

Réponse à Perron et al concernant leur texte :
Quelques remarques méthodologiques à propos du Rapport INSERM
« Psychothérapie. Trois approches évaluées »

Philippe CIALDELLA, psychiatre, LYON
p.cialdella@free.fr

Remerciements : Messieurs les docteurs et professeurs Chabrol, Cottraux, Cucherat, Guelfi, Thurin, et Pierre-Henri Castel, psychanalyste et philosophe, ont bien voulu relire ce manuscrit alors qu'il était en cours d'élaboration, les docteurs Thurin et Cottraux étant deux des auteurs de l'expertise de l'INSERM. Cependant, les opinions exprimées dans le texte qui suit n'engagent que moi-même.

Introduction

L'expertise INSERM sur les psychothérapies (2004) a fait l'objet d'un grand nombre de critiques, parfois passionnées (Guelfi, 2004). Les enseignants de psychologie Roger Perron, Bernard Brusset, Clarisse Baruch, Dominique Cupa, et Miccele Emmanuelli ont proposé sur le site www.techniques-psychotherapiques.org géré par notre collègue Jean-Michel Thurin, une critique très détaillée qui aborde des questions méthodologiques (2004), mais leurs arguments ne me semblent absolument pas convaincants, aussi une contre-argumentation m'a-t-elle semblé utile.

Ayant une expérience de la méta-analyse, notamment à propos de l'efficacité des antidépresseurs sérotoninergiques (Bech et Cialdella, 1992, Cucherat et Cialdella, 1996, Bech et al, 2000, Bech et al, 2004), je répondrais principalement sur le terrain méthodologique, avec cependant un petit détour sur celui de l'éthique, pour répondre à certaines affirmations et questions abordés par Perron et al, mais qui sont liés à des aspects méthodologiques.

Une version plus courte de ce texte limitée aux seuls aspects méthodologiques a été proposée à la revue de psychiatrie l'Encéphale.

L'expertise INSERM ne concerne que l'efficacité de quelques psychothérapies, dont certains types de thérapies psychodynamiques (TP), en général de brève durée et plutôt éloignés de ce qui se pratique en France, ainsi que de thérapies cognitivo-comportementales (TCC), dans des pathologies psychiques clairement identifiées, ce qui laisse de côté un large champ d'utilisation non médicale de la psychothérapie. La psychanalyse en tant que telle n'est pas

évaluée par l'expertise et son efficacité ne peut être extrapolée par induction de celle des thérapies psychodynamiques.

J'ai sélectionné 18 citations dans le texte de Perron et al, qui me semblaient problématiques. Elles sont présentées in extenso en annexe et résumées dans le texte qui suit sous forme de questions numérotées. Les questions 16 à 18 sont plus spécifiquement éthiques.

Question 1. L'expertise INSERM satisfait-elle à des critères d'objectivité scientifique et ses méthodes sont-elles biaisées par des préjugés non scientifiques ?

Deux points épistémologiques ne seront pas détaillés dans ce texte, qui concernent la validité de l'opposition entre objectivité et subjectivité et la question de la scientificité et de ses critères. Quant à l'existence de préjugés, ils sont malheureusement inévitables, et on a même tenté de cerner cette réalité de façon précise dans les études sur les psychothérapies au moyen de la notion d'allégeance du thérapeute à sa méthode (Luborsky et al, 1999).

L'expertise de l'INSERM est une revue de littérature (niveau 3) et non une étude scientifique empirique, mais elle repose sur des études scientifiques empiriques (niveau 1) ou des méta-analyses (niveau 2), assemblées selon une méthodologie posée a priori, comme dans une étude empirique. L'expertise n'est cependant pas une méta-analyse (MA), car ses méthodes sont narratives alors qu'une MA utilise des méthodes statistiques. Les déclarations suivantes figurant dans le texte de Perron et al « *Il s'agit d'une " méta-analyse de méta-analyses "* ainsi que « *On dira qu'il s'agit d'une " méta-analyse " s'il s'y ajoute un effort de systématisation et d'évaluation des recherches ainsi colligées* » sont donc erronées.

Le jugement humain subjectif tient donc une place essentielle dans l'expertise INSERM, mais celle-ci donne au lecteur toutes les informations sur la revue de littérature effectuée, et semble avoir été quasiment complète, ce qui permet au lecteur de retourner aux sources (niveau 1) et d'apprécier par lui-même les choix des experts. Ces éléments confèrent à l'expertise une certaine objectivité.

Les règles méthodologiques de l'expertise n'ont pas eu l'assentiment de tous ses auteurs, puisque l'un d'entre eux (Dr JM Thurin) l'a publiquement signalé et a estimé avoir été trompé sur les véritables objectifs de l'expertise. Ceci n'enlève pas toute validité ni scientificité à l'expertise, qui a adopté les valeurs de l'Evidence-Based Medicine (EBM), approche certes contestable et contestée, mais néanmoins appuyée sur des démarches scientifiques empiriques. Adopter ces valeurs ne signifie pas automatiquement s'obliger à passer par les fourches caudines de méthodologies stéréotypées, car il y a encore de la place pour de la créativité en matière d'essais thérapeutiques, même si les grandes lignes méthodologiques

(comparaisons entre groupes, tirage au sort, utilisation des statistiques) semblent incontournables.

Malgré la restriction concernant le rôle de la subjectivité dans une expertise narrative, il est étonnant de lire que l'expertise a été menée avec des préjugés peu scientifiques, pour aboutir à un résultat biaisé programmé par avance.

Si nous suivons l'argumentation de Perron et al, qui se situe au niveau de la scientificité, alors le terme de biais ne peut être utilisé que dans son sens statistique. Rappelons ce qu'est un biais statistique dans le contexte des essais thérapeutiques comparatifs et par extension, de la MA ou d'une revue de littérature : Selon le Dr M Cucherat, pharmacologue à Lyon, « *un biais peut être introduit par une différence systématique entre les deux groupes, dû à autre chose qu'au traitement étudié et qui ne permet donc pas d'attribuer au seul traitement la différence observée entre les résultats* ». Le biais ou erreur systématique est différent de « l'erreur pure », liée à de multiples phénomènes, dont l'erreur de mesure, et qui n'avantage pas un des groupes. A l'inverse, le biais produit une différence avantageant un des groupes comparés.

Les sources de biais sont très nombreuses dans les études comparatives (Sackett, 1979), et les extirper totalement relève de l'idéal utopique. La plupart sont connues, leur impact est plus ou moins sérieux, certaines sont quantifiables et une partie est contrôlable. Les dénoncer de façon très générale, comme le font Perron et al, est un exercice facile, qui revient à une critique de la méthode scientifique, dont l'abandon pourrait engendrer des biais par faiblesse méthodologique.

Question 2. Est-il acceptable ou non sur le plan scientifique d'avoir réduit le champ de l'expertise INSERM aux situations pathologiques ?

Les troubles retenus dans cette expertise (schizophrénie, dépression, certains troubles anxieux, anorexie mentale, dépendance alcoolique, par exemple) ne sont pas des problèmes « bénins », et ont fait l'objet d'essais thérapeutiques de psychothérapie, mais sont également souvent traités à l'aide de médicaments. De ce fait, ces études sont importantes pour les patients qui souhaitent ne pas prendre de pharmacothérapie ou pour lesquels la pharmacothérapie est insuffisante, ce qui est le cas pour un très grand nombre de patients sinon pour la majorité.

La restriction de l'expertise INSERM à ces situations pathologiques n'a rien de surprenant, dans la mesure où la vocation de cet organisme est de promouvoir la recherche médicale, et où une très grande partie des psychothérapies est effectuée par des psychiatres et par des psychologues exerçant dans un champ médical (CMP, psychiatrie libérale, CHS), reposant sur une prise en charge par l'assurance maladie.

Sur un plan plus méthodologique, la réduction du champ d'investigation à des pathologies avérées permet de mettre en évidence une différence de résultats en fonction d'une thérapeutique plus facilement que chez des personnes présentant des problèmes plus légers, où l'effet placebo risque d'être plus important.

La remarque de Perron et al pose également un problème délicat en ce qu'elle assimile les troubles psychopathologiques à des « maladies ». Or ce terme, qui date de l'époque kraepelinienne, n'est plus utilisé dans les systèmes de classification psychiatrique modernes, et on est passé par exemple du terme de « névrose » ou de « psychose » dans les systèmes d'inspiration psychodynamique qui prévalaient avant le DSM-III (Bergeret et al, 1979) à celui de « trouble » dans les systèmes descriptifs et statistiques (DSM). On peut toutefois admettre que le DSM-IV est très influencé par un modèle médical décrivant des entités pathologiques discrètes (catégorielles) et que le terme de trouble est plutôt une précaution de langage.

Il est logique que Perron et al critiquent le choix d'un modèle purement descriptif (DSM) au profit d'un système plus psychodynamique. Mais les différences sont-elles si importantes si l'on s'en tient aux seuls aspects descriptifs ? Les psychodynamiciens ont sans doute une conception plus continue des troubles, avec la névrose à l'un des pôles et les sujets « névrotico-normaux » à l'autre, avec en position intermédiaire les personnalités névrotiques, mais ne font-ils pas une distinction claire entre les pôles ? Dans la conception psychodynamique, la « névrose » ne se réfère-t-elle pas à des particularités symptomatiques repérables et régulières ?

