

Déprimer par les nombres

Le façonnage statistique et psychométrique des dépressions contemporaines

Depress by Numbers

Statistical and Psychometric Shaping of Contemporary Depressions

Xavier Briffault et Olivier Martin

Volume 43, numéro 2, automne 2011

La statistique en action
Statistics in Action

URI : <https://id.erudit.org/iderudit/1008239ar>
DOI : <https://doi.org/10.7202/1008239ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0038-030X (imprimé)
1492-1375 (numérique)

[Découvrir la revue](#)

Citer cet article

Briffault, X. & Martin, O. (2011). Déprimer par les nombres : le façonnage statistique et psychométrique des dépressions contemporaines. *Sociologie et sociétés*, 43(2), 67–89. <https://doi.org/10.7202/1008239ar>

Résumé de l'article

Cet article se propose d'examiner le statut et le rôle d'une échelle quantitative d'évaluation des effets thérapeutiques des antidépresseurs. Cette échelle, dite HAM-D (échelle de Hamilton), joue un rôle central dans les pratiques médicales, les normes thérapeutiques, les politiques de santé et finalement les conceptions de la dépression. Tout en examinant les conditions sociales et scientifiques expliquant la centralité et l'importance de cette échelle, l'article détaille les conséquences des usages de cette échelle. L'échelle façonne la notion de dépression en proposant une représentation unidimensionnelle, une définition quasi opérationniste et performative. En numérisant la dépression, elle offre aux acteurs du marché des antidépresseurs la possibilité d'une « mise en scène du nombre » qui contribue largement à ce que les patients traités par ces molécules soient bien plus nombreux que ce qu'autoriseraient logiquement les caractéristiques de ceux sur lesquels la (toute relative) efficacité clinique des antidépresseurs a été montrée.



Déprimer par les nombres

Le façonnage statistique et psychométrique
des dépressions contemporaines

XAVIER BRIFFAULT

CERMES 3
Université Paris Descartes
Sorbonne Paris Cité
45, rue des Saints-Pères
75006 Paris
Courriel : xavier.briffault@parisdescartes.fr

OLIVIER MARTIN

CERLIS
Université Paris Descartes
Sorbonne Paris Cité
45, rue des Saints-Pères
75006 Paris
Courriel : olivier.martin@parisdescartes.fr

Dans aucun [] cas il n'est envisageable de rouvrir la boîte et d'en regarder le contenu, et ceci non par machiavélisme ou souci de tromper, mais simplement par économie. La fermeture de la boîte a eu un coût. Elle a été un investissement, dont on peut espérer tirer les fruits de façon machinale, comme on le ferait pour d'autres outils. Défaire une statistique est au moins aussi coûteux que de la faire.

Alain Desrosières (2008 : 93)

OBSERVATION DES DOCUMENTS SCIENTIFIQUES COMMUNIQUÉS PAR LES FIRMES

Considérons un document intitulé « Valdoxan : dossier scientifique » de 81 pages au format A4 sur papier glacé, distribué sur le stand d'une société pharmaceutique lors du Congrès de l'Encéphale, le plus important congrès de psychiatrie français, en janvier 2011. Ce document concerne une molécule à visée antidépressive récemment commercialisée (2009) par la société Euthérapie (Groupe Servier) : l'Agomélatine (Valdoxan®).

Ce « dossier scientifique » contient une section d'une vingtaine de pages intitulée « Efficacité clinique » (p. 29-47). Cette section expose six études d'efficacité de l'Agomélatine publiées dans des revues « classiques » dont les « *impact factors* » sont de très bon niveau (l'*International Journal of Neuropsychopharmacology*, l'*International Clinical*

Psychopharmacology, ou encore le *Journal of Clinical Psychiatry*) et par des auteurs occupant des positions situées au sommet de la hiérarchie de la psychiatrie française (Goodwin *et al.*, 2009; Hale *et al.*, 2010; Kennedy et Emsley, 2006; Lemoine *et al.*, 2007; Loo *et al.*, 2002; Olie et Kasper, 2007). Chacune de ces études est présentée sur une à trois pages.

Chaque présentation est introduite par un encart mis en évidence par un code couleur différent ainsi qu'un texte en gras et une police de taille supérieure au reste du texte, intitulé « Les points à retenir ». Par exemple: « Valdoxan a démontré son efficacité antidépressive versus placebo sur l'échelle HAM-D chez les patients souffrant d'épisode dépressif majeur » (pour la présentation de Kennedy et Emsley, 2006; Loo *et al.*, 2002; Olie et Kasper, 2007); « cette étude a montré une efficacité statistiquement supérieure (différence de 1, 49 point sur le score HAM-D) de Valdoxan par rapport à la fluoxétine à 8 semaines sans différence significative en termes de tolérance » (pour la présentation de Hale *et al.*, 2010); « la mesure de l'efficacité sur le score HAM-D (critère secondaire) après 6 semaines de traitement a montré une différence en faveur de Valdoxan par rapport à la sertraline de 1, 68 point » (p. 43); « L'efficacité a également été observée chez des patients présentant une dépression plus sévère (score HAM-D initial > 24) dans toutes les études positives contrôlées vs placebo ».

L'observation de ces documents permet d'établir un premier constat: l'échelle de Hamilton, identifiée par le sigle HAM-D, est l'échelle admise pour mesurer l'efficacité de traitements de la dépression. En l'occurrence, elle est mentionnée dans 5 des 6 sections « Points à retenir ». Comme nous le montrerons, ce constat dépasse largement ces documents consacrés au Valdoxan: l'échelle de Hamilton est l'échelle quasi universelle de mesure de l'efficacité thérapeutique des traitements de la dépression.

Le second constat est que la mesure de cette efficacité se résume à une évaluation quantitative unique: l'échelle est unidimensionnelle. L'hypothèse sous-jacente est que les traitements ne peuvent avoir que des effets unidimensionnels; seule leur intensité varie d'un traitement à l'autre. L'efficacité d'un traitement peut être plus ou moins grande mais est toujours de même nature. Nous retrouvons ici un processus déjà mis en évidence dans d'autres contextes (Martin, 1997): le développement d'outils de mesures quantitatives conduit implicitement à des définitions unidimensionnelles des grandeurs à mesurer.

Enfin, l'examen de ce dossier permet d'établir un troisième constat: après avoir été quantifiés, les effets thérapeutiques du traitement sont présentés sans faire intervenir de chiffres ou d'évaluation quantitative précise. Les expressions et les précisions numériques des résultats sont oubliées dans les synthèses, au profit d'énoncés généraux. Ainsi, alors que le corps du texte (section « Critères d'inclusion ») indique que seuls les patients présentant un score HAM-D supérieur à 22 (25 pour Hale *et al.*, 2010) ont été inclus dans les études, l'encart « Points à retenir » mentionne l'efficacité du traitement de manière générale en parlant de son efficacité sur les « patients souffrant d'épisode dépressif majeur ». Alors que la taille d'effet de la molécule par rapport au placebo et par rapport à d'autres molécules, et son évolution au cours du temps,

sont bien présentées graphiquement dans le « Dossier scientifique », aucune mention de cette taille d'effet n'est faite dans le dépliant en trois volets « Résumé des caractéristiques du produit ». Dans la « Notice : information de l'utilisateur », ni la taille d'effet ni le niveau de sévérité à partir duquel la molécule a montré la taille d'effet retenue comme significative ne sont précisés : « Valdoxan [...] vous a été prescrit pour traiter votre dépression [...]. Valdoxan commence à agir sur les symptômes de la dépression dans les deux premières semaines de traitement ».

L'ÉCHELLE DE HAMILTON : UN OUTIL INCONTOURNABLE

Au-delà du cas particulier de cet antidépresseur, il apparaît que l'échelle d'Hamilton est l'échelle quasi unanimement utilisée pour mesurer les bénéfices des thérapies sur les dépressions. Elle a donc une importance critique dans le façonnage de la dépression et de sa prise en charge. Comme toute autre mesure quantitative d'une variable d'intérêt clinique, elle est utilisée dans les essais contrôlés pour calculer l'amplitude de l'effet du traitement, ou « taille d'effet », généralement calculée comme la différence entre le score pré-traitement et le score post-traitement. Arrêtons-nous brièvement sur l'origine de ce test et sur l'histoire de ses développements.

L'échelle a vu le jour en 1960 dans un article intitulé « *A rating scale for depression* » (Hamilton, 1960), de Max Hamilton, un psychiatre anglais versé dans les techniques statistiques (Roth, 1990). L'introduction de l'article précise que « la présente échelle a été conçue pour être utilisée uniquement chez les patients déjà diagnostiqués comme présentant un trouble de l'humeur de type dépressif » (traduction) et que par ailleurs « elle est utilisée pour quantifier les résultats d'un entretien, et sa valeur dépend entièrement des compétences de la personne qui réalise l'entretien à éliciter les informations nécessaires » (traduction). Il s'agit donc d'une échelle hétéro-administrée, qui nécessite le recours à un professionnel de la santé mentale formé à la passation de l'échelle.

Dans sa forme originelle, l'échelle comprend 17 variables participant à la cotation. Elles sont mesurées sur une échelle à 5 niveaux (0 à 4 : absent, léger, modéré —2 et 3 —, sévère) si une quantification est possible (8 d'entre elles) ou à 3 niveaux (0 à 2 : absent, léger ou douteux, clairement présent) sinon (les 9 autres). Le score global de l'échelle est donné par la sommation des scores de chacune des questions, soit donc un score entre 0 et 50 (mais dans différentes versions ultérieures [Kriston et von Wolff, 2011] le score maximum peut être de 52, 54, 55...).

