 1975