Or, il semble bien que ces caractéristiques régulières soient globalement les mêmes que celles qui figurent dans le DSM, avec la différence que les caractéristiques du concept de névrose sont éclatées dans le DSM en des syndromes cliniques réparties sur plusieurs têtes de chapitre de l'axe I (troubles anxieux, troubles de l'humeur, troubles somatoformes) et des troubles de personnalité placés sur l'axe II. Il faut également ajouter l'individualisation plus récente de troubles tels que les névroses dépressives ou les phobies sociales.

Hormis ces différences, on ne peut guère nier l'utilité d'un moyen de repérage des troubles et la différenciation entre syndromes cliniques classiques et formes moins sévères. Certaines névroses ont en effet un potentiel fortement handicapant, parfois égal ou supérieur des psychoses.

Question 3. Les investigateurs ayant réalisé des études contrôlées de TP ont-ils réellement la qualité de psychanalystes ?

Les TP mentionnées dans l'expertise sont d'une extrême variété et il est parfaitement logique de pointer cette hétérogénéité et de mettre en question la généralisabilité des résultats. Il est

également normal de formuler des doutes sur la façon dont les TP ont été appliquées dans bon nombre d'études, car les thérapeutes n'étaient parfois ni des médecins ni des psychologues, et parfois leur expérience de la thérapie était assez faible.

En revanche, il est étonnant que Perron et al doutent de la qualité de psychanalyste des promoteurs de ces études, pour ceux d'entre eux qui se présentent comme tels. Cette donnée est vérifiable notamment auprès des sociétés de psychanalyse, et au moins trois promoteurs des psychothérapies brèves cités dans l'expertise (Edmond Gilliéron, Lester Luborsky, Peter Sifnéos) sont incontestablement des psychanalystes renommés.

Question 4. Les manuels de psychothérapie utilisés dans les études sont-ils inacceptables ?

L'émergence des manuels ne peut se comprendre que dans un contexte historique : la MA des psychothérapies a été proposée par Glass (1976), créateur du terme de méta-analyse, en réponse à une revue de littérature du comportementaliste Hans Eysenck (1952) qui avait conclu à l'inefficacité des psychothérapies.

La MA de Smith et Glass (1977) fit apparaître que les psychothérapies étaient efficaces contre des conditions contrôle (en général liste d'attente), mais ses résultats furent obscurcis par le fait que toutes les thérapies semblaient marcher y compris des thérapies placebo (simulacres de thérapie). Cette MA faisait apparaître des différences de tailles d'effet en faveur des thérapies appartenant au champ des TCC, (tableau page 41 de l'expertise INSERM 2004). Ces différences furent confirmées par la MA d'Andrews et Harvey (1981) qui a repris les données de l'étude Smith et Glass en ne conservant que les vrais patients névrotiques. Le tableau I reprend les résultats de Smith et Glass

Tableau I : Résultats de la méta-analyse de Smith et Glass. Tailles d'effet selon la méthode thérapeutique rangées par ordre décroissant (d'après Rachman et Wilson, 1980)

| Méthode | Taille d'effet (effect-size) |
|---|------------------------------|
| 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 |

Mais ce tableau procède à une comparaison indirecte des thérapies entre elles, qui n'est pas probante au plan statistique car les conditions expérimentales peuvent avoir été très différentes selon les thérapies.

Les critiques de Glass lui ont reproché d'avoir confirmé le « verdict du dodo », du nom d'un passage « d'Alice aux pays des merveilles » où l'oiseau dodo déclare : « tout le monde a gagné et tous doivent avoir un prix ».

L'enjeu des recherches ultérieures fut donc de différencier facteurs thérapeutiques spécifiques et non-spécifiques. Pour démontrer que les psychothérapies vraies (bona fide) pouvaient avoir des effets différents des traitements placebo (Luborsky et al 1975) il est apparu important aux auteurs, de bien déterminer les ingrédients actifs de chaque thérapie, d'où l'utilisation des manuels.

C'est donc entre autres pour tenter de réfuter le « verdict du dodo » que des auteurs psychodynamiciens tels que Luborsky ont promu des manuels. Ces manuels permettent aussi de bien s'assurer que les types d'intervention diffèrent entre les diverses thérapies bona fide, la définition de chaque thérapie bona fide faisant intervenir une théorie explicative psychologique particulière. Lorsqu'il existe un manuel il est possible de vérifier par des enregistrements de séance si le thérapeute respecte ou non l'esprit et/ou la lettre du manuel.

Cependant, de nombreux auteurs, dont Wampold pensent que la différence ne se fait pas entre les ingrédients actifs mais entre les thérapeutes.

Question 5. La MA fait-elle perdre en précision ?

Non, la MA fait plutôt gagner en précision statistique car l'augmentation du nombre de sujets pris en compte (ou du nombre d'études) permet une réduction des écarts-types des estimations, ce qui est synonyme d'augmentation de précision de l'estimation de l'effet commun. Trop souvent les essais contrôlés de psychothérapies impliquent des échantillons de très petite taille et la MA est la seule façon scientifique d'augmenter la puissance statistique.

La divergence entre Perron et al et nous-même tient à ce que ces auteurs emploient le terme de précision dans un sens qui n'est pas statistique mais plutôt clinique. Perron et al se soucient plutôt des conséquences de l'hétérogénéité des méthodologies qui gommant les petits détails des prises en charge. Mais les chercheurs n'ont pas tous les mêmes objectifs et même si la méthode de l'essai contrôlé oblige à une certaine standardisation, la façon d'aborder le même thème clinique peut être très variable. On retrouve cette problématique dans la question suivante.

Question 6. Les méta-analyses accumulent-elles des biais liés à des présupposés communs à tous les chercheurs ?

L'expertise de l'INSERM n'étant ni une étude empirique, ni une MA, le terme de biais statistique est peu adéquat, mais une expertise est exposée à des risques d'erreur de conclusion similaires aux biais observés en MA.

En MA, les biais les plus sérieux proviennent de la sélection des études : en priorité le biais de publication, qui fait qu'une étude négative (qui ne va pas dans le sens souhaité par l'auteur) a moins de chances d'être publiée et donc incluse dans une MA. Ensuite le biais de sélection, qui fait qu'un auteur inclut dans une MA un sous-ensemble des études réalisées, en fonction de critères de sélection avantageant le traitement qui a sa préférence. Il est donc essentiel lorsqu'on lit une MA de scruter les indices pouvant témoigner de l'exclusion de certaines études, et d'obtenir des informations sur les raisons des exclusions, qui peuvent être liées à la mauvaise qualité méthodologique de certaines études. Ces notions sont bien repérées dans le texte de l'expertise et les raisons de l'exclusion semblent assez bien précisées. Toutefois, Perron et al ne mentionnent pas le potentiel biaisant très important des critères de sélection, non pas simplement au niveau de l'expertise elle-même, mais aussi au niveau de chacune des MA citées.

Enfin, parmi les autres problèmes fréquemment rencontrés dans les MA, on peut citer l'hétérogénéité clinique (problèmes des « oranges et des pommes ») et celui de l'hétérogénéité des résultats observés. Ce dernier problème peut être testé statistiquement et aboutir soit à invalider la MA, soit à la formulation de nouvelles hypothèses, si des études donnent des résultats non conformes à ce qui est observé dans la masse des autres études.

En revanche, Perron et al nous proposent un nouveau type de biais avec la théorie des préjugés communs à tous les chercheurs. Leur argument ne paraît pas fondé sur des faits observés, mais plutôt sur une opinion : Sont-ils en effet susceptibles de fournir un seul exemple où comme ils l'évoquent, cent études auraient été victimes du même biais lié à ces présupposés ?

Suivons les néanmoins : nous avons deux hypothèses, soit celle d'un aveuglement des chercheurs, soit celle d'un mimétisme dans la réalisation des protocoles de recherche.

L'hypothèse du mimétisme ou du conformisme est vraisemblable, pour des raisons de facilité, mais les chercheurs ont également intérêt à se singulariser, et à innover. Cet impératif favorise la critique, ce qui peut contrebalancer l'effet de conformisme. La littérature scientifique en général abonde en disputes méthodologiques et en dénonciation de biais possibles, et celle qui est consacrée à l'évaluation des psychothérapies n'est pas exempte de cette tendance.

Si le conformisme joue un rôle important en psychopharmacologie, notamment en raison des exigences standardisées des agences d'enregistrement du médicament, le mimétisme est moins vraisemblable dans les recherches sur les psychothérapies, car les études sont très disparates, même s'il existe actuellement une tendance à la convergence des méthodes.