Les items de l'échelle sont restés globalement inchangés depuis plus de 40 ans, tandis que les conceptualisations des troubles dépressifs ont évolué de façon importante, en particulier avec les versions successives du DSM, et spécialement l'actuel DSM-IV (APA, 2000a) qui fait fonction de nosologie internationale « officielle », en particulier pour les travaux de la psychiatrie scientifique. Aujourd'hui, de nombreux critères du DSM-IV ne se trouvent pas dans l'échelle de Hamilton, tout autant que de nombreux critères de l'échelle ne se trouvent plus dans les critères du DSM-IV (APA, 2000a). Nous reviendrons sur ce que cet exemple illustre parfaitement : la « force de la norme ».

Tableau 1 : Les 17 variables de l'échelle de Hamilton

N°	Cotation	Symptôme évalué
1	0-4	Humeur dépressive
2	0-4	Culpabilité
3	0-4	Suicide
4	0-2	Insomnie de début de nuit
5	0-2	Insomnie de milieu de nuit
6	0-2	Insomnie de fin de nuit
7	0-4	Travail et centres d'intérêt
8	0-4	Ralentissement
9	0-2	Agitation
10	0-4	Anxiété psychique
11	0-4	Anxiété somatique
12	0-2	Symptômes somatiques, gastro-intestinaux
13	0-2	Symptômes somatiques généraux
14	0-2	Symptômes génitaux
15	0-4	Hypocondrie
16	0-2	Perte d'insight
17	0-2	Perte de poids

Pour autant, il n'existe aucune norme communément admise sur l'interprétation qualitative du niveau de sévérité sur l'échelle de Hamilton (Kriston et von Wolff, 2011). Selon la référence utilisée parmi les cinq recommandations disponibles (Bech, 1996; APA, 2000b; Furukawa *et al.*, 2007; Baer et Blais, 2008; NICE, 2009), un même score de 23 ou 24 peut ainsi être interprété comme modéré, majeur, sévère, ou très sévère! Les conséquences thérapeutiques du choix d'un référentiel plutôt que d'un autre sont drastiques: si l'on adopte le point de vue de l'APA, un score Hamilton de 23 imposera une prescription d'antidépresseurs de synthèse pour une dépression « très sévère », tandis que l'adoption du référentiel de Baer (2010) autorisera une psychothérapie, voire un traitement phytothérapeutique utilisant le millepertuis (*Hypericum perforatum*), une plante en vente libre dont différentes études ont montré l'intérêt pour des dépressions dont le niveau de sévérité se situe aux alentours de 22 (Linde *et al.*, 2008).

L'ÉCHELLE DE HAMILTON : CENTRALE DANS LES PRATIQUES... ET DANS LES « NORMES »

L'échelle de Hamilton fait l'objet de multiples usages ou services. Par exemple, un rapide dénombrement en langue française montre que des dizaines de sites Web proposent des évaluations de la valeur d'un individu sur l'échelle: sites destinés au grand public et offrant divers conseils ou tests psychologiques, tout autant que des sites

destinés aux médecins généralistes (www.medsyn.fr/perso/g.perrin/cyberdoc/doc/Hamilton.htm)... L'échelle n'est pas réservée aux seuls thérapeutes ou chercheurs spécialisés dans le traitement des dépressions : elle est installée dans le paysage et constitue une référence commune à tous ceux s'intéressant, à titres divers, à la dépression et à ses traitements. (www.mgfrance.org/documents/outils/Hamilton.htm)

L'échelle de Hamilton est devenue l'outil incontournable pour la mesure des effets des antidépresseurs, non seulement dans la pratique, mais également dans les normes édictées par les autorités de santé. En France, la passation du test dispose d'une cotation et d'une tarification à la Caisse d'assurance maladie, et les recommandations de bonne pratique et les référentiels des autorités de régulation et d'assurance maladie, qui s'appuient sur les études publiées, reposent en dernier ressort sur l'échelle de Hamilton. Une réponse récente de la Haute Autorité de Santé Française (HAS) à une saisine de la Caisse nationale d'assurance maladie des travailleurs salariés française (CNAMTS) en donne un exemple illustratif (HAS, 2010). Selon les termes de la saisine : « La CNAMTS a saisi la HAS d'une demande de validation de deux documents rédigés dans le cadre des actions de maîtrise médicalisée de l'Assurance Maladie et concernant l'épisode dépressif caractérisé (EDC) isolé de l'adulte. Ces deux documents sont destinés à être utilisés en 2010 pour une action d'accompagnement auprès des médecins généralistes sur le thème de l'EDC isolé de l'adulte. Leur objectif est de rappeler aux médecins généralistes les points forts du diagnostic et de la prise en charge de l'EDC. Il s'agit :

- d'un document d'aide à la décision ;
- d'un document présentant de manière simplifiée les critères diagnostiques de l'EDC selon les classifications CIM-10 et DSM-IV (p. 4).

La vocation des documents évoqués à façonner les pratiques de prise en charge de la dépression — des médecins généralistes en l'occurrence — est manifeste. Il est par ailleurs précisé par la CNAMTS que les documents transmis pour validation à la HAS reposent sur trois références, et uniquement celles-ci :

- les recommandations de l'ANAES de mai 2002 : Prise en charge d'un épisode dépressif isolé de l'adulte en ambulatoire (ANAES, 2003) ;
- les recommandations de l'AFSSAPS d'octobre 2006 : Bon usage des médicaments antidépresseurs dans le traitement des troubles dépressifs et des troubles anxieux de l'adulte. Recommandations de bonne pratique (AFSSAPS, 2006) ;
- une méta-analyse récente : Jay C. Fournier ; Robert J. DeRubeis ; Steven D. Hollon *et al.* Antidepressant drug effects and depression severity : a patient-level meta-analysis. JAMA 2010 ; 303(1) : 47-53 (Fournier *et al.*, 2010 : 5).

Comme indiqué plus loin, la méta-analyse de Fournier *et al.* (2010) utilise comme critère d'exclusion des études à analyser le fait de ne pas utiliser l'échelle de Hamilton. Les documents produits par la CNAMTS et validés par la HAS reposent donc uniquement sur des études utilisant comme mesure d'efficacité l'échelle de Hamilton. Ainsi, l'échelle est à ce point devenue l'outil central que les recherches et évaluations n'y

recourant pas sont parfois exclues, sans autre forme de procès, des publications et des référentiels pratiques.

L'échelle de Hamilton, au cœur de l'évaluation de l'efficacité des traitements de la dépression, se trouve donc également au cœur du façonnage des pratiques médicales de prise en charge de la dépression par les recommandations de bonne pratique formulées par les autorités de santé. Recommandations dont l'influence a été récemment largement accrue par la mise en place des procédures obligatoires d'évaluations des pratiques professionnelles par la HAS¹.

LES « BESOINS » D'ÉCHELLES DE LA « MÉDECINE DES PREUVES »

Le recours à des échelles quantitatives est d'autant plus fréquent qu'elles répondent bien aux exigences modernes de la « médecine des preuves », de l'*Evidence based medicine* (EBM, « Médecine fondée sur les faits » en français).

Selon les termes mêmes de ses développeurs, l'EBM n'est pas une théorie scientifique, ni une nouvelle approche des sciences biomédicales, mais une théorie de l'optimisation de la pratique clinique (Djulgovic *et al.*, 2009). Selon les termes de Sackett *et al.*, « Pratiquer l'Evidence-Based Medicine signifie intégrer l'expertise clinique individuelle avec les meilleures données cliniques extérieures disponibles issues de la recherche systématique » (Sackett *et al.*, 1996, traduction). Expertise clinique du praticien et résultats de la recherche scientifique systématique sont deux des trois pieds du tripode de l'EBM (avec les valeurs et les préférences en matière de santé du patient) : « un bon médecin utilise à la fois son expertise clinique individuelle et les meilleures données extérieures disponibles, et aucune des deux n'est suffisante à elle seule » (*ibid.*, traduction).

La pratique de l'EBM consiste donc en une *intégration* des données issues de la recherche scientifique systématique et de l'expérience clinique du praticien, qui façonne les décisions et actions concrètes au sein d'une relation thérapeutique avec un patient singulier.

On perçoit donc dès maintenant que les catégories utilisées, au nombre desquelles le score de Hamilton, pour produire « les meilleures données disponibles » ont une importance structurante critique pour l'ensemble du processus EBM, qui vise *in fine* à permettre une décision thérapeutique reposant sur une analyse complète des probabilités relatives qu'a un patient spécifique de bénéficier ou de pâtir de différentes stratégies thérapeutiques envisageables pour les problèmes qu'il présente. L'objectif du processus est de pouvoir classer les traitements selon leurs rapports bénéfiques-risques. Pour cela, chaque traitement peut être résumé par un indicateur quantitatif unique : la PRBP (la « probabilité relative de bénéficier ou de pâtir »). Cet indicateur est calculé à l'aide des données issues des essais thérapeutiques et des enquêtes épidémiologiques. Ayant calculé les PRBP pour un patient donné des différents traitements possibles, on

1. Site de la HAS, section « La politique d'évaluation des pratiques de la HAS », www.has-sante.fr/portail/jcms/c_434225/la-politique-devaluation-des-pratiques-de-la-has, consulté le 21 février 2011.

choisira le traitement ayant la meilleure PRBP (voir par exemple Strauss *et al.*, 2007 : 118-137) pour les détails sur les modalités de calcul).

