

2

Aspects méthodologiques de l'évaluation

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 précision 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

À la suite de Claude Bernard (1865), les psychologues américains 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.

12 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 psychothérapeutique (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 AB-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 lignes de base multiples 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-sujet 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, et peuvent venir pour cette raison 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.

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,

Tableau 2.1 : Principales échelles d'évaluation et questionnaires disponibles en français et dates de leur publication**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 Asberg (1979) ; Échelle de désespoir (Beck, 1974) ; Échelle d'attitudes dysfonctionnelles (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 (Asberg, 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 Schedule* (Tyrer, 1992) ; SCIDII-PQ (First et coll., 1997) ; PDQ-4 *Personality disorder questionnaire* (Hyler, 1994) ; *International personality disorders examination*, IPDE (Loranger, 1991) ; Inventaire de personnalité d'Eysenck, EPI (1968) ; Entretien diagnostique structuré 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)

Tableau 2.I (suite) : Principales échelles d'évaluation et questionnaires disponibles en français et dates de leur publication

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., 1998 ; Lecrubier et coll., 1998) ; 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)

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

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 méta-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).

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.

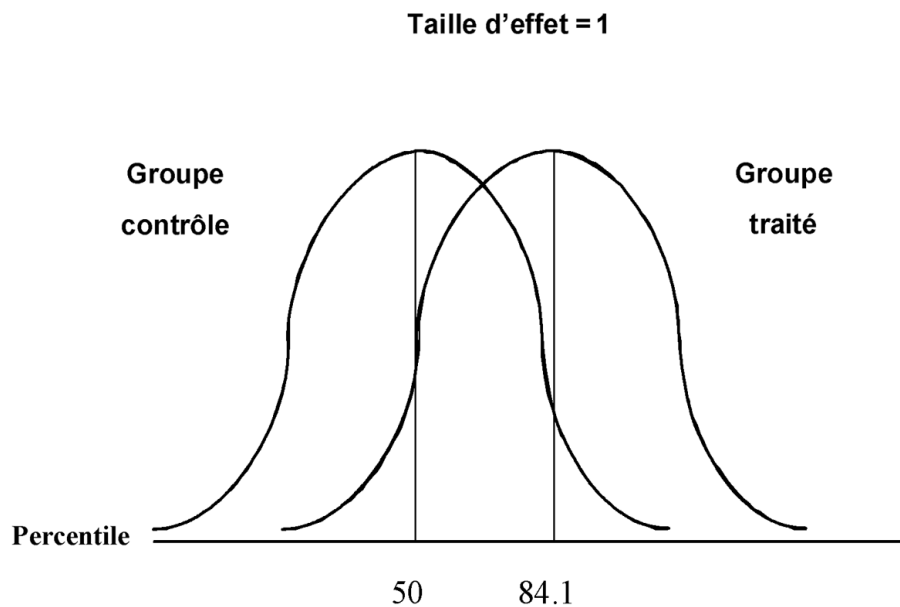


Figure 2.1 : Distribution des scores et taille d'effet

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

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.

Dans certaines études, la taille d'effet est calculée par la comparaison des scores pré-post-traitement. Elle est égale à la différence du score après traitement moins le score avant traitement divisée par l'écart-type. Cette taille d'effet tend à avoir une valeur supérieure à celle de la taille d'effet du traitement actif *versus* le contrôle puisque l'effet placebo n'est pas soustrait.

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 » la 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 soit 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 fournis 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 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'ÉVALUATION EN SANTÉ). 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 GUELFY 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, LÉGERON 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, LÉGERON 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, MOLLARDE 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