Envisageons maintenant l'hypothèse d'un aveuglement généralisé des chercheurs sur un point essentiel passé inaperçu. Mais lequel ? La grande variété des approches psychothérapeutiques retrouvées par l'expertise de l'INSERM autorise à penser que tous les auteurs ne partagent pas les mêmes préjugés concernant ces dernières mais que chacun peut avoir à cœur de démontrer l'efficacité de la thérapie qu'il défend, ce qui est une forme de préjugé parfaitement identifié notamment par Luborsky et al (1999) sous le terme « d'allégeance à la méthode ».

Un biais pourrait en résulter en faveur de la psychothérapie en général, par rapport aux groupes contrôle en liste d'attente. Un exemple fameux de ce présupposé peut être trouvé chez Glass lui-même qui raconte (Glass, 2000) qu'il a pu se sortir d'un état qui rendait sa vie « misérable » selon ses termes, grâce à l'aide d'une psychothérapie qui a duré huit ans. C'est de cette expérience personnelle qu'il a tiré sa motivation en faveur de la recherche psychothérapeutique et a pu réaliser sa méta-analyse de 1977.

Mais la remarque de Perron et al porte plutôt, si on anticipe la suite de leurs arguments sans forcer l'interprétation, sur la méthodologie des essais thérapeutiques : ils pensent que l'acceptation « aveugle », par les chercheurs en psychothérapie, des règles méthodologiques des essais thérapeutiques, en fonction d'un « présupposé » favorable à cette méthodologie, serait néfaste pour les TP, et bienfaitrice pour les TCC. Je reviendrai sur ces arguments en examinant la question n°10.

Mais dans ce cas, la thèse des études qui seraient « *biaisées de la même façon* » serait réfutée du fait que le biais serait favorable à certaines thérapies et défavorable pour d'autres.

Étant donnée la quasi-impossibilité de s'abstraire de tout biais dans les essais thérapeutiques, la critique générale de Perron et al aboutit à jeter le discrédit sur la recherche psychothérapeutique dans son ensemble, sans reconnaître les avantages qu'elle a pu apporter, notamment dans la contestation du point de vue négativiste d'Eysenck.

Question 7. Les MA risquent-elles de faire perdre le contact avec les études de niveau 1 et de renforcer les biais des études de ce niveau ?

L'éloignement évoqué par Perron et al entre niveau 3 et 1 est un réel problème, mais le lecteur est informé très précisément sur les références bibliographiques et peut dès lors retourner aux sources des rapports des études de niveau 1. Ces études constituent d'ailleurs une sorte de pool, dans lequel différents auteurs ont pioché et produit, selon des critères différents, des

assemblages d'études dont la MA a pu parfois donner des résultats différents. Ces différentes MA sur des données communes permettent d'ailleurs de se faire une idée des biais de sélection (études de sensibilité).

Rappelons aussi qu'une MA n'élimine pas les biais des études incluses, lesquels peuvent jouer dans des sens très différents et ne se renforcent ni ne s'annulent forcément (Matt et Navarro, 1997). Mais, s'il y a un grand nombre d'études peu biaisées, l'impact de ces biais peut être dilué, en raison du nombre élevé des sujets dans les MA.

Question 8. La comparabilité des groupes peut-elle vraiment être assurée dans les essais comparatifs ?

Ce problème est bien connu : il est nécessaire que les groupes soient comparables sur tous les critères possibles sauf celui qu'on va faire varier expérimentalement, en général le traitement. Cette comparabilité est difficile à garantir, mais l'objectif recherché est la comparabilité, pas l'identité. Les essais thérapeutiques ne pouvant reposer sur des clones, aucune garantie de copie exacte des deux groupes ne pourra jamais être obtenue.

Cependant, avant d'examiner les résultats d'une étude, on évalue statistiquement les différences entre les groupes sur des critères importants qui pourraient perturber la comparabilité. Lorsque les auteurs d'une étude négligent de produire cette comparaison, l'étude est de qualité médiocre.

Question 9. Le DSM utilisé comme critère de sélection des sujets dans les études, est-il suffisamment accepté dans notre pays et peut-il être la cause d'un biais favorisant les TCC ?

Le DSM-IV est utilisé par de très nombreux psychiatres et reconnu par beaucoup comme un progrès par rapport à d'autres approches nosographiques. L'avantage scientifique du DSM-IV ne réside bien sûr pas dans le fait qu'il s'agit de la « *meilleure démarche possible* » (INSERM, 2004) ce qui est effectivement un peu court, mais plutôt dans le fait que cette classification, depuis l'édition III (1980), a fait l'objet de nombreuses études de fidélité inter-juges des catégories diagnostiques, d'où son intérêt pour constituer des groupes suffisamment homogènes, ce qui est un critère tout à fait scientifique.

De plus, le DSM présente l'intérêt d'être quasi-exclusivement descriptif et donc, acceptable depuis sa version III, par à peu près tous les courants de la psychiatrie, bien qu'au sens de plus petit commun dénominateur et non de classification consensuelle.

On ne voit pas, à l'heure actuelle, quel autre système de classification pourrait faire l'objet d'une acceptation aussi large, ce qui ne signifie pas non plus que le DSM est dorénavant figé,

car c'est un système qui a connu un certain nombre de changements significatifs, le prochain pouvant être majeur s'il inclut une meilleure prise en compte de la dimensionnalité des troubles mentaux.

Bien sûr, les psychodynamiciens sont très critiques à l'égard du DSM, mais des critiques existent aussi chez certains cognitivo-comportementalistes qui reprochent au DSM son aspect trop catégoriel dans la description des troubles de personnalité de l'axe II, alors qu'un modèle dimensionnel aurait beaucoup plus leur faveur. Certains chercheurs en psychiatrie biologique (par ex Van Praag, 2000) préféreraient eux aussi un modèle dimensionnel et reprochent au DSM de freiner les avancées de la recherche biologique. Si le DSM peut avantager une partie de la psychiatrie c'est plutôt la psychopharmacologie, qui est hors du domaine considéré par l'expertise.

Un des problèmes importants du DSM réside dans la catégorisation des troubles : les patients des études ne ressemblent pas assez aux patients reçus dans les cabinets de consultation : symptômes en dessous du seuil de détection d'un trouble (états subsyndromiques), ou appartenant à deux catégories proches mais sans satisfaire aux critères de l'un ou l'autre, et surtout comorbidités (associations de plusieurs troubles) qui semblent être la règle plutôt que l'exception. Les résultats de l'étude épidémiologique européenne en population générale ESEMeD 2000 (Alonso et al, 2004) montrent par exemple qu'on observe une comorbidité chez 81% des patients souffrant d'agoraphobie, chez 73,3% des patients souffrant de dysthymie, chez 69,4% des patients souffrant d'un trouble anxieux généralisé, et 63,8 % des patients souffrant de trouble panique. Or les patients présentant des comorbidités sont rarement inclus dans des études.

Question 10. L'utilisation de la quantification dans l'évaluation des psychothérapies produit-elle automatiquement un biais en faveur des TCC, en raison d'un clivage épistémologique entre approches nosologiques et approches fonctionnelles ?

Le premier niveau de réponse possible est épistémologique, mais sera juste esquissé dans cet article.

La conception d'une psychiatrie clivée en seulement deux grands modèles opposés est très contestable. Dès qu'on envisage l'étiologie la diversité règne, y compris au sein des écoles psychodynamiques et c'est une banalité de dire que la psychiatrie compte non pas deux mais trois grands courants étiologiques, biologique, psychologique et social. Le courant de psychiatrie sociale ne peut par exemple pas se réduire à l'approche systémique, (évaluée au titre des thérapies familiales dans l'expertise INSERM) qui n'en est qu'un des représentants.

Un deuxième point de désaccord concerne l'affirmation selon laquelle les approches descriptives-symptomatiques et psychodynamiques-fonctionnelles seraient exclusives l'une de l'autre. On ne voit pas bien pourquoi la description des symptômes et des syndromes serait de nature à empêcher de formuler des théories fonctionnelles ou étiologiques à leur sujet. Tout psychiatre a besoin de modèles étiologiques qui peuvent inclure l'approche psychodynamique mais sans forcément s'y résumer. Nous posons ici un lien entre « fonctionnel » et « étiologique », car le terme « fonctionnel » suppose une fonction donc une transformation d'un input en output, et si la structure responsable de la fonction est défectueuse, c'est dû à une ou plusieurs causes.

Au fond ce qui distingue l'approche descriptive du DSM de celle des psychodynamiciens, c'est que le DSM décrit les output de façon statique (les symptômes comme déviations par rapport aux output attendus) là où les psychodynamiciens décriraient le processus de dysfonctionnement de la structure productrice d'output. Ce dernier point est confirmé lorsque Perron et al parlent d'une approche psychodynamique « *qui (sur le modèle de la recherche en physiologie depuis Claude Bernard) décrit des structures fonctionnelles et en analyse les troubles* », mais sans qu'ils précisent ce que le terme « fonctionnel » recouvre pour eux. La liste figurant dans le texte de l'expertise, qui n'engage bien sûr pas Perron et al, mentionne comme fonctions possibles les notions d'insight, d'identification, la distanciation, l'autonomie psychique, la qualité des relations interpersonnelles, l'utilisation des capacités, la nature des investissements. Ces fonctions sont-elles parfaitement identifiées et obéissent-elles à des lois précises ou font-elles l'objet de débats dans les communautés psychanalytiques ?