En matière d'essais thérapeutiques, l'étalon-or méthodologique est l'essai contrôlé randomisé (ECR) en triple aveugle² grâce auquel on pourra comparer le taux d'évènements à éviter — la dépression par exemple — du groupe de référence (notons TER, pour taux d'évènements du groupe de référence) et celui du groupe expérimental recevant le traitement testé (notons TEE). La différence des deux, différence des risques ou réduction absolue de risque (notons RAR = TER - TEE) est un indicateur de la taille de l'effet du traitement. Le calcul pour les effets négatifs est similaire, mais on s'intéresse alors à la comparaison des évènements négatifs dans les groupes et à l'augmentation des risques. Pour un traitement T, la PRPB « des études » se calcule simplement comme le ratio de la variation des risques de présenter la maladie sur la variation des risques de présenter les effets indésirables.

Mais ce calcul ne contient rien de spécifique au patient concret auquel on envisage de prescrire le traitement testé. En particulier, ses risques de présenter les évènements qu'est censé éviter le traitement ne sont peut-être pas les mêmes que ceux des patients du groupe contrôle des essais. Par exemple, les risques de présenter une dépression selon que l'on soit un homme ou une femme, d'une CSP élevée ou faible, d'un niveau d'éducation supérieur ou inférieur au baccalauréat... ne sont pas identiques dans les études épidémiologiques. Il faut donc « singulariser » la formule en y intégrant les taux d'évènements attendus spécifiques au patient. Ces données « spécifiques » au patient peuvent par exemple être estimées à partir des études épidémiologiques disponibles en intégrant les facteurs de risque du patient, *qui sont disponibles dans les études*. On peut ainsi calculer pour chaque traitement un ratio bénéfices-risques spécifique au patient.

Des calculateurs³ et des outils graphiques (par exemple des nomogrammes) sont à la disposition des praticiens pour intégrer ces données dans leur pratique.

Cette démarche d'application des résultats des études — résultats statistiques, puisqu'il s'agit d'études sur des groupes expérimentaux ou des populations — au patient singulier conduit certes le clinicien à porter un regard clinique sur les caractéristiques du patient, mais c'est un regard *conformé* par les seules caractéristiques dont il a été tenu compte dans les essais thérapeutiques et dans les enquêtes épidémiologiques sur la pathologie.

Le point de vue du patient est intégré, *in fine*, dans le calcul de la PRBP. En effet, chacun n'attribue pas la même valeur à différents états de santé. Par exemple, l'importance relative des effets de la dépression et des effets des traitements antidépresseurs peut varier selon les patients et leur situation : on n'accordera ainsi sans doute pas la même importance aux effets respectifs de la dépression et des antidépresseurs sur la libido et l'érection selon que l'on soit un homme ou une femme, que l'on soit en

2. Ni les patients, ni les médecins, ni les évaluateurs ne savent à quels groupes appartiennent les patients.

3. Voir par exemple www.ktclearinghouse.ca/cebm/toolbox.

couple ou pas, que l'on ait 20 ans ou 80... Pour tenir compte de ces différences d'évaluation, on peut estimer la valeur qu'accorde le patient aux différents états problématiques évités ou provoqués par le médicament en lui demandant par exemple d'évaluer la gravité de chaque état sur une échelle numérique allant de 0 (la mort) à 1 (la pleine santé). Le rapport entre la valuation des problèmes évités par le médicament et ceux qu'ils provoquent fournit une estimation de la valeur relative de la gravité de chaque état pour le patient. On peut alors intégrer ce coefficient (notons gP pour gravité évaluée par le patient P) dans le calcul de la PRBP pour tenir compte de ses préférences.

Le point de vue du patient est donc certes intégré dans le calcul de la PRBP, mais c'est un point de vue lui aussi *conformé* par les catégories des études, car c'est sur la base des états de santé que ces catégories définissent qu'il lui est demandé de se prononcer.

En fin de compte, comme nous le précisons plus haut, c'est sur la base d'une analyse comparative des PRBP des différents traitements envisageables que s'opère le choix du traitement qui va être retenu : c'est celui qui a la meilleure PRBP. On comprend donc l'importance critique des indicateurs servant à définir l'efficacité des traitements et, s'agissant de la dépression, de l'échelle de Hamilton.

Cette logique a clairement vocation à s'appliquer non seulement à la prescription pharmacologique en médecine somatique, mais aussi à tout type de traitement dans toute spécialité médicale, dont la médecine mentale. C'est ainsi que Drake *et al.*, 2005, dans leur manuel d'EBM en médecine mentale, précisent que « chaque praticien a une obligation morale et éthique de comprendre et de partager les connaissances scientifiques sur les traitements [...]. Continuer à proposer des interventions en santé mentale sans comprendre les données disponibles actuelles n'est pas acceptable. De même, proposer les traitements qui sont familiers aux praticiens, ou pour lesquels ils ont une préférence personnelle, plutôt que les traitements en santé mentale qui correspondent aux données scientifiques n'est plus acceptable » (p. 62, traduction).

LES ESSAIS CONTRÔLÉS RANDOMISÉS ET LES MÉTA-ANALYSES

La qualité des « *evidences* » qui sont produites par la littérature scientifique dépend de nombreux critères, que l'EBM formalise, mais la méthodologie expérimentale utilisée joue un rôle particulièrement important. S'agissant de l'efficacité des traitements, une hiérarchie des méthodologies a été définie, au sommet de laquelle se trouvent l'essai contrôlé randomisé de forte puissance, déjà évoqué, et la méta-analyse d'essais contrôlés randomisés (HAS, 2000 : 15, 20, 47-50), dont l'existence permet de conclure à la « preuve scientifique établie », grade A des niveaux de preuve (p. 50). Au grade B — « présomption scientifique » — se trouvent les ECR de faible puissance, les études de cohorte... ; au grade C — « faible preuve scientifique » — on trouve au niveau 3 les études cas-témoins, et au niveau 4 les études rétrospectives et les études épidémiologiques descriptives, ainsi que les séries de cas, longtemps considérées selon d'autres logiques comme le *nec plus ultra* de l'analyse des effets des traitements, en psychanalyse par exemple. S'ajoute parfois à ces niveaux un 5^e degré encore plus faible, celui de « l'opinion d'expert », qui souvent n'est même plus mentionné.

L'importance de l'essai contrôlé randomisé est telle que Strauss *et al.*, 2007 n'hésitent pas à écrire que « les investigateurs doivent être fiers d'avoir utilisé la randomisation, et souvent elle est mentionnée explicitement dans le résumé (voire le titre !). Si une étude n'est pas randomisée, on peut interrompre la lecture et passer à l'article suivant. Ainsi, pour évaluer rapidement et de façon critique les articles, il est conseillé de commencer par examiner le résumé pour voir si l'étude est randomisée. Si elle ne l'est pas, l'étude peut être rejetée.

Ces niveaux de preuve s'appliquent tout aussi bien en médecine mentale. Par exemple, la recommandation de bonne pratique dans la « Prise en charge d'un épisode dépressif isolé de l'adulte en ambulatoire » de la Haute Autorité de Santé française (ANAES, 2003) les mobilise ainsi : « Dans l'épisode dépressif léger à modéré, les antidépresseurs et les psychothérapies sont efficaces (grade A pour les antidépresseurs, grade B pour les psychothérapies cognitivo-comportementales, grade C pour les autres psychothérapies, accord professionnel pour la psychanalyse). L'association antidépresseurs-psychothérapie n'a pas fait la preuve d'une plus grande efficacité que la psychothérapie seule dans ces formes légères à modérées (grade C) [...]. Dans l'épisode dépressif sévère, les antidépresseurs sont indispensables (grade A). L'association antidépresseurs-psychothérapie peut être proposée (grade C). Les antidépresseurs peuvent être associés aux neuroleptiques dans les formes psychotiques (accord professionnel) » (p. 6).

Dans les méta-analyses comme dans les revues de synthèse, l'ECR a la même valeur d'étalon-or que dans les manuels EBM ou les recommandations de bonne pratique des institutions. Par exemple, dans une récente revue de la fondation Cochrane portant sur la comparaison de l'efficacité des antidépresseurs et du placebo dans la dépression en soins primaires (Arroll *et al.*, 2009), la section « Méthodes » évacue la question des « critères de prise en considération des études dans cette revue » d'un laconique « seuls les essais contrôlés randomisés ont été inclus » (p. 3).

Dans deux grandes méta-analyses récentes, la procédure de sélection est la même. Kirsch *et al.*, 2008 n'intègrent en entrée du processus de sélection des études que des ECR. De même, Fournier *et al.*, 2010 spécifient que « les critères d'inclusion requerraient que les études soient des essais contrôlés randomisés » (traduction) et intègrent très explicitement le critère de randomisation directement dans la programmation de leur requête de recherche.

