

1

Réflexion sur l'évaluation

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

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

« Expériences » dans un cadre psychothérapeutique

Ce point concerne en fait la recherche thérapeutique en général et pas seulement le domaine des psychothérapies. Ainsi, dans un essai thérapeutique comparant l'efficacité d'un antiulcéreux à un placebo à l'aide d'une méthodologie de « double aveugle » et d'allocation aléatoire du traitement, il

Il y a effectivement une modification du contexte de soin par rapport à la situation clinique habituelle : ni le médecin ni le malade ne savent quel traitement est administré. Rien ne dit alors que la différence d'efficacité constatée en situation expérimentale est véritablement superposable à la différence d'efficacité existant en pratique clinique quotidienne. Certains champs de la clinique sont plus sensibles que d'autres à ce biais potentiel, la psychiatrie en est sûrement un, mais elle n'est pas la seule. Les autorités de santé sont conscientes de ce problème et suggèrent, pour apporter des éléments de réponse à cette question, la réalisation d'études observationnelles de terrain, il est vrai moins rigoureuses et plus difficilement interprétables, mais dont les résultats permettent de corroborer utilement les données issues des essais thérapeutiques classiques.

L'évaluation des psychothérapies relève de la même logique : l'essai thérapeutique comparatif apporte une information capitale du fait de la rigueur et de la transparence de la méthodologie mise en jeu. Cette information ne se suffit néanmoins pas à elle-même, elle doit toujours être interprétée à la lumière des spécificités de la pratique clinique réelle et pourra éventuellement gagner à être étayée par des études complémentaires, qualitatives ou épidémiologiques.

Mesures objectives, mesures subjectives

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

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

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

avantages et ses inconvénients : il permet d'une part de bénéficier des formidables outils que les mathématiciens ont développé au cours des temps pour manipuler les nombres, il nous fait perdre d'autre part un peu de cette intimité que le langage naturel nous procure à partir de l'évocation du mot « tristesse ».

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

À propos des différences entre mesures subjectives et mesures objectives, certains pourraient remarquer qu'un attribut subjectif ne correspond à rien de concret, tout juste à un mot, qui plus est au sens généralement imprécis. Prenons la tristesse. Il est vrai que, curieusement, il est difficile d'en obtenir une définition¹, mais est-ce pour cela que le sens de ce mot est imprécis ? Bien au contraire. La tristesse est un sentiment élémentaire que tout être humain a déjà éprouvé. Il suffit, en fait, que je dise : « je suis triste... » pour que n'importe quelle personne sache exactement ce que je ressens. Pour le moins, on ne peut voir ici aucune imprécision. À l'opposé, le temps est une variable objective que chacun mesure quotidiennement. Mais comment définir le temps ? Philosophes ou physiciens ont en pratique bien des difficultés pour répondre à cette question.

À propos des ressemblances entre mesures subjectives et mesures objectives, on remarquera une évolution parallèle des paradigmes² de mesure dans les

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

2. Dans son sens courant, le terme paradigme désigne un « mot-type » qui est donné comme modèle pour une déclinaison, une conjugaison (le paradigme de la conjugaison des verbes du premier groupe est le verbe « chanter ») ; il est, en ce sens, synonyme de modèle ou d'exemple. En épistémologie, il a été utilisé par Kuhn pour désigner les « accomplissements passés pouvant servir d'exemples et remplacer les règles explicitées, en tant que bases de

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

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

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

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

Pour conclure sur ce point, si nous venons de voir que les différences entre les mesures subjectives et les mesures objectives ne sont pas aussi évidentes que l'on pouvait le penser a priori, nous avons aussi constaté que les instruments de mesure utilisés le plus fréquemment en psychopathologie (les instruments

solution pour les énigmes qui subsistent dans la science normale » ; par exemple, l'expérience dite du labyrinthe en « T » est un paradigme de la théorie du conditionnement opérant de Skinner, l'auscultation est le paradigme de l'examen clinique tel qu'on le pratique depuis Laennec. C'est en ce sens que ce mot est utilisé ici.

constitués d'une liste d'items) dépendaient fortement d'un système théorique. Or d'une part ces systèmes théoriques ne sont pas toujours bien explicités, mais d'autre part ils peuvent ne pas être congruents avec les théories psychopathologiques sous-jacentes aux psychothérapies évaluées. Ainsi, il n'est pas rare de voir des thérapies d'inspiration psychanalytique évaluées à l'aide d'instruments développés dans un cadre purement symptomatique. Les résultats de telles études sont donc nécessairement à interpréter avec prudence, même si certains pourraient alors argumenter que l'objectif n'est pas tant d'évaluer la psychothérapie en tant que telle, mais plutôt ses effets (son efficacité). Les instruments développés pour étudier les changements symptomatiques seraient alors adaptés aux examens de toute forme de psychothérapie, à la condition que le contexte de cette comparaison soit bien admis.

Reproductibilité statistique

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

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

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

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

La réponse à cette question est de nouveau affirmative. En effet, dans un essai thérapeutique avec allocation aléatoire des traitements, la statistique ne fait que chiffrer la part possiblement attribuable au hasard dans la différence d'efficacité observée entre les traitements comparés. Plus précisément, chaque patient a, sans que l'on puisse généralement le savoir a priori, une aptitude particulière à répondre à un traitement donné. Il est possible que le tirage au sort attribue, par le simple fait du hasard, un plus grand nombre de

fois un traitement particulier à des patients au potentiel de guérison le plus élevé. Les tests statistiques évaluent dans quelle mesure ce phénomène aléatoire pourrait aboutir à lui seul à la différence d'efficacité observée entre les traitements. Ce calcul se fait de la même façon que le traitement soit médicamenteux ou psychothérapeutique.

La reproductibilité statistique des résultats issus des essais thérapeutiques en psychiatrie repose en grande partie sur l'adoption de critères diagnostiques comme ceux du DSM ou CIM-10. Ces critères obéissent cependant à une logique interne qui n'est pas sans conséquences sur les interprétations que l'on peut réaliser de ces résultats.

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

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

8 E , nous venons de voir qu'il n'existe pas de frein conceptuel évident à l'évaluation scientifique de l'efficacité d'une psychothérapie. Il est

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