L'évocation des travaux de Claude Bernard comme référence de l'approche fonctionnelle est tout de même surprenante, car ce dernier cherchait des lois physiologiques très précises, déterministes, et concevait la maladie comme déviation quantitative d'une fonction normale. Si Claude Bernard a toujours été assez opposé aux statistiques et à la notion de moyenne, ses travaux notamment sur la régulation de la glycémie montrent bien qu'il utilisait la mesure et qu'il accordait un rôle utile à la quantification. Claude Bernard nous semble donc plutôt conduire à Pavlov qu'à Freud, et aurait sûrement hésité à condamner la méthodologie des essais thérapeutiques s'il vivait de nos jours, car à son époque la statistique était embryonnaire.

Le DSM, quant à lui, n'est pas dépourvu de possibilités d'explications fonctionnelles, à commencer par la définition même d'un trouble qui se rapporte explicitement au terme de dysfonctionnement à la fois dans une perspective causale hypothétique, (une fonction normale perturbée) et dans une perspective de conséquence sociale (gêne occasionnée par le trouble). Si l'on constate que l'étiologie est peu présente dans le DSM, elle n'est pas non plus absente,

si l'on se réfère aux pathologies réactionnelles et secondaires à des problèmes physiques ou à des abus de substance, ou aux troubles spécifiques à l'enfant. Lorsque l'étiologie est absente, l'organisation structurale du DSM par regroupement de troubles reposant sur des ressemblances ou des noyaux symptomatiques communs au sein des têtes de chapitre permet de supposer une origine fonctionnelle commune. Ceci permet d'expliquer en partie les phénomènes que j'ai proposé ailleurs (Cialdella, 2001) de qualifier de comorbidité intra-classe (association de troubles au sein d'une même tête de chapitre).

Le clivage épistémologique proposé par Perron et al entre approche descriptive et fonctionnelle n'est donc pas si étanche.

Toujours au niveau épistémologique, Perron et al considèrent que les « *traitements statistiques, ne peuvent jouer qu'un rôle secondaire, et difficilement de preuve, dans l'approche fonctionnelle* ». En statistique la preuve ne provient que du rejet des hypothèses nulles (hypothèses d'équivalence des effets des traitements) et ce rejet comporte un risque d'erreur alpha (α). Mais, dans les expressions « médecine fondée sur les preuves » ou « niveau de preuve », le mot « preuve », traduisant le terme anglais « evidence », signifie plutôt « fait » (« on parle aussi de médecine factuelle »), par opposition à « opinion » ou « théorie ».

Les réticences de Perron et al portent probablement plus sur les mesures que sur les statistiques à proprement parler, qu'ils assimilent de façon restrictive à des comptages de fréquence. Ces auteurs se réfèrent visiblement à une définition étroite de la mesure. Dans la conception représentationnelle très large de Stevens (1946), au moins trois types de mesures sur les quatre décrits peuvent être utilisées en psychiatrie et psychologie : le premier type de mesure est binaire (un phénomène se produit ou non), le second catégoriel (un même phénomène apparaît sous divers aspects) et le troisième est ordinal (un phénomène se présente avec des différences de degré), chacun autorisant des traitements statistiques différents. Les mesures ordinales sont parmi les plus employées car elles représentent bien la perception que nous avons de nombreux phénomènes psychiques, comme étant affectés d'une variabilité de degrés, mais sans que nous puissions supposer que la structure du phénomène soit susceptible d'une plus grande précision quantitative.

Selon cette conception large de la mesure et en accord avec la supposition de fonctions psychiques, il n'est pas surprenant qu'existent des instruments d'évaluation psychodynamiques, parmi lesquels sont cités dans l'expertise : le Karolinska Psychodynamic Profile, le Core Conflictual Relationship Theme, le Defense Mechanism Rating Scale, et bien d'autres. Tous ces instruments ainsi que les échelles de symptômes mentionnées dans

l'ouvrage de Bouvard et Cottraux (2002) permettent d'évaluer le changement des patients quelle que soit la thérapeutique appliquée (médicamenteuse ou psychothérapique).

Le rôle des mesures est-il, comme le prétendent Perron et al, secondaire pour évaluer les changements fonctionnels ou psychodynamiques ? En clinique ordinaire peut-être puisqu'on raisonne sur l'individu, mais en recherche sur des groupes de patients, la mesure apporte des avantages essentiels pour synthétiser l'information évaluative et faciliter la communication au sein de la collectivité des chercheurs.

Un deuxième niveau de réponse est méthodologique : Perron et al semblent faire une confusion entre les critères de sélection, pour lesquels on utilise en général un instrument diagnostique catégoriel tel que le DSM-IV, qui permet de sélectionner un groupe relativement homogène d'individus en vue de l'essai thérapeutique, et les critères d'efficacité, qui reposent en général sur des échelles dimensionnelles symptomatiques ou non. Si les critères de sélection permettent en effet un comptage, c'est uniquement celui du nombre de patients inclus dans chaque groupe.

Perron et al ne semblent pas clairement percevoir que l'emploi d'échelles symptomatiques (type Hamilton ou Beck) comme critère d'efficacité pourrait être plus biaisant que l'emploi du DSM, le biais supposé étant celui favorisant les TCC. En effet ces thérapies sont en principe fortement axées sur la réduction symptomatique, que les TP plus axées sur les processus et les causes putatives. Cet aspect n'a pas été oublié dans la méthodologie de l'expertise et on ne peut en faire grief à leurs auteurs. De plus, de nombreuses études retenues dans l'expertise ont utilisé des critères d'évaluation autres que syndromiques, par exemple des évaluations du processus ou des changements psychodynamiques, ou des évaluations de la personnalité, des rôles sociaux ou de la qualité de vie, même s'il n'est pas évident que ces critères aient été placés en position équivalente à celle des symptômes.

On aurait pu attendre de l'expertise qu'elle fasse le point sur l'efficacité des psychothérapies avec deux types de critères en parallèle : des critères symptomatiques et des critères non-symptomatiques, comme par exemple la qualité de vie, mais si les critères symptomatiques sont assez standardisés et utilisés dans la grande majorité des études, les critères non-symptomatiques ne le sont pas. La multiplication des critères d'évaluation pose d'ailleurs de sérieux problèmes en raison des tests statistiques multiples et expose à un risque élevé d'erreur statistique de première espèce.

Un troisième niveau de réponse à la question « biais-quantification-approche fonctionnelle » est empirique : Le biais supposé favorisant les TCC lié aux échelles d'évaluation symptomatiques peut-il être évalué à partir des données existantes ?

Tout d'abord, un biais ne peut vraiment apparaître que dans des comparaisons directes randomisées TCC versus TP, les comparaisons indirectes étant faussées par des conditions expérimentales parfois très différentes.

Les psychanalystes Luborsky et al (2002) citent deux MA comparant directement TCC et TP, dans des pathologies regroupant troubles anxieux et dépressifs : la MA de Svartberg et Stiles (1991) montrait une différence significative en faveur des TCC sur les TP (effect size ou ES = 0,47), mais paradoxalement pas de différence entre TP et thérapie comportementale (ES = 0,10). Cette différence provenait d'un petit nombre d'études aussi la conclusion est-elle fragile. Luborsky et al (1999) ne trouvaient pas de différence significative entre TP et thérapie comportementale (ES = 0,03) ainsi qu'entre TP et thérapie cognitive (ES = -0,02). Les tailles d'effet (effect-size) dans ces MA de comparaisons directes dépassent rarement 0,20 ce qui est peu. On ne trouvait que 4 études communes à ces 2 MA, d'où l'importance des critères de sélection. La MA de Wampold et al (1997) qui concernait, pour les trois quarts des études, des comparaisons de thérapies cognitives versus comportementales, trouvait également un ES de la différence d'environ 0,20.

Pour les seules études de comparaisons directes TCC-TP par pathologie spécifique, décrites dans les pages 387-413 de l'expertise, on trouve dans le Trouble Anxieux Généralisé une supériorité significative des TCC, mais pas dans le Stress Post-Traumatique, l'anorexie et la boulimie. Même situation dans la dépression avec des résultats contradictoires entre les études, la MA la plus récente citée (2001) ne trouvait pas de différence significative entre TCC et TP chez des sujets âgés, mais deux autres MA plus anciennes trouvaient un ES de la différence proche de 0,20 en faveur des TCC.

Bien que sur des données un peu fragmentaires, il y a peu d'indication d'un biais majeur favorisant les TCC, puisque les différences entre TCC et TP, qui sont en général en faveur des TCC sont faibles, dépassent rarement un ES de 0,20.

Question 11. L'évaluation des psychothérapies suit-elle un modèle pharmacologique (comparaisons avant-après et emploi des grilles) inadapté ?