L'objectif des ECR est de comparer l'efficacité des traitements d'une problématique donnée, en les comparant entre eux et/ou avec un placebo. Pour cela, on cherche à calculer la différence entre les groupes sur des indicateurs pertinents standardisés dont les propriétés sont connues, et que les études utilisent toutes, afin de garantir la comparabilité des résultats. Lorsque l'indicateur est catégoriel (avoir ou non un épisode dépressif caractérisé du DSM-IV par exemple), on pourra comparer les taux d'événements survenant dans les différents groupes (cf. *supra* le calcul de la PRBP). Lorsque l'indicateur est continu (le niveau de sévérité d'une dépression par exemple), on pourra comparer la moyenne des résultats entre les groupes, ou discrétiser l'échelle pour retomber sur une logique catégorielle.

L'indicateur retenu est évidemment d'une importance critique: l'efficacité observable d'un traitement dans un ECR est en fait concrètement son efficacité à faire varier l'indicateur utilisé.

À ce titre, l'efficacité des traitements de la dépression est leur efficacité à faire évoluer l'échelle de Hamilton (HAM-D ou HDRS, Hamilton, 1960). C'est en effet l'instrument de référence des ECR dans l'évaluation de la sévérité dépressive: «L'échelle de Hamilton est devenue l'échelle d'évaluation de la sévérité de la dépression la plus largement utilisée dans le monde» (Williams, 2001, traduction); «L'échelle de Hamilton est l'un des instruments les plus communément utilisés pour mesurer la dépression» (Trajkovic *et al.*, 2011, traduction); «Depuis plus de quatre décennies, l'échelle de Hamilton est considérée comme l'étalon-or de la mesure de la sévérité dépressive, et l'on estime que c'est la mesure de la sévérité dépressive la plus utilisée dans les essais cliniques» (Santor *et al.*, 2008, traduction).

La sévérité dépressive est elle-même généralement la principale variable utilisée pour attester de l'effet des traitements de la dépression, comme en atteste la section «résultats mesurés» de la méta-analyse Cochrane déjà mentionnée:

«Les principaux résultats mesurés étaient la réduction des symptômes de dépression, mesurés de la façon suivante:

1. les résultats continus, rapportés en termes de réduction des symptômes dépressifs post-traitement, exprimés sur une échelle validée de mesure de la dépression (les plus utilisées étaient l'échelle de dépression de Hamilton et l'échelle de Montgomery-Absberg);
2. les résultats dichotomiques, rapportés sous forme de réponse clinique post-traitement. Les résultats étaient considérés comme positifs pour la rémission lorsqu'une réduction de 50% du score initial ou un score inférieur à 8 sur l'échelle de Hamilton était atteint (traduction).

L'échelle de Hamilton a quasiment acquis le même statut d'évidence méthodologique que l'essai contrôlé randomisé: à l'instar des raisons pour lesquelles une étude *doit* utiliser un dispositif contrôlé randomisé pour être digne de considération, les raisons pour lesquelles une étude doit impérativement utiliser l'échelle de Hamilton n'ont plus à être fournies; cela s'impose comme une évidence: «Les études étaient limitées aux échantillons de patients adultes en ambulatoire [...]. Par ailleurs, les études devaient inclure une comparaison traitement/placebo sur une durée d'au moins 6 semaines, ainsi que les scores HDRS au début et à la fin du traitement» (Fournier *et al.*, 2010, traduction). Dans la méta-analyse de Kirsch sur l'efficacité des ISRS, les 35 ECR intégrés utilisaient tous l'échelle de Hamilton comme mesure de sévérité initiale et finale (Kirsch *et al.*, 2008).

Dans une certaine mesure, on peut même considérer que l'échelle de Hamilton est progressivement devenue non seulement une référence de la mesure, mais aussi de la définition de la dépression: «En raison de son usage extrêmement répandu et de sa longue tradition clinique, l'échelle de Hamilton semble définir la dépression tout autant qu'elle la mesure. On pourrait critiquer le DSM-IV pour ne pas adéquatement

rendre compte de l'échelle de dépression de Hamilton, tout autant qu'on pourrait critiquer l'échelle de Hamilton pour ne pas couvrir totalement la dépression du DSM-IV» (Bagby *et al.*, 2004, traduction).

De la même manière que l'intelligence a progressivement été définie par les outils de sa mesure — les échelles psychométriques d'intelligence puis le QI — (Martin, 1997), la dépression et les variations des états dépressifs ont été progressivement définies par leur valeur sur une échelle quantitative unidimensionnelle. La définition de la dépression devient inutile et les débats sur cette notion deviennent superflus si un accord sur les conditions de sa mesure ou de son évaluation existe. Les discussions portent donc moins sur la définition de la notion ou des états dépressifs que sur les conditions de son évaluation, voire, en raison des objectifs prioritaires que constituent les visées thérapeutiques, sur l'efficacité comparée des différents traitements.

LA RARETÉ DES DÉBATS ET LE FAIBLE IMPACT DES CRITIQUES

Affirmer que l'échelle de Hamilton n'est pas contestée serait abusif. Des critiques ont été formulées. Ainsi, dans la mesure où l'échelle a été conçue pour une utilisation *uniquement* (Hamilton, 1960) sur des personnes diagnostiquées comme dépressives, cette non-congruence entre les critères diagnostiques actuels et les critères de sévérité de l'échelle, parmi d'autres difficultés, amène certains auteurs à une position très critique. Ainsi, Bagby *et al.* dans *l'American Journal of Psychiatry*, l'une des publications de référence du domaine, affirment : « En conclusion, nous avons été frappés du contraste marqué entre les efforts et la sophistication scientifique engagés dans la conception de nouveaux antidépresseurs et le recours persistant à des concepts et à des méthodes antiques pour évaluer les changements dans la sévérité de la dépression que ces médicaments ont pour objectif de modifier » (Bagby *et al.*, 2004).

Nonobstant ces critiques, l'échelle et son usage quasi exclusif et universel n'ont pas à ce jour été remis en cause. De manière plus générale, la contestation d'un indicateur profondément ancré dans les pratiques et dans les institutions a peu d'effet sur l'évolution des pratiques. Par exemple, les critiques adressées aux mesures de l'intelligence (Martin, 1997) ou celles adressées au PIB (Gadrey et Jany-Catrice, 2007) n'ont qu'un impact marginal sur les usages de ces notions et des outils de mesure associés. Remettre en cause de tels indicateurs suppose de profondes et larges transformations : pour atténuer le recours à l'échelle de Hamilton, il faudrait changer les pratiques des laboratoires, des équipes scientifiques, des revues scientifiques, les recommandations des autorités médicales, les manuels d'enseignement, les savoir-faire des thérapeutes... à une échelle non seulement nationale mais internationale. Pour reprendre une expression d'Alain Desrosières, l'échelle de Hamilton fait « tenir ensemble » beaucoup de choses (Desrosières, 2010 : 17-18). Une contestation efficace de l'échelle supposerait des modifications des pratiques, des connaissances et des normes implicites ou explicites dans un très grand nombre de secteurs.

L'échelle est à ce point solidement installée dans les pratiques que l'évolution des critères pathologiques qui l'ont initialement fondée ne la déstabilise pas. Née à partir

d'un cadre de connaissances pathologiques précises, elle s'est progressivement autonomisée pour devenir un outil à part entière, détaché de son origine et du cadre théorique qui lui a donné naissance. L'échelle de Hamilton, tout comme d'autres indicateurs et outils quantitatifs, a une vie autonome (Desrosières, 2005) : sa légitimité ne provient plus de son origine ni de son rapport à l'objet qu'elle est censée mesurer, mais de ses usages ancrés dans les institutions de santé et dans les pratiques thérapeutiques.

LA PRODUCTION DU NOMBRE

La méta-analyse de Kirsch en 2008 (Kirsch, *op. cit.*) portant sur 35 essais cliniques soumis à la Food and Drugs Administration (FDA) américaine pour l'agrément de quatre inhibiteurs sélectifs de la sérotonine — ISRS — (fluoxétine, venlafaxine, néfazodone, paroxétine), la classe d'antidépresseurs la plus prescrite, a calculé l'effet moyen de ces quatre molécules. La différence moyenne de score Hamilton avant et après le traitement (différence pré-post) était de 7,8 dans les groupes placebo, et de 9,6 dans les groupes traités, soit une différence des résultats pré-post entre les deux groupes de 1,8. Calculée en termes de *d* de Cohen (différence standardisée des moyennes), la différence était de 0,32, ce qui signifie que la différence des résultats représentait 32 % d'écart type. Ce n'est que pour les niveaux de sévérité supérieurs à 28 que l'écart entre les groupes contrôles et les groupes traités atteignait 3 sur l'échelle de Hamilton, soit 0,5 en termes de *d* de Cohen, seuil de significativité clinique retenu par le NICE (NICE, 2004). Par ailleurs, cette augmentation de l'efficacité semble être mieux expliquée par une diminution de l'effet placebo dans les niveaux les plus sévères de dépression que par une augmentation de l'effet thérapeutique.

