Comprendre le débat sur les antidépresseurs dans le traitement de l’épisode dépressif majeur [Understanding the Antidepressant Debate in the Treatment of Major Depressive Disorder]

Florian Naudet, Rémy Boussageon, Clément Palpacuer, Laurent Gallet, Jean-Michel Reymann, Bruno Falissard

To cite this version:
Florian Naudet, Rémy Boussageon, Clément Palpacuer, Laurent Gallet, Jean-Michel Reymann, et al.. Comprendre le débat sur les antidépresseurs dans le traitement de l’épisode dépressif majeur [Understanding the Antidepressant Debate in the Treatment of Major Depressive Disorder]. Thérapie, 2015, 9 (19), pp.140028. 10.2515/therapie/2014228. hal-01147040

HAL Id: hal-01147040
https://univ-rennes.hal.science/hal-01147040
Submitted on 23 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Understanding the Antidepressant Debate in the Treatment of Major Depressive Disorder

Florian Naudet1,2,3, Rémy Boussageon4, Clément Palpacuer2, Laurent Gallet3, Jean-Michel Reymann2,5 and Bruno Falissard1,6,7,8

1 INSERM U669, Paris, France
2 Centre d’investigation clinique CIC-P INSERM 1414, Hôpital de Pontchaillou, Centre hospitalier universitaire de Rennes & Université de Rennes 1, Rennes, France
3 Centre hospitalier Guillaume Régnier, Service hospitalo-universitaire de psychiatrie, Rennes, France
4 Faculté de médecine de Poitiers, Département de médecine générale, Poitiers, France
5 Laboratoire de pharmacologie expérimentale et clinique, Faculté de médecine de Rennes, Rennes, France
6 Université Paris-Sud et Université Paris Descartes, UMR-S0669, Paris, France
7 AP-HP, Hôpital Paul Brousse, Département de santé publique, Villejuif, France

Text received November 11th, 2014; accepted November 20th, 2014

Keywords: antidepressants; randomised controlled trials; placebo; publication bias; outcomes

Abstract – There is a long-standing polemic concerning the usefulness of antidepressants in the treatment of major depressive disorder. In this paper we propose to highlight some aspects of this controversy by exploring the mutual influence of psychopharmacology and trial methodologies. Indeed, antidepressant efficacy, if not proved, was accepted before antidepressant randomised controlled trials (RCTs) were run. While RCTs became a gold standard to meet the requirements of the regulatory bodies, methodological tools were required to measure outcomes and to test whether antidepressants provide statistically significant benefits as compared with a placebo. All these methodological options have nonetheless introduced fuzziness in our interpretation of study results, in terms of clinical meaningfulness and in terms of transposability to a real life settings. Additionally, selective publication raises concerns about the published literature, and results in many paradoxes. Instead of providing easy answers, the application of the RCT paradigm in MDD raises numerous questions. This is probably in the nature of all scientific studies, but it can be in contradiction with clinicians’ expectations, who want to be sure that the treatment will (or will not) work for their individual patients.

Mots clés : antidépresseurs ; essais randomisés contrôlés ; placebo ; biais de publication ; critères de jugements

Résumé – Comprendre le débat sur les antidépresseurs dans le traitement de l’épisode dépressif majeur. Il existe un vieux débat à propos de l’utilité des antidépresseurs dans le traitement de l’épisode dépressif majeur. Dans cet article, nous présentons certains aspects de la controverse en explorant l’influence mutuelle de la psychopharmacologie et de la méthodologie des essais. En effet, l’efficacité des antidépresseurs était, sinon prouvée, admise avant que les premières études contrôlées randomisées (ECR) ne soient conduites. Alors que les ECR devenaient, du point de vue des autorités sanitaires, le “gold standard” pour l’évaluation des médicaments, il devenait nécessaire d’adopter des outils méthodologiques permettant de mesurer des critères de jugement et de tester si les antidépresseurs permettaient l’obtention d’une différence statistiquement significative par rapport au placebo. Ces options méthodologiques ont néanmoins introduit du flou quand à l’interprétation des résultats des ECR, notamment en terme de significativité clinique et de transposabilité « à la vraie vie ». Au-delà, la publication sélective des ECR impacte la validité de la littérature publiée et résulte en de nombreux paradoxes. Ainsi, au lieu de fournir des réponses simples, l’application du paradigme de l’ECR à l’épisode dépressif majeur soulève de nombreuses questions. Il en va probablement de même pour toutes les études scientifiques, mais dans ce cas précis, cela rentre en contradiction avec les attentes des cliniciens qui veulent être sûr que leur traitement sera efficace (ou pas) pour leurs patients.

Abbreviations : see end of article.
1. Introduction

There is a long-standing but still active polemic concerning the usefulness of antidepressants in the treatment of major depressive disorder. Recently, some opinion leaders stated that antidepressants have no place in evidence-based medicine, while others consider that this is an "irrational polemic" and have disputed psychological interventions for depression. This debate could lead to a major public health problem, since treatments that are offered to patients (pharmacological or psychological) are being discredited by partisans of either side, and this risks depriving some patients with depression of useful treatments. The subject is too important to reduce to a mere opposition between "pro" and "anti" antidepressants; and it deserves careful examination from different points of view. In this paper we propose to highlight some aspects of this controversy.

2. Birth of the concepts of antidepressant and major depressive disorder

In the case of depression, stimulants were used as the treatment during the 1940s. In the 1950s, new substances such as iproniazid and imipramine were viewed as specific to treating depression, whereas earlier stimulants were