Cette méthodologie comparative avant-après est appliquée dans de très nombreux contextes scientifiques, et ne trouve pas ses origines en psychopharmacologie. Concernant les fameuses « grilles », elles ne sont pas très anciennes, puisque durant une première étape de la discipline (années 50-70), les résultats ont plutôt été appréciés selon des méthodes globales ou semi-globales et les critères du DSM ne sont apparus qu'en 1980.

Il n'y a donc pas lieu de faire d'amalgame entre la méthode scientifique comparative utilisée en psychopharmacologie et l'utilisation de grilles de symptômes.

Question 12. La perte de comparabilité entre les groupes au cours des études est-elle inévitable ?

L'attribution d'un sujet au groupe traité ou au groupe contrôle selon un tirage au sort évite l'arbitraire du chercheur, et empêche que les présupposés de ce dernier n'influencent trop les résultats et n'aboutissent à un biais.

Toutefois, une mauvaise comparabilité après tirage au sort peut entraîner un biais. Des événements peuvent modifier la comparabilité en cours d'étude, notamment les interruptions de traitement pour les patients en thérapie ou la perte de vue des patients en liste d'attente. Les interruptions dues à des raisons indépendantes du traitement et/ou de la maladie, comme un déménagement par exemple, n'entraînent pas de biais en général. Celles dues à des raisons dépendantes du traitement ou de la maladie comme l'aggravation de l'état clinique, sont une cause potentielle de biais. Il faut donc connaître les causes d'interruption de participation au protocole et comparer statistiquement les groupes selon ces causes.

Un problème connexe est que, dans un centre de recherche, tous les consultants ne sont pas forcément inclus dans le protocole de recherche en cours : certains seront inclus tandis que d'autres recevront un traitement hors protocole, peut-être ceux jugés les plus urgents ou correspondant le mieux à une thérapie connue, comme l'évoquent Perron et al. Cette inégalité de traitement devrait faire l'objet d'une description comparative, car susceptible d'introduire un biais. Malheureusement, il est extrêmement rare en médecine que les patients non inclus dans le protocole soient comparés aux inclus. Une fois le tirage au sort effectué, les constatations cliniques ne peuvent être généralisables qu'à des patients strictement similaires au groupe de patients retenus dans le protocole.

Question 13 : Dans les études, les sujets du groupe contrôle sont-ils évalués uniquement à l'inclusion, et n'est-ce pas donc nier la possibilité de changements psychologiques même sans traitement ?

L'assertion selon laquelle le groupe contrôle n'est évalué qu'une fois est inexacte, comme deux des auteurs de l'expertise ont pu me le confirmer.

Question 14 : les statistiques utilisées dans les études de niveau 1 et 2 (MA) sont-elles invalides (Student, d de Cohen) ?

Perron et al critiquent l'emploi du « t » de Student sans vérification des distributions dans les études de niveau 1, ce qui est justifié. Les chercheurs en psychologie devraient employer plus souvent des méthodes non-paramétriques, surtout lorsque l'effectif de patients est petit. Les

rapports d'étude ne mentionnent souvent que des valeurs de « t », ou de probabilité « p », les détails en termes de moyennes et les propriétés des distributions étant plus rarement fournis (Matt et Navarro, 1997).

Il s'agit donc d'une limite réelle de toute MA, et de l'expertise de l'INSERM, mais si l'on avait rejeté la littérature reposant sur des statistiques paramétriques inadéquates, la base de données aurait été considérablement appauvrie et de très nombreuses questions seraient restées sans réponse. De plus, lorsque les effectifs croissent, les distributions ont tendance à se rapprocher de la normalité.

Concernant la statistique « d », il en existe effectivement différents types, le plus souvent utilisé étant la statistique de Smith et Glass (1977), souvent mentionnée dans la littérature sous le nom d'« ES » (Effect Size ou taille d'effet), qui utilise la seule variance du groupe contrôle. Lorsque le terme « d de Cohen » est employé il s'agit en général d'une variance poolée des deux groupes. L'expertise de l'INSERM signale le type de statistique utilisée, en parlant soit du « d » de Glass ou du « d » de Cohen. L'utilisation d'un type de statistique à la place d'un autre ne résulte le plus souvent qu'en une différence minime d'ordre de grandeur (environ +/- 10%). Les lecteurs attirés par ces aspects techniques peuvent se reporter à l'ouvrage de Hedges et Olkin (1985).

Un problème qui aurait mérité clarification dans l'expertise et qui n'est pas évoqué par Perron et al, est celui de la combinaison des critères d'évaluation : dans une étude de niveau 1 il y a souvent plusieurs critères d'évaluation, donc plusieurs statistiques finales, lesquelles peuvent être combinées lors d'une MA pour produire des « ES ». Mais il peut y avoir plus de statistiques « ES » que d'études et cela pose problème : faut-il choisir une seule statistique et si oui, laquelle ? Faut-il faire la moyenne des « ES » ou des « d » ou les combiner d'une façon ou d'une autre ? Comment tenir compte des corrélations entre critères d'efficacité qui rompent l'assomption d'indépendance des critères ? Ces problèmes techniques sont traités dans un article de Rosenthal et Rubin (1986).

Enfin, il n'est pas exact que l'utilisation des différences de moyennes serve à mimer la science. Lorsqu'on constate que des patients s'améliorent de façon incomplète, l'utilisation d'une évaluation ordinale rend compte de cette réalité d'une façon plus précise que la simple appréciation « il va mieux ». La recherche de précision dans l'évaluation est tout à fait scientifique.

Question 15 : la différence de nombre d'études en faveur des TCC sur les TP produit-elle un artefact ou un biais en faveur des TCC ?

Plus on dispose de sujets dans une étude de niveau 1, ou d'études dans une MA, plus la puissance statistique (probabilité de mettre en évidence des différences entre des groupes) est élevée. Ceci n'est pas un problème mais un atout. Perron et al pensent qu'une différence de puissance statistique entre deux groupes d'études aboutit à un biais, ce qui est faux.

Dans cette expertise, il y a plus d'études pour les TCC que pour les TP dans certains troubles DSM et, quelles qu'en soient les causes, ceci n'entraîne pas que l'efficacité absolue (comparaison des thérapies versus contrôles en liste d'attente) des TCC soit plus grande que celle des TP. Les conditions expérimentales au sein de chaque étude étant équivalentes pour les deux groupes comparés (TP ou TCC versus liste d'attente), il n'y a théoriquement pas plus de biais que dans un étude comparative classique, et il faut rappeler que les comparaisons indirectes d'efficacité absolue sont sans valeur probante.

Un nombre de patients plus élevé permet seulement d'augmenter la précision de l'estimation de l'ES. En revanche, pour une petite taille d'effet, il faut plus de patients pour démontrer l'efficacité.

Pour en finir avec cette hypothèse de biais, il suffit d'examiner les études d'efficacité relative, c'est-à-dire les comparaisons directes entre TCC et TP, examinées à la question précédente, où le nombre d'études est le même pour les deux types de thérapies. Malheureusement, ces études de comparaison directe ne couvrent pas tout le champ des catégories cliniques DSM et n'envisagent pas les problèmes de comorbidité.

S'il n'y a donc pas de biais lié au nombre d'études, il y a un problème de présentation des résultats dans l'expertise INSERM, qui sont donnés en fonction du niveau de preuve, lequel n'est pas une statistique. Or, le plus grand nombre d'études d'efficacité absolue réalisées avec les TCC qu'avec les TP fait que les premières franchissent plus souvent la barre du niveau de preuve alors que les secondes la franchissent moins souvent.

Pour les différents troubles anxieux, certaines études de TP ont été réalisées avant le DSM-III, lequel a éclaté les anciennes névroses (surtout l'hystérie et la névrose d'angoisse) en diverses rubriques sur les axes I et II ; de plus le DSM-IV a introduit récemment la notion de phobie sociale. De ce fait, le manque apparent de preuve d'efficacité absolue pour les TP serait peut-être partiellement contrebalancé par les preuves obtenues pour les névroses, qui n'ont pas été sélectionnées dans le chapitre « synthèse » de l'expertise, même s'il y a incertitude sur le contenu du terme de « névroses » dans des études souvent anciennes, dont la méthodologie n'est peut-être pas aussi bonne que les études actuelles.

Des solutions pourraient être donc trouvées pour présenter en parallèle les résultats des TP et TCC dans les névroses et les troubles anxieux.

Ceci dit, on ne peut pas reprocher aux TCC leurs méthodes d'évaluation, logiques dans leur conception du trouble mental ni leurs succès, qui sont maintenant bien démontrés. Il revient en fait aux tenants de l'approche psychodynamique de procéder à plus d'études évaluatives, et si possible, en comparaison directe TP-TCC.

De même on ne peut reprocher aux auteurs de l'expertise d'avoir utilisé la notion de niveau de preuves, maintenant largement acceptée dans le monde, lorsque des décisions thérapeutiques sont à prendre.

Question 16 : les groupes contrôles avec placebo de psychothérapie sont-ils non éthiques ?