Une méta-analyse ultérieure (Fournier *et al.*, 2010, *op. cit.*) confirme ces résultats en les raffinant : pour les patients se situant dans les niveaux léger à modéré (critères APA, soit entre 8 et 18), la taille d'effet était $d=0,11$; pour les patients dans le niveau sévère (19 à 22), $d = 0,17$; pour le groupe très sévère (>23), $d = 0,47$. Selon les termes des auteurs, les effets des antidépresseurs étaient inexistantes ou négligeables pour les patients présentant un niveau de sévérité initial léger, modéré, ou même sévère (soit donc entre 8 et 22) et ne devenaient cliniquement significatifs qu'au-dessus d'un score Hamilton initial de 25. On constate par ailleurs que ledit effet cliniquement significatif peut être obtenu par une diminution de score de 1 point sur seulement trois des dix-sept variables de l'échelle, diminution qui peut largement être due à des causes autres que la seule molécule testée. De fait, par exemple, dans 25 des 35 études, un autre psychotrope a pu être consommé, et dans la plupart des études, l'usage d'un sédatif était autorisé.

LA MISE EN SCÈNE DU NOMBRE

Au vu de ces résultats, Fournier et ses co-auteurs s'étonnent de l'incohérence entre deux constats : d'une part, l'inefficacité des antidépresseurs sur des dépressions dont les scores Hamilton sont inférieurs à 25 ; d'autre part, le fait que la majorité des patients consultant « en ville » se voient prescrire ces molécules, alors que la majorité

d'entre eux présentent des dépressions d'un niveau de sévérité inférieur. Entre 1987 et 1997 aux États-Unis, la proportion de personnes traitées en ville pour dépression est passée de 0,73 pour 100 personnes à 2,33 pour 100 personnes. Parmi celles-ci, la proportion utilisant des antidépresseurs est passée de 37 % à 75 %, tandis que la proportion de celles recevant une psychothérapie est passée de 71 % à 60 %. Dans le même temps, la proportion traitée par un médecin généraliste est passée de 69 % à 87 %, le nombre moyen de séances par an de 13 à 9, et la proportion de prise en charge par une assurance de 39 % à 55 % (Olfson *et al.*, 2002). Or, dans une étude portant sur 2500 patients de l'importante étude STAR*D sur l'efficacité des antidépresseurs en situation naturelle (Sinyor *et al.*, 2010), Gaynes *et al.*, en 2007, montrent que les patients consultant pour un épisode dépressif caractérisé au sens du DSM-IV, initial ou récurrent, en soins primaires comme en soins spécialisés, ont un score HAM-D moyen de 19,6 (écart-type de 6,5), très en dessous du score à partir duquel l'efficacité des antidépresseurs commence à être significative. Zimmerman *et al.*, en 2005, ont eux examinés 600 patients présentant également un épisode dépressif majeur au sens du DSM-IV : le score Hamilton moyen était de 20,3, et 48 % se trouvaient en dessous de 20, et donc en dessous du plancher d'inclusion dans la majorité des essais. Plus généralement, en tenant compte de tous les critères d'exclusion habituels des essais (risque suicidaire, abus ou dépendance à l'alcool ou autres drogues, autre trouble comorbide de l'axe I), 79 % des 600 patients auraient été exclus des essais. Les essais thérapeutiques des antidépresseurs sont généralement limités aux patients ayant la plus grande probabilité de démontrer une différence médicament-placebo, qui ne sont pas ceux que l'on rencontre en pratique courante, mais les antidépresseurs sont approuvés et diffusés sans tenir compte de ce problème majeur de transposition. Pour Zimmerman et ses co-auteurs, les antidépresseurs ne devraient être approuvés et utilisés que pour le sous-groupe limité de patients chez lesquels leur efficacité a été démontrée, ce qui impliquerait des études spécifiques pour d'autres groupes.

Pour Fournier et ses co-auteurs ces différents chiffres, et leur incohérence avec les résultats empiriques sur l'efficacité des antidépresseurs, s'expliquent par le fait que les prescripteurs, les concepteurs de politiques de santé, et les patients ne sont pas conscients du fait que l'efficacité des antidépresseurs n'est avérée que pour les cas les plus sévères de dépression, et que même dans ce cas, leur efficacité reste limitée, car cette caractéristique essentielle n'est pas mentionnée dans les messages marketing de ces médicaments à destination du public et des praticiens.

En effet, bien que l'on apprenne effectivement, en lisant par exemple dans les détails le dossier scientifique du Valdoxan analysé plus haut, que l'efficacité contre placebo de l'agomélatine dans les études présentées est d'environ 3 points HAM-D sur une population de patients dont le score HAM-D était d'au moins 26 en moyenne, cela n'apparaît pas dans les « Points à retenir » pour lesquels il est simplement précisé que « Valdoxan a démontré son efficacité antidépressive versus placebo sur l'échelle HAM-D chez les patients souffrant d'épisode dépressif majeur ». De surcroît, la formulation utilisée pour présenter l'efficacité sur les cas plus sévères laisse entendre que

cette efficacité existe également sur les cas moins sévères: « Valdoxan est efficace [...] dans le traitement des épisodes dépressifs majeurs, y compris chez les patients présentant une intensité accrue des symptômes dépressifs »; « L'efficacité a également été observée chez des patients présentant une dépression plus sévère (score HAM-D initial > 24) dans toutes les études positives contrôlées vs placebo. »

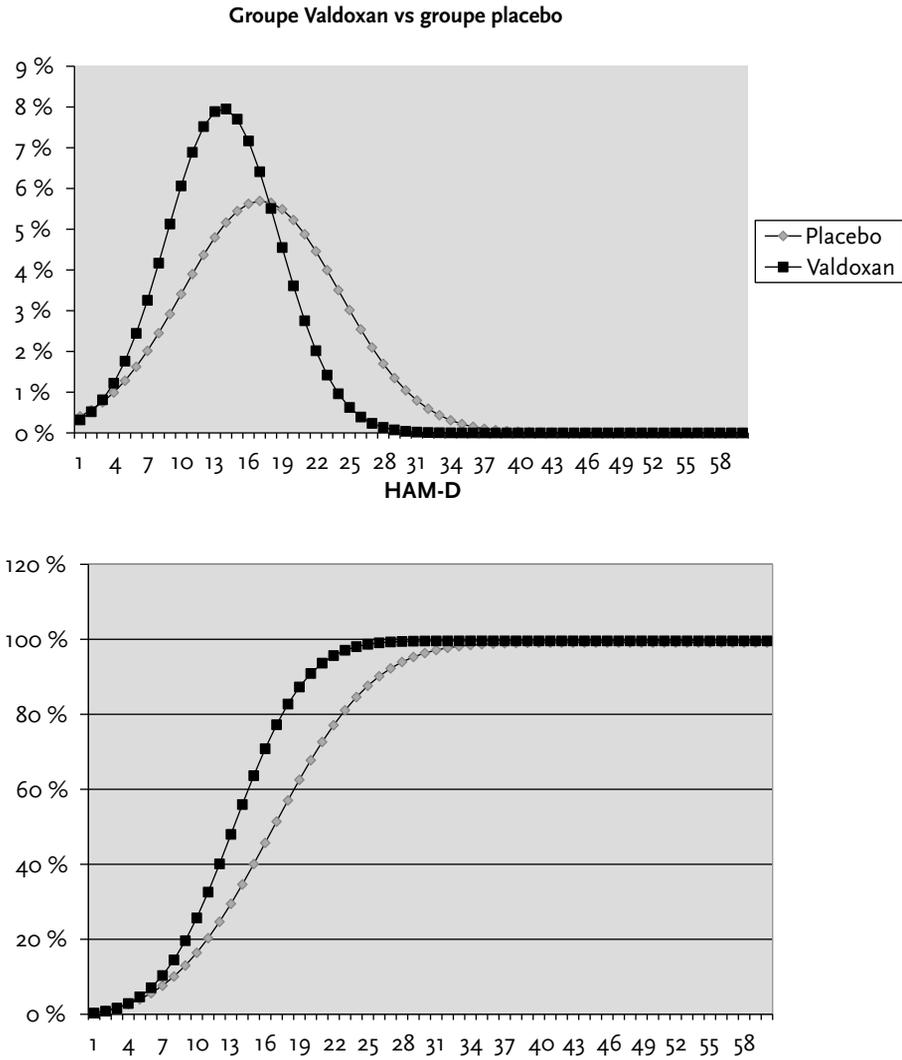
Cette logique de présentation est également présente dans le document « Résumé des caractéristiques du produit », qui donne pour indications thérapeutiques « traitement des épisodes dépressifs majeurs (c'est-à-dire caractérisés) chez l'adulte », et dont la section « Efficacité clinique et sécurité d'emploi » précise uniquement que « l'efficacité de l'agomélatine a été démontrée de façon significative dans trois des six études court terme, réalisées en double insu et contrôlées versus placebo ».

Quant à la « notice patient », l'indication y est introduite de la façon suivante: « Valdoxan appartient à un groupe de médicaments appelés antidépresseurs et il vous a été prescrit pour traiter votre dépression. La dépression est un trouble de l'humeur qui perturbe la vie quotidienne. Les symptômes de la dépression varient d'une personne à l'autre, mais comprennent souvent une profonde tristesse, une impression d'inutilité, une perte d'intérêt pour ses activités favorites, des troubles du sommeil, une impression de ralentissement, un sentiment d'anxiété, des variations de poids. » Aucune mention n'est faite des tailles d'effet ou de quelque indication d'efficacité, à l'exception du fait que « Valdoxan commence à agir sur les symptômes de la dépression dans les deux premières semaines de traitement ».

Une autre stratégie de présentation, décrite dans Kirsch, 2003, est également fréquemment utilisée. Il s'agit d'exprimer les résultats non plus en termes de différences des scores Hamilton, mais en différence des taux de répondants, « être répondant » étant classiquement défini comme le fait de présenter une réduction du score initial HAM-D d'au moins 50 %. C'est ce qui est utilisé dans le « dossier scientifique » pour la présentation de Olie et Kasper, 2007: « les taux de réponse pour les patients sévèrement déprimés étaient de 49 % vs 34 % respectivement pour les groupes Valdoxan et placebo ». La différence, 15 %, en impose davantage que la différence de score de 3 %. Pour autant, il s'agit — par définition — de la même chose, c'est-à-dire d'une différence de score Hamilton de 3,4 (17,1 vs 13,7). Le décodage de l'astuce rhétorique demande un peu de technique. Considérons (Figure 1) la représentation graphique des données de Olie et Kasper, 2007 présentées dans le dossier scientifique: le score HAM-D moyen initial est de 27 (1); le score HAM-D final est de 13,7 pour le groupe Valdoxan (2) et de 17,1 pour le groupe placebo (3); l'écart entre les deux groupes est de 3,44 (4); le score en deçà duquel les patients sont considérés comme répondant est de 13,5 — 50 % de 27 — (5) -score qui se trouve être identique à la moyenne du groupe Valdoxan —; 49 % des patients du groupe Valdoxan sont donc considérés comme répondants (6), contre seulement 34 % des patients du groupe placebo (7). L'importance de la différence entre les taux de répondants, malgré le faible écart absolu en termes de score, découle donc des propriétés de la loi normale, dont la pente de la fonction de répartition est très accentuée aux alentours de la moyenne: tout décalage

vers les scores HAM-D supérieurs du groupe placebo entraîne donc une diminution proportionnellement importante du taux de répondants dans ce groupe.

Figure 1 : Différence des taux de répondants pour une taille d'effet de 3 points HAM-D



LA DISPARITION DU NOMBRE

Ce qui ressort de cette analyse c'est que, conformément aux conclusions de Fournier *et al.*, 2010 (cf. *supra*), les caractéristiques essentielles d'efficacité de l'antidépresseur et de niveaux de sévérité des patients sur lesquels cette efficacité a été obtenue ne sont pas mentionnées de façon lisible et systématique dans les documents auxquels ont accès en routine prescripteurs et patients. Au contraire, l'indication est toujours pré-

sentée de façon plus large que ce que permettent de dire les études (épisode dépressif caractérisé plutôt que HAM-D >25) ainsi que l'efficacité (significativité statistique plutôt que signification clinique par exemple, comparaison des taux de répondants plutôt que différence des scores). Bien que l'information sur l'indication et l'impact des antidépresseurs existe et soit systématiquement mesurée sur l'échelle de Hamilton, les stratégies rhétoriques utilisées par le marketing des antidépresseurs vont donc dans le sens d'une présentation de l'information apportée par cette mesure qui tend à rendre plus larges les conditions de prescription des antidépresseurs, sans que pour autant il ne puisse être reproché aux laboratoires de dissimuler l'information, puisqu'elle se trouve bien dans le dossier scientifique.

On comprend facilement l'intérêt de ces stratégies rhétoriques et de présentation pour les firmes pharmaceutiques. Mais les autres sources d'informations disponibles ne sont pas pour autant mobilisées par les médecins. Dans une étude sur 650 médecins italiens (De Vito *et al.*, 2009), seuls 40 % lisent souvent des essais contrôlés randomisés et 18 % des méta-analyses ; 39 % n'utilisent pas les résultats des essais contrôlés randomisés (56 % pour les méta-analyses) pour prendre des décisions dans leur pratique clinique. Une étude récente portant sur 1800 étudiants français en 6^e année de médecine s'est intéressée à leur utilisation de la littérature scientifique (Roupret *et al.*, 2010). Il s'avère que si près de la moitié d'entre eux lisent « les articles médicaux dans les revues françaises » (mais seulement 8,8 % plus d'une fois par mois), 56 % ne lisent jamais « les articles médicaux dans les revues internationales », et ils ne sont que 5,9 % à les lire plus d'une fois par mois. Par ailleurs, seuls 7,3 % utilisent les bases de recherche médicale sur Internet, mais ils sont près de 95 % à trouver les conférences de consensus et les recommandations de bonne pratique essentielles ou très utiles.

Or, si l'on se réfère aux contenus des recommandations de bonne pratique présentées plus haut, il est clair que le seul ajout de ces recommandations aux supports marketing de l'industrie pharmaceutique, sans accès à la littérature princeps, ne peut permettre aux praticiens de disposer de l'information nécessaire et suffisante pour comprendre le périmètre exact de l'efficacité des antidépresseurs qu'ils envisagent de prescrire à leurs patients dépressifs. Moins encore pourraient-ils sur cette seule base déterminer les probabilités relatives de bénéficier ou de pâtir (PRBP) des différents traitements envisageables, en y intégrant les données scientifiques et les valeurs du patient.

En effet, si l'échelle dispose d'une gradation de la sévérité sur un continuum de 50 niveaux, dans la majorité des cas, les recommandations de bonne pratique la réduisent à 3 ou 4 niveaux discrets. La recommandation de la HAS sur la prise en charge de l'épisode dépressif isolé de l'adulte en ambulatoire, déjà évoquée, se fonde ainsi sur les niveaux ambigus légers, modérés, et sévères pour recommander l'usage de la psychothérapie ou des antidépresseurs.

CONCLUSION

Cette analyse de la place et des rôles de l'échelle de Hamilton dans le traitement et la conception des dépressions permet de tirer plusieurs enseignements essentiels, qui

dépassent largement le seul cas de cette échelle, et ont déjà pu être établis dans d'autres contextes et pour d'autres outils quantitatifs (Desrosières, 2010 ; Gadrey et Jany-Catrice, 2007 ; Martin, 1997).

Sans les déterminer de manière univoque, le recours à un indicateur quantitatif façonne les pratiques, les représentations et les notions évaluées. Plusieurs aspects de ce façonnage ont été évoqués au fil de l'article.

L'unidimensionnalité de la mesure, qui masque la complexité et la variabilité internes du phénomène, et tend à le réifier en une entité monolithique résultant d'une cause unique (un dysfonctionnement des monoamines dans la dépression) lors même que le phénomène peut être polymorphe et résulter de mécanismes multiples.

La disparition du besoin de la définition, le phénomène finissant par être assimilé à l'instrument qui le mesure (selon une conception implicitement opérationniste), au point que l'on perd de vue le caractère artefactuel de l'entité ainsi construite.

La standardisation de la manière dont on évalue les grandeurs psychiques, qui repose *in fine* sur l'hypothèse fort peu crédible d'une homogénéité des psychismes individuels.

La dimension performative de la définition, particulièrement critique dans le cas des troubles mentaux, puisque la constitution même de ces troubles chez le sujet intègre un façonnage social médié par la réflexivité du sujet et la « grammaire relationnelle » qui organise ses interactions lorsque le trouble mental considéré y intervient.

L'existence du standard produit un effet de mise en cohérence de tous les niveaux des instances concernées par les dépressions et leur traitement, qui détermine un système extrêmement cohérent et très interdépendant. La puissance de la norme ainsi engendrée entraîne une forte inertie au changement. De la définition de l'échelle de mesure à la manière dont les personnes pensent, vivent, et agissent leur « dépression », en passant par la définition des traitements, les essais thérapeutiques, les recommandations de bonne pratique, la formation des praticiens, les campagnes nationales d'information (Briffault *et al.*, 2010), les médias, les catégories de la grammaire sociale commune... c'est toute une « construction sociale » que fait « tenir ensemble » l'échelle de Hamilton et « la dépression » qu'elle fait exister : « l'objet, initialement produit d'une convention, devient réel après avoir été transmis clé en main, et réutilisé par d'autres » (Desrosières, 2010 : 139).

Trois raisons peuvent expliquer la « résistance » d'un outil ou d'une norme, c'est-à-dire sa capacité à résister à la critique interne et à la mise en cause de sa centralité sociale. Premièrement, un changement rend délicat le travail de comparaison et donc de validation d'une nouvelle méthode ou d'une nouvelle substance thérapeutique : l'échelle agit comme une norme dont il est très difficile de s'écarter ; tenter une nouvelle approche, c'est prendre le risque de devoir « tirer un trait » sur 50 ans de résultats disponibles. Deuxièmement, l'échelle se situe à l'intersection de plusieurs espaces sociaux régulés : les diverses communautés scientifiques et professionnelles (médecins, psychiatres, psychothérapeutes, pharmaciens, chercheurs...), les industriels du secteur pharmaceutique, les autorités de la santé publique, les associations d'usagers et de

malades... Critiquer ou modifier cette échelle revient à déstabiliser, d'une manière ou d'une autre, les pratiques en vigueur dans ces divers espaces sociaux, dont beaucoup d'acteurs ont par ailleurs un intérêt spécifique à maintenir le *statu quo*. Troisièmement, la logique numérique de l'échelle de Hamilton s'inscrit parfaitement dans cette « confiance dans les nombres » grâce à laquelle « le succès des mesures vient de ce qu'elles donnent des directions aux activités mêmes qui sont mesurées. De cette manière, les individus sont rendus gouvernables [...]. Les nombres créent — et peuvent être comparés à — des normes, qui sont parmi les formes de pouvoir les plus douces, et pour autant les plus influentes, dans les démocraties modernes » (Porter, 1995, traduction).