L'approche minimale ou le placebo de psychothérapie rentrent dans le cadre large des thérapies dites non-bona fide, c'est-à-dire qui ne reposent pas sur une théorie psychologique consistante et dont l'efficacité semble liée aux éléments non-spécifiques de toute thérapie. La plupart des thérapies bona fide ont fait preuve d'une efficacité contre l'abstention thérapeutique ou la liste d'attente, mais il est actuellement recommandé de tester leur efficacité contre des thérapies non bona-fide, afin de valider les éléments actifs de la méthode. Dans la MA inaugurale de Smith et Glass de 1977, les « thérapies placebo » obtenaient des résultats apparemment aussi satisfaisants que des thérapies adlériennes ou la Gestalt, et presque aussi bons que les TP, même s'il faut garder à l'esprit que ces comparaisons indirectes n'ont guère valeur de preuve.

L'usage de groupes contrôle reposant sur une thérapie non bona fide est donc utile pour tenter de réfuter « le verdict du dodo ». Est-il pour autant éthiquement justifié ? Il serait non-éthique, comme dans l'ensemble de la médecine, de diffuser largement une méthode thérapeutique sans s'être assuré qu'elle est efficace de façon absolue (versus abstention thérapeutique ou versus placebo) et/ou relative (c'est-à-dire qu'elle peut apporter un avantage par rapport à une autre méthode utilisée avant elle). Cette vérification d'efficacité doit reposer sur l'état des connaissances scientifiques du moment sans faire courir de risque majeur au patient.

Or ce risque est minimisé si l'on sait, par des résultats antérieurs, que la thérapie placebo donne des résultats assez satisfaisants, et si l'on prend des précautions visant à protéger le patient en cas d'aggravation clinique. En fait c'est souvent lors des essais thérapeutiques et études comparatives que l'on est le plus prudent avec le patient. En France, tous les protocoles d'intervention thérapeutique doivent être au préalable approuvés par un comité de protection des personnes dans la recherche biomédicale (CCPPRB). L'équivalent (souvent nommé

comité d'éthique) existe dans tous les pays industrialisés. Dans les essais thérapeutiques de médicaments, on ne prendra pas le risque d'inclure les patients présentant les pathologies les plus sévères dans les étapes initiales de démonstration d'efficacité. On revoit le patient très régulièrement, des possibilités d'interruption du protocole sont prévues afin de basculer sur une autre thérapeutique, et en cas de sortie du protocole, l'observation du patient rentrera dans les résultats en tant qu'échec du produit X selon le principe de l'intention de traiter.

Selon le vieux principe qui consiste d'abord à ne pas nuire, on ne peut qu'être en accord avec Perron et al pour condamner sans appel les anti-thérapies.

Le glissement de la méthodologie vers l'éthique s'opère chez Perron et al, dans une assimilation du registre de la recherche clinique avec celui du soin ordinaire, en supposant que le problème de l'efficacité des TP est d'ores et déjà réglé. Les résultats moyens observés pour les TP brèves à l'issue de cette expertise devraient inciter les praticiens et les chercheurs à plus de modestie : on supposait à l'avance que ces thérapies étaient efficaces et leurs résultats sont décevants du moins pour certaines des pathologies considérées.

Question 17 : Chercher à rendre convaincant le placebo de psychothérapie est-il éthique ?

Il est dans la nature même du placebo d'être convaincant et dès lors qu'on pense à instituer un groupe contrôle intitulé « placebo de psychothérapie », il doit être convaincant pour le patient. La croyance du thérapeute qui pratique une thérapie placebo concerne plus son rôle en tant qu'individu que la méthode utilisée, et cet exercice est assez différent de celui du médecin qui sait qu'une boîte de traitement sur les deux contient un placebo, mais ne sait pas laquelle.

Dans les thérapies placebo, le thérapeute sait qu'il se prive des ingrédients actifs de la thérapie reposant sur la théorie, mais il lui reste les ingrédients non-spécifiques comme le contact humain chaleureux et bienveillant avec le patient, l'écoute voire la facilitation de la parole par des méthodes rogéiennes. Le thérapeute qui a accepté de participer à une étude en faisant une thérapie non bona fide sait qu'il aura quand même une certaine action thérapeutique.

Question 18 : le tirage au sort d'un traitement psychothérapeutique est-il pire sur le plan éthique que ce qui se pratique en médecine ?

On ne voit pas bien pourquoi les règles éthiques qui ont été élaborées dans le domaine médical et qui concernent la totalité des essais thérapeutiques de médicaments ne pourraient pas s'appliquer également aux psychothérapies. Or, la randomisation ou tirage au sort ne pose pas de problème particulier aux CCPPRB dès lors qu'il s'inscrit dans une démarche logique visant à une démonstration d'efficacité et que les conditions de la protection du patient sont

envisagées et décrites. A priori, si l'on suppose qu'un nouveau traitement va faire du bien au patient, il arrive soit qu'il n'apporte aucun bénéfice par rapport à l'abstention thérapeutique ou aux traitements pratiqués auparavant, soit qu'il soit plus nocif que les conditions contrôles. Il est utile de rappeler que de nombreux essais thérapeutiques en médecine comportent comme critère de jugement des variables lourdes comme la durée de survie, et que le risque de décès d'un patient au cours d'un essai n'empêche absolument pas que le traitement expérimenté soit proposé sur tirage au sort.

En fait, la méprise de Perron et al vient de ce qu'ils assimilent les essais thérapeutiques à un soin ordinaire, et qu'ils pensent que l'efficacité des psychothérapies est donnée par avance. Or ce n'est pas le cas dans un essai thérapeutique qui doit justement répondre à cette question d'efficacité.

Concernant le terme de pseudo-traitement à propos du placebo il faut énergiquement le rejeter dans la mesure où les taux d'amélioration qu'on observe en l'utilisant sont souvent élevés. On pourrait qualifier le placebo de « traitement ne comportant pas de produit biologiquement ou physiologiquement actif ».

Conclusion

Perron et al font dès le départ une confusion sur la nature de l'expertise INSERM en qualifiant ce travail de méta-analyse de méta-analyses. Plusieurs parmi leurs critiques se sont avérées fausses ou infondées, comme l'assertion d'un cumul automatique de biais dans les méta-analyses, ou celle concernant l'existence d'un biais qui serait lié au nombre d'études, où la non prise en compte de l'évolution des groupes contrôle.

L'obstacle épistémologique évoqué par eux à propos de l'évaluation quantifiée des phénomènes psychiques ne repose probablement que sur une vision restrictive de la mesure et du champ de la psychiatrie.

L'avantage supposé procuré par les instruments d'évaluation symptomatiques en faveur des TCC n'est pas retrouvé dans les données disponibles : l'examen des études ou MA de comparaisons directes TP-TCC ne montre pas de supériorité écrasante des TCC sur les TP, donc peu d'argument pour un biais favorisant les TCC.

Perron et al ont des positions concernant les biais qui sont contradictoires par rapport à leur parti-pris méthodologique de départ : certains biais les gênent, mais si on les suivait en éliminant certaines règles méthodologiques (étude comparative, tirage au sort) on aboutirait certainement à d'autres biais plus graves. D'autres biais fréquents en MA leur échappent, comme les biais de publication et de sélection.

Il semble donc que les critiques de Perron et al prennent leur origine dans une série de préjugés contre la méthodologie scientifique et son application à l'évaluation des psychothérapies. Ces préjugés apparaissent lorsque les auteurs dérivent vers l'éthique, qui leur permet de justifier, selon les conceptions qu'ils s'en font, le rejet des groupes contrôle, de la randomisation et du placebo. Or, cette conception éthique est très largement plus stricte que celle qui prévaut dans les essais thérapeutiques en médecine dans tous les pays développés.

L'expertise INSERM en elle-même, du fait de son statut de revue narrative et des choix méthodologiques, peut être critiquée. La méthodologie et la présentation des résultats sont denses et complexes et il est parfois difficile de s'y retrouver. De plus il y a beaucoup de comparaisons indirectes d'efficacité absolue et de nombreux résultats figurant dans les tableaux portent sur des critères d'évaluation qui semblent secondaires plutôt que principaux. Les phénomènes de comorbidité sont mal appréhendés dans les études de niveau 1 et la généralisabilité des conclusions tirées d'études sur troubles purs apparaît faible.

L'utilisation du niveau de preuve, l'accent mis sur l'efficacité absolue plutôt que relative et l'éclatement des névroses classiques en catégories de troubles DSM disjoints peut renforcer l'impression de supériorité des TCC sur les TP.

Un fait, rendu apparent au travers de l'expertise, ne souffre cependant pas de discussion : le manque d'études de comparaison directe TCC-TP pour certaines catégories cliniques est patent et il reste aux psychodynamiciens un important travail d'évaluation à réaliser, tant au niveau des études comparatives que des études naturalistes, indispensables pour mieux apprécier la généralisabilité des résultats obtenus dans les études.