Pour des opérateurs de santé publique qui se préoccupent de gérer des masses et de « réduire le niveau moyen de dépressivité de la population », comme pour des firmes pharmaceutiques qui se préoccupent de disposer d'un indicateur scientifié de l'efficacité de leurs produits, l'échelle de Hamilton n'a en effet que des avantages.

Pour des individus qui se préoccupent de sortir de leur état dépressif et de ne pas y retomber, la situation est sans doute différente. Ce qui se dégage des innombrables travaux disponibles, c'est avant tout que la dépression est le résultat d'un enfermement dans une situation (psychologique, relationnelle, sociale...) intolérable dont on ne peut plus sortir, et au sein de laquelle s'enclenche un cercle vicieux dépressogène dans lequel les mécanismes biologiques, psychologiques, relationnels, sociaux s'articulent de façon délétère pour produire de la dépression (Briffault, 2010). C'est sans doute là le point commun de toutes les dépressions. Mais il y a d'innombrables moyens différents de se trouver enfermé dans des situations intolérables et d'y être maintenu ! C'est pourquoi, malgré leur apparente similarité (ils ont tous un « épisode dépressif majeur »), les sujets des essais contrôlés randomisés et de l'échelle de Hamilton — à partir desquels est construite toute la connaissance sur les traitements dans la logique de l'*evidence-based medicine* — sont en vérité très différents. Mais dans un dispositif tel que celui des essais contrôlés randomisés, dont la finesse d'analyse ne dépasse pas quelques variables en raison des limitations statistiques liées à l'exigence d'obtenir des résultats « statistiquement significatifs », il est absolument impossible de tenir compte de cette complexité et de ces spécificités.

Soumises aux impératifs de la « médecine des preuves » et à la demande sociale et politique d'indicateurs quantitatifs standardisés d'évaluation des bénéfices thérapeutiques, toutes les thérapies nécessitant des approches cliniques rencontrent des difficultés à exister et à trouver une légitimité sociale et académique. C'est notamment le cas de la psychanalyse, des psychothérapies, ou encore de l'ostéopathie, de l'homéopathie ou de l'acupuncture, dont les pratiques et théories cliniques ne se prêtent que difficilement à l'expérimentation contrôlée standardisée.

RÉSUMÉ

Cet article se propose d'examiner le statut et le rôle d'une échelle quantitative d'évaluation des effets thérapeutiques des antidépresseurs. Cette échelle, dite HAM-D (échelle de Hamilton), joue un rôle central dans les pratiques médicales, les normes thérapeutiques, les politiques de santé et finalement les conceptions de la dépression. Tout en examinant les conditions sociales et scientifiques expliquant la centralité et l'importance de cette échelle, l'article détaille les conséquences des usages de cette échelle. L'échelle façonne la notion de dépression en proposant une représentation unidimensionnelle, une définition quasi opérationniste et performative. En numérisant la dépression, elle offre aux acteurs du marché des antidépresseurs la possibilité d'une « mise en scène du nombre » qui contribue largement à ce que les patients traités par ces molécules soient bien plus nombreux que ce qu'autoriseraient logiquement les caractéristiques de ceux sur lesquels la (toute relative) efficacité clinique des antidépresseurs a été montrée.

Mots clés : statistique, psychométrie, mesure, dépression, échelle de Hamilton

ABSTRACT

This article aims to review the status and role of a quantitative scale for assessing the therapeutic effects of antidepressants. This scale, known as HAM-D (Hamilton Scale), plays a central role in medical practices, standards of care, health policy and finally the concepts of depression. While examining the social and scientific conditions, explaining the centrality and importance of this scale, the article details the consequences of the uses of it. The scale shapes the concept of depression by providing a one-dimensional representation, an almost operational and performative definition. Digitizing the depression offers antidepressants market participants the ability to “stage the number” which significantly contributes to the fact that treated patients with these molecules are far more numerous than what logically would allow the characteristics of those on which the (relative) clinical efficacy of antidepressants has been shown.

Key words : statistics, psychometrics, measurement, depression, Hamilton Rating Scale

RESUMEN

Este artículo propone examinar el estatus y el papel de una escala cuantitativa de evaluación de los efectos terapéuticos de los antidepresivos. Esta escala, llamada HAM-D (Escala de Hamilton), juega un papel central en las prácticas médicas, las normas terapéuticas, las políticas de salud, así como en las concepciones de la depresión. Examinando las condiciones sociales y científicas que explican la centralidad y la importancia dada a esta escala, el artículo detalla las consecuencias de los usos de la misma. La escala elabora la noción de depresión, proponiendo una representación unidimensional, una definición casi operacionista y performativa. La cuantificación de la depresión ofrece a los actores del mercado de los antidepresivos la posibilidad de producir una “escenificación del número”, lo que contribuye ampliamente a que los pacientes tratados por medio de estas moléculas sean mucho más numerosos que aquellos que producirían lógicamente las características de aquellos a partir de los cuales ha sido presentada la (muy relativa) eficacia clínica de los antidepresivos.

Palabras clave : estadísticas, psicometría, medida, depresión, Escala de Hamilton

BIBLIOGRAPHIE

- AFSSAPS (2006), « Le bon usage des médicaments antidépresseurs dans le traitement des troubles dépressifs et des troubles anxieux de l'adulte », [www.afssaps.fr/Infos-de-securite/Recommandations-de-bonne-pratique/Bon-usage-des-medicaments-dans-le-traitement-des-troubles-depressifs-et-des-troubles-anxieux-de-l-adulte-recommandations-de-bonne-pratique/\(language\)/fre-FR](http://www.afssaps.fr/Infos-de-securite/Recommandations-de-bonne-pratique/Bon-usage-des-medicaments-dans-le-traitement-des-troubles-depressifs-et-des-troubles-anxieux-de-l-adulte-recommandations-de-bonne-pratique/(language)/fre-FR).
- ANAES (2003), « Prise en charge d'un épisode dépressif isolé de l'adulte en ambulatoire », www.has-sante.fr/portail/jcms/c_608233/prise-en-charge-d-un-episode-depressif-isole-de-l-adulte-en-ambulatoire.
- APA (2000a), *Diagnostic and Statistical Manual of Mental Disorders: DSM IV-TR*, American Psychiatric Association.
- APA (2000b), *Handbook of Psychiatric Measures*, New-York, American Psychiatric Press Inc.
- ARROLL, B., C. R. ELLEY *et al.* (2009), « Antidepressants Versus Placebo for Depression in Primary Care », *Cochrane Database Syst Rev*, n° 3, p. CD007954, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19588448.
- BAER, L. et M. A. BLAIS (dir.) (2008), *Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health*, Totowa (NJ), Humana Press Inc.
- BAGBY, R. M., A. G. RYDER, D. R. SCHULLER et M. B. MARSHALL (2004), « The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight? », *American Journal of Psychiatry*, vol. 161, n° 12, p. 2163-2177, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15569884.
- BECH, P. (1996), *The Bech, Hamilton and Zung Scales for Moods Disorders Screening and Listening: A Twenty Years Update With Reference to Dsm-IV and Icd-10*, Springer-Verlag Berlin and Heidelberg GmbH & Co. K% @ 3540612459% 7, 2^e édition revue.
- BECK, A. T., C. H. WARD, M. MENDELSON, J. MOCK et J. ERBAUGH (1961), « An Inventory for Measuring Depression », *Arch Gen Psychiatry*, vol. 4, p. 561-571, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=13688369.
- BRIFFAULT, X. (2010), *La fabrique de la dépression: Observer, comprendre, agir*, Paris, Armand Colin.
- BRIFFAULT, X., Y. MORVAN et E. ROSCOAT (2010), « [National Information Campaigns on Depression. A Biopsychosocial Anthropology?] », *Encephale*, vol. 36, Suppl 2, p. D124-132, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20513455.
- CHATELLIER, G., E. ZAPLETAL, D. LEMAITRE, J. MENARD et P. DEGOULET (1996), « The Number Needed to Treat: A Clinically Useful Nomogram in its Proper Context », *Bmj*, vol. 312, n° 7028, p. 426-429, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8601116.
- DE VITO, C., C. G. NOBILE *et al.* (2009), « Physicians' Knowledge, Attitudes and Professional Use of RCTs and Meta-Analyses: A Cross-Sectional Survey », *Eur J Public Health*, vol. 19, n° 3, p. 297-302, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19129347.
- DESROSIÈRES, A. (2005), *Comparer l'incomparable: essai sur les usages sociaux des probabilités et des statistiques*, Communication présentée Augustin Cournot, les sciences sociales et la mathématisation de l'économie, Paris, www.centrecournot.org/pdf/Desrosieres.pdf.
- DESROSIÈRES, A. (2008), *Pour une sociologie historique de la quantification*, Paris, Presses des Mines Paris Tech.
- DESROSIÈRES, A. (2010), *La politique des grands nombres: histoire de la raison statistique*, Paris, La Découverte.
- DJULBEGOVIC, B., G. H. GUYATT et R. E. ASHCROFT (2009), « Epistemologic Inquiries in Evidence-Based Medicine », *Cancer Control*, vol. 16, n° 2, p. 158-168, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19337202.
- DRAKE, R. E., M. R. MERRENS et D. W. LYNED (dir.) (2005), *Evidence-Based Mental Health Practice*, New York, Norton.