Plus pragmatiquement, les conclusions de l'expertise ne sont pas réellement surprenantes pour les psychiatres : on savait déjà que dans la schizophrénie et l'anorexie mentale des approches multimodales fonctionnaient mieux que les TP isolées. On savait également que certains syndromes névrotiques graves échappent aux TP en raison de phénomènes divers (chronicité et enkystement du symptôme, bénéfices secondaires, etc.). Dans les cas de dépressions sévères, peu de psychanalystes de formation médicale entreprendraient une analyse sans que le patient ne prenne également un traitement antidépresseur. Des échanges de patients entre certains psychanalystes et certains thérapeutes cognitivo-comportementaux existent depuis longtemps, sans reposer sur des expertises autres que l'expérience des praticiens concernées.

En revanche, dans les troubles de la personnalité, les dépressions légères, les troubles anxieux et bon nombre d'autres conditions, les TP sont encore largement pratiquées, même si le traitement de ces états n'est pas toujours purement psychothérapique en raison des comorbidités. Or, l'efficacité des TP brèves telles qu'évaluées par l'expertise semble ne pas être au niveau qu'on pouvait espérer. Faut-il pour autant changer de modèle étiologique et

basculer vers le modèle cognitivo-comportemental ? On fera remarquer que seules les thérapies brèves ont été évaluées et que rien n'invalide ni ne valide a priori les TP plus longues, et que d'autre part le nombre de thérapeutes cognitivo-comportementalistes est trop faible. De plus, la faible différence observée entre TCC et TP, même si elle semble indiquer une supériorité des TCC dans certains troubles, n'est pas suffisante pour justifier une condamnation sans appel des TP, d'autant que les phénomènes d'allégeance à la méthode, l'importance de l'effet thérapeute (peut-être supérieur à l'effet méthode) semblent de nature à compliquer les données d'évaluation.

RÉFÉRENCES

1. AMERICAN PSYCHIATRIC ASSOCIATION. DSM-IV-Text Revision : manuel diagnostique et statistique des troubles mentaux ; trad. française coordonnée par GUELFY JD et CROCQ. MA - Paris : Masson, 2003.
2. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, Girolamo G, Graaf R, Demyttenaere K, Gasquet I, Haro JM, Katz SJ, Kessler RC, Kovess V, Lepine JP, Ormel J, Polidori G, Russo LJ, Vilagut G, Almansa J, Arbabzadeh-Bouchez S, Autonell J, Bernal M, Buist-Bouwman MA, Codony M, Domingo-Salvany A, Ferrer M, Joo SS, Martinez-Alonso M, Matschinger H, Mazzi F, Morgan Z, Morosini P, Palacin C, Romera B, Taub N, Vollebergh WA; ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. 12-Month comorbidity patterns and associated factors in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand Suppl.* 2004;(420):28-37.
3. ANDREWS G, HARVEY R : Does psychotherapy benefit neurotic patients ? *Arch Gen Psychiatry* 1981 38 (11) 1203-1208
4. BECH P, CIALDELLA P. Citalopram in depression--meta-analysis of intended and unintended effects. *Int Clin Psychopharmacol.* 1992 Jun;6 Suppl 5:45-54.
5. BECH P, CIALDELLA P, HAUGH MC, BIRKETT MA, HOURS A, BOISSEL JP, TOLLEFSON GD. : Meta-analysis of randomised controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *Br J Psychiatry.* 2000 May;176:421-8.
6. BECH P, TANGHOJ P, CIALDELLA P, ANDERSEN HF, PEDERSEN AG. Escitalopram dose-response revisited: an alternative psychometric approach to evaluate clinical effects of escitalopram compared to citalopram and placebo in patients with major depression. *Int J Neuropsychopharmacol.* 2004;7(3):283-90
7. BERGERET J, ACHAINTE A, BECACHE A, BOULANGER JJ, CHARTIER JP, DUBOR P, HOUSER M, LUSTIN JJ : *Psychologie pathologique.* Masson, 1979. Paris.
8. BOUVARD M, COTTRAUX J. *Protocoles et échelles d'évaluation en psychiatrie et en psychologie.* Masson, Paris 1996, troisième édition 2002
9. CIALDELLA P., CHIGNON JM : comorbidité psychiatrique et nosologie. In Lempérière T (Ed) *Dépression et comorbidités psychiatriques.* Collection PRID. Masson ed, Paris, pp 3-26. 2001

10. CUCHERAT M, CIALDELLA P. La méta-analyse des essais thérapeutiques : application en psychiatrie. *Encephale* 1996, XXII : 378-387
11. EYSENCK H. : The effects of psychotherapy : an evaluation. *Journal of Consulting Psychology* 1952, 16:319-324
12. GLASS GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976, 5 : 3-9
13. GLASS GV : Meta-Analysis at 25 2000, papier publié sur le site <http://glass.ed.asu.edu/gene/papers/meta25.html>
14. GUELFY JD. Psychiatres et psychanalystes d'aujourd'hui. A paraître in *Topiques*. 2004 ; 88
15. HEDGES LV, OLKIN I. *Statistical methods for meta-analysis*. Academic Press, New York 1985
16. KÄCHELE H : Analyse du discours du patient et du thérapeute par des analyses du contenu informatisées. In GERIN P ET DAZORD A (eds) *Recherches cliniques planifiées sur les psychothérapies*. INSERM (Ed) Paris. 1992
17. INSERM : *Psychothérapie : Trois approches évaluées*. Expertise Collective (O. CANCEIL, J. COTTRAUX, B FALISSARD, M. FLAMENT, J. MIERMONT, J. SWENDSEN, M. TEHERANI, J.M THURIN), Inserm, 2004, 553 pages. Résumé en ligne www.inserm.fr
18. LUBORSKY L., SINGER B., LUBORSKY L. : Comparative studies of psychotherapies : is it true that « everyone has won and all must have prizes ? » *Archives of General Psychiatry*, 1975, 32:995-1008
19. LUBORSKY L, DIGUER L, SELIGMAN DA et al. The researcher's own therapeutic allegiances : "a wild card" in comparisons of treatment efficacy. *Clin Psychol Sci Prac* 1999 ; 6 : 95-132
20. LUBORSKY L, ROSENTHAL R, DIGUER L et al. The dodo bird verdict is alive and well – mostly. *Clin Psychol Sci Prac* 2002 ; 9: 2-12
21. MATT GE, NAVARRO AM. : What meta-analyses have and have not taught us about psychotherapy effects: a review and future directions. *Clin Psychol Rev*. 1997;17(1):1-32
22. PERRON R, BRUSSET B, BARUCH B, CUPA D, EMMANUELLI M. Quelques remarques méthodologiques à propos du Rapport INSERM " Psychothérapie. Trois approches évaluées ". Texte publié en ligne en 2004 sur le site www.techniques-psychotherapies.org
23. ROSENTHAL R RUBIN DB : Meta-analytic procedures for combining studies with multiple effect sizes. *Psychol Bull* 1986, 99,3 400-406
24. SACKETT DL. Bias in analytic research. *Journal of Chronic Diseases* 1979, 32:51-63
25. SMITH M.L., GLASS G.V.. Meta-analysis of psychotherapy outcome studies. *American Psychologist* 1977, 32:752-760

26. STEVENS SS. On the theory of scales of measurement. Science 1946 ; 103 : 677-680
27. SVARTBERG M, STILES TC. Comparative effects of short-term psychodynamic psychotherapy: a meta-analysis. J Consult Clin Psychol. 1991 ; 59(5) :704-14
28. VAN PRAAG HM. Nosologomania: a disorder of psychiatry. World J Biol Psychiatry 2000 ; 1(3) : 151-8
29. WAMPOLD B.E., MONDIN G.W., MOODY M., STICH F., BENSON K., AHN H-N. .A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, « all must have prizes » Psychological Bulletin 1997, 122,3:203-215

Annexes : Citations extraites

1. «... Encore faut-il qu'une telle évaluation soit objectivement conduite, et que, prétendant établir la valeur scientifique des thérapies examinées, elle soit elle-même scientifique. Tel n'est pas le cas. Ce Rapport scandalise par ses méthodes, de toute évidence biaisées à chaque étape de la démarche d'évaluation par des préjugés bien peu scientifiques. Tout y semble fait pour déboucher sur les conclusions souhaitées au départ...»
2. « Dans le Rapport, ce terme [psychothérapie] est réduit à des démarches médicales ou paramédicales visant à guérir des maladies (en ce cas des maladies psychiques). Les professionnels concernés ne peuvent accepter une définition aussi restrictive qui désigne par principe les personnes dont ils s'occupent comme des « malades » ».
3. «.. [psychothérapies] conduites selon des procédures parfois fort éloignées de la cure psychanalytique proprement dite, même si leurs promoteurs se disent psychanalystes (assertion que les auteurs du Rapport ne contrôlent ni ne peuvent contrôler en aucune façon). »
4. «Ainsi, on trouve, p. 19, à propos d'une étude de Luborsky et coll. (1985), ceci : « Les prédicteurs 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 ». Cette indication, « suivre consciencieusement le Manuel », reviendra très souvent dans le texte de ce Rapport à titre de critère du sérieux d'une étude de validation, jugée bonne s'il y est dit que les thérapeutes procédaient ainsi.»
5. « On dira qu'il s'agit d'une « méta-analyse » s'il s'y ajoute un effort de systématisation et d'évaluation des recherches ainsi colligées. Mais il faut bien voir qu'on gagne alors en étendue ce qu'on perd en précision ».
6. « En effet, si cent recherches aboutissent à des résultats de même type, ce peut être pour deux ordres de raisons bien différents : ou bien parce qu'alors le réel répond de la même façon à tous ces sondages... ou bien parce que toutes ces recherches sont biaisées de la même façon en vertu de présupposés communs à tous les chercheurs. L'histoire des sciences abonde en exemples.»
7. « En analysant ainsi un ensemble d'ensembles, on est très loin de la réalité de chacune des recherches de base et de ses conditions réelles de réalisation. Le lecteur de ce Rapport se trouve alors, en fait, hors d'état d'évaluer la pollution de l'ensemble par des recherches de niveau 1 éventuellement biaisées pour des raisons qui auraient échappé à leurs auteurs, mais aussi aux experts en fonction de leurs propres présupposés. »

8. « Après un temps de traitement jugé suffisant, on compare l'état de ces divers groupes. Soit... mais cela suppose que, au départ, ces groupes soient comparables. Or rien n'est plus difficile à garantir ».
9. « La principale difficulté réside en fait dans la caractérisation des sujets en cause, avant et après traitement. Ceci introduit un présupposé majeur. En effet, la plupart de ces travaux utilisent des « grilles » qui repèrent des symptômes allégués par le sujet et notés par le consultant ; l'instrument le plus souvent utilisé est le DSM-IV, élaboré par l'Association Américaine de Psychiatrie. Le Rapport INSERM remarque que les critères de ce type «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» (p.22). Il est exact que ce type d'évaluation des troubles, où l'on s'attache uniquement aux symptômes, a suscité en France de très vives objections ; mais écarter ces objections en déclarant qu'il s'agit « pourtant » de la « meilleure » démarche possible constitue la simple expression d'une opinion, qui n'a rien de scientifique....»
10. « En fait, cette option, en apparence simplement technique, suppose une prise de parti sur une question scientifique majeure. Il existe en ce domaine une divergence épistémologique, méthodologique et technique, peut-être irréductible, entre deux types de démarches possibles pour étudier les troubles et les difficultés psychiques. La première position est celle d'une démarche nosographique qui décrit des symptômes regroupés en syndromes, pour distinguer et classer des maladies (sur le modèle taxinomique, classificatoire, de la botanique et de la zoologie) ; une tout autre position (sur le modèle de la recherche en physiologie depuis Claude Bernard) décrit des structures fonctionnelles et en analyse les troubles. Or il est patent que si les comptages de fréquences, et de façon générale les traitements statistiques, sont pertinents dans la première optique, ils ne peuvent jouer qu'un rôle secondaire, et difficilement de preuve, dans la seconde. » Et, plus loin « En effet, les thérapies «cognitivo-comportementales», en définissant les cas à traiter par des symptômes puis en traitant ces symptômes, se situent explicitement dans la première optique, taxinomique ; au contraire, les approches psychodynamiques – ou, plus particulièrement et précisément, l'approche psychanalytique – se situent dans la seconde, celle de l'analyse fonctionnelle. Ainsi, d'emblée le choix d'instruments qui caractérisent les troubles à traiter dans l'optique taxinomique biaise massivement toute recherche en faveur des premières, au détriment des secondes.»

11. « Comment évaluer l'effet d'une psychothérapie ? La solution simple paraît évidemment, en s'inspirant des études pharmacologiques (selon un modèle qui pèse lourdement sur tout ceci), de comparer l'état des patients « avant » et « après » : on applique ces grilles de repérage des symptômes avant le traitement, puis après, et l'on évalue la différence »
12. « Or, supposer ces deux échantillons comparables, c'est ignorer les multiples facteurs qui risquent de les rendre hétérogènes : priorité de traitement légitimement donnée à certains types de cas, jugés plus urgents ou plus accessibles au traitement proposé ; poids des possibilités de prise en charge par l'institution ou / et le thérapeute ; fonte de l'un et l'autre échantillons en cours d'étude (ruptures de traitement pour le groupe traité, lassitude et disparition des personnes laissées en liste d'attente), etc. Tout ceci peut jouer, avoir joué, plus ou moins à l'insu du chercheur, dans la réalité d'un fonctionnement institutionnel. En fait, il est extrêmement difficile de garantir la comparabilité de deux groupes, et le maintien de cette comparabilité tout au long de l'étude. En fait, si fréquentes sont les taches aveugles que rares sont les publications qui s'attachent réellement à éclairer le lecteur sur ces points pourtant majeurs. Qu'on n'objecte pas que tout ceci ne peut jouer lorsqu'on tire au hasard la répartition des sujets en groupe traité et groupe contrôle, non traité.. »
13. « La rigueur méthodologique voudrait qu'on évalue l'état des personnes en cause deux fois (avant - après) aussi bien pour le groupe contrôle (B) que pour le groupe traité (A). En fait, la plupart des recherches recensées ne procèdent pas ainsi : on se contente d'évaluer le groupe contrôle une seule fois, supposant que, en l'absence de traitement, les personnes en cause ne changent en aucune façon...»
14. « Dans la majorité de ces études, il s'agit ... le plus souvent, parce que cela semble plus “ scientifique ”, d'une variation de moyennes. Rappelons qu'il existe une technique paramétrique tout à fait classique (la statistique “ t de Student ”) pour estimer si une telle différence entre les moyennes de deux séries d'observations est statistiquement significative ; cela suppose qu'on tienne compte de la variabilité (plus exactement la “ variance ”) des deux échantillons. Il faut se rappeler que, en principe, cette statistique n'est utilisable que si les deux distributions sont « normales » (gaussiennes), ce qui est rarement vrai dans ce domaine où sont fréquentes des distributions fortement asymétriques, et plus rarement encore considéré par les auteurs de ces travaux. Passons, car il y a plus inquiétant. En effet, nous dit-on, dans beaucoup des études recensées la différence des moyennes est rapportée à la variance du seul groupe de contrôle ; l'aléatoire ainsi introduit dans la démarche est considérable.

D'ailleurs, deux pages plus loin (p. 26), il nous est dit qu'il " existe de très nombreuses versions de ce " d ", sans véritable consensus sur son utilisation ". Alors, que vaut la "preuve" statistique abstraite d'un ensemble d'études ainsi déclarées sans cohérence quant à l'instrument même de cette preuve ? ».

15. « Il suffit d'augmenter la taille d'un échantillon pour qu'une faible différence de moyennes, si elle se maintient, devienne très significative (en termes statistiques). Cela découle simplement de la procédure de calcul et ne recouvre aucun mystère, ainsi que le savent tous les étudiants débutants en statistiques. Les auteurs du Rapport le savent eux aussi, et le disent (p.20). Mais ensuite ils n'en tiennent pas compte, laissant ainsi s'introduire un biais qui fait planer sur tout leur travail un doute très sérieux. Il y a là en effet la source d'un artefact majeur : pour ces raisons purement statistiques, si trois études d'efficacité font apparaître une légère amélioration de l'état des patients, mais à un seuil statistiquement non significatif, la même légère amélioration moyenne, établie sur vingt études, deviendra hautement significative. Ainsi, plus une psychothérapie d'un certain type sera représentée dans un recensement d'études d'efficacité, plus elle aura de chances d'être déclarée efficace... »
16. « Cependant, ne craignant pas la contradiction, il déclare quelques lignes plus loin (p. 18): «plusieurs solutions ont été proposées pour résoudre le problème du placebo en psychothérapie, en particulier : 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, et le contrôle par une pseudo-thérapie ou une anti-thérapie»... Ceci laisse le lecteur effaré. Comment des gens, s'il s'en trouve, qui s'appliquent ainsi à ne pas soigner peuvent-ils se croire et se dire «psychothérapeutes»? Comment peut-on mentir de façon délibérée à des gens qui souffrent et demandent de l'aide ? De telles aberrations éthiques et déontologiques sont simplement considérées dans ce texte comme pouvant «générer des problèmes éthiques et pratiques». Admirable sens de la litote ! ».
17. « Et que dire des recherches évoquées quelques lignes plus loin, où l'on décide par tirage au sort du traitement à " appliquer " à un consultant ? Car, nous dit-on (p. 18 encore), « 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 ». Qui méprise-t-on alors le plus, le patient ainsi tiré au sort, le «psychothérapeute » qui s'attache à ne pas le

soigner et cependant « y croit » (mais il croit à quoi ?), ou la crédulité de l'un et de l'autre ? Et peut-on fonder une démarche qui se prétend scientifique sur le partage d'une croyance ? ».

18. « On ne peut manquer de remarquer que tout cela revient à nier l'idée même d'un choix de la thérapeutique selon ses indications spécifiques et en fonction des troubles à prendre en charge. Imagine-t-on que dans le cas de troubles cardiaques graves, d'un cancer curable, etc., on tire au sort l'affectation du malade à un traitement espéré efficace ou à un pseudo-traitement ?»