- EVIDENCE-BASED MEDICINE WORKING GROUP (1992), «Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine», *JAMA*, vol. 268, n° 17, p. 2420-2425, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1404801.
- FOURNIER, J. C., R. J. DERUBEIS *et al.* (2010), «Antidepressant Drug Effects and Depression Severity: A Patient-Level Meta-Analysis», *JAMA*, vol. 303, n° 1, p. 47-53, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20051569.
- FURUKAWA, T. A., T. AKECHI, H. AZUMA, T. OKUYAMA *et* T. HIGUCHI (2007), «Evidence-Based Guidelines for Interpretation of the Hamilton Rating Scale for Depression», *J Clin Psychopharmacol*, vol. 27, n° 5, p. 531-534, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17873700.
- GADREY, J. *et* F. JANY-CATRICE (2007), *Les nouveaux indicateurs de richesse*, La Découverte.
- GAYNES, B. N., A. J. RUSH *et al.* (2007), «Major Depression Symptoms in Primary Care and Psychiatric Care Settings: A Cross-Sectional Analysis», *Ann Fam Med*, vol. 5, n° 2, p. 126-134, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17389536.
- GOODWIN, G. M., R. EMSLEY, S. REMBRY *et* F. ROUILLON (2009), «Agomelatine Prevents Relapse in Patients with Major Depressive Disorder Without Evidence of a Discontinuation Syndrome: A 24-Week Randomized, Double-Blind, Placebo-Controlled Trial», *J Clin Psychiatry*, vol. 70, n° 8, p. 1128-1137, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19689920.
- HALE, A., R. M. CORRAL *et al.* (2010), «Superior Antidepressant Efficacy Results of Agomelatine Versus Fluoxetine in Severe MDD Patients: A Randomized, Double-Blind Study», *Int Clin Psychopharmacol*, vol. 25, n° 6, p. 305-314, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20856123.
- HAMILTON, M. (1960), «A Rating Scale for Depression», *J Neurol Neurosurg Psychiatry*, vol. 23, p. 56-62, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14399272.
- HAS (2000), «Guide d'analyse de la littérature et des recommandations», www.has-sante.fr/portail/jcms/c_434715/guide-danalyse-de-la-litterature-et-gradation-des-recommandations.
- HAS (2010), «Épisode dépressif caractérisé isolé de l'adulte : Réponse à une saisine de la CNAMTS», www.has-sante.fr/portail/jcms/c_986540/referentiel-concernant-lepisode-depressif-caracterise-isole-de-ladulte-saisine-du-3-aout-2010.
- KENNEDY, S. H. *et* R. EMSLEY (2006), «Placebo-Controlled Trial of Agomelatine in the Treatment of Major Depressive Disorder», *Eur Neuropsychopharmacol*, vol. 16, n° 2, p. 93-100, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16249073.
- KIRSCH, I. (2003), «St John's Wort, Conventional Medication, and Placebo: An Egregious Double Standard», *Complement Ther Med*, vol. 11, n° 3, p. 193-195, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14659385.
- KIRSCH, I., B. J. DEACON *et al.* (2008), «Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food and Drug Administration», *PLoS Med*, vol. 5, n° 2, p. e45, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18303940.
- KNEKT (2011), «Quasi-Experimental Study on the Effectiveness of Psychoanalysis, Long-Term and Short-Term Psychotherapy on Psychiatric Symptoms, Work Ability and Functional Capacity During a 5-year Follow-Up», *Journal of Affective Disorders*.
- KRISTON, L. *et* A. VON WOLFF (2011), «Not as Golden as Standards Should Be: Interpretation of the Hamilton Rating Scale for Depression», *J Affect Disord*, vol. 128, n° 1-2, p. 175-177, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20696481.
- LEMOINE, P., C. GUILLEMINAULT *et* E. ALVAREZ (2007), «Improvement in Subjective Sleep in Major Depressive Disorder with a Novel Antidepressant, Agomelatine: Randomized, Double-Blind Comparison with Venlafaxine», *J Clin Psychiatry*, vol. 68, n° 11, p. 1723-1732, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18052566.

- LINDE, K., M. M. BERNER et L. KRISTON (2008), « St John's Wort for Major Depression », *Cochrane Database Syst Rev*, n°4, p. CD000448, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18843608.
- LOO, H., A. HALE et H. D'HAENEN (2002), « Determination of the Dose of Agomelatine, a Melatoninergic Agonist and Selective 5-HT(2C) Antagonist, in the Treatment of Major Depressive Disorder: A Placebo-Controlled Dose Range Study », *Int Clin Psychopharmacol*, vol. 17, n° 5, p. 239-247, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12177586.
- MARTIN, O. (1997), *La mesure de l'esprit: Origines et développements de la psychométrie, 1900-1950*, Paris, L'Harmattan
- NICE (2004), « Depression: Management of Depression in Primary and Secondary Care. », n° 23, www.nice.org.uk/CG23.
- NICE (2009), « Depression: the Treatment and Management of Depression in Adults (update) », www.guidance.nice.org.uk/CG90.
- OLFSON, M., S. C. MARCUS *et al.* (2002), « National trends in the outpatient treatment of depression », *Jama*, vol. 287, n° 2, p. 203-209, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11779262
- OLIE, J. P. et S. KASPER (2007), « Efficacy of Agomelatine, a MT1/MT2 Receptor Agonist with 5-HT2C Antagonistic Properties, in Major Depressive Disorder », *Int J Neuropsychopharmacol*, vol. 10, n° 5, p. 661-673, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17477888.
- PORTER, T. M. (1995), *Trust in Numbers*, Princeton, New Jersey, Princeton University Press.
- ROTH, M. (1990), « Max Hamilton: A Life Devoted to Psychiatric Science », *Psychopharmacol Ser*, vol. 9, p. 1-9, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2204918.
- ROUPRET, M., L. KARILA, S. KERNEIS et J. H. LEFEVRE (2010), « [Is the Learning Process of "Evidence-Based Medicine" Influenced by Determined Factors? Results from a National Survey of 1870 French Students in their Sixth Year] », *Presse Med*, vol. 39, n° 6, p. e126-133, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20466510.
- SACKETT, D. (1995), « Evidence-Based Medicine », *Lancet*, vol. 346, n° 8983, p. 1171, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7475646.
- SACKETT, D. L. (1997a), « Evidence-Based Medicine », *Semin Perinatol*, vol. 21, n° 1, p. 3-5, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9190027.
- SACKETT, D. L. (1997b), « Evidence-Based Medicine and Treatment Choices », *Lancet*, vol. 349, n° 9051, p. 570; author reply 572-573, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9048806.
- SACKETT, D. L. (1998), « Evidence-Based Medicine », *Spine (Phila Pa 1976)*, vol. 23, n°10, p. 1085-1086, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9615357.
- SACKETT, D. L. et W. M. ROSENBERG (1995a), « The Need for Evidence-Based Medicine », *J R Soc Med*, vol. 88, n° 11, p. 620-624, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8544145.
- SACKETT, D. L. et W. M. ROSENBERG (1995b), « On the Need for Evidence-Based Medicine », *J Public Health Med*, vol. 17, n° 3, p. 330-334, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8527187.
- SACKETT, D. L. et W. M. ROSENBERG (1995c), « On the Need for Evidence-Based Medicine », *Health Econ*, vol. 4, n° 4, p. 249-254, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8528427.
- SACKETT, D. L., W. M. ROSENBERG, J. A. GRAY, R. B. HAYNES et W. S. RICHARDSON (1996), « Evidence Based Medicine: What it is and what it isn't », *Bmj*, vol. 312, n° 7023, p. 71-72, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8555924.

- SANTOR, D. A., D. DEBROTA, N. ENGELHARDT et S. GELWICKS (2008), « Optimizing the Ability of the Hamilton Depression Rating Scale to Discriminate Across Levels of Severity and Between Antidepressants and Placebos », *Depress Anxiety*, vol. 25, n° 9, p. 774-786, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17935212.
- SINYOR, M., A. SCHAFFER et A. LEVITT (2010), « The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Trial: A Review », *Can J Psychiatry*, vol. 55, n° 3, p. 126-135, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20370962.
- STOUT, E. et R. A. HAYES (dir.) (2005), *The Evidence-Based Practice: Methods, Models, and Tools for Mental Health Professionals*, Wiley.
- STRAUSS, S. E., W. S. RICHARDSON, P. GLASZIOU et B. HAYNES (2007), *Médecine fondée sur les faits*, Paris, Elsevier.
- TRAJKOVIC, G., V. STARCEVIC *et al.* (2011), « Reliability of the Hamilton Rating Scale for Depression: A Meta-Analysis Over a Period of 49 years », *Psychiatry Res*, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21276619.
- WILLIAMS, J. B. (2001), « Standardizing the Hamilton Depression Rating Scale: Past, Present, and Future », *Eur Arch Psychiatry Clin Neurosci*, vol. 251, Suppl 2, p. I16-12, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11824839.
- ZIMMERMAN, M., I. CHELMINSKI et M. A. POSTERNAK (2005), « Generalizability of Antidepressant Efficacy Trials: Differences Between Depressed Psychiatric Outpatients Who Would or Would not Qualify for An Efficacy Trial », *Am J Psychiatry*, vol. 162, n° 7, p. 1370-1372, www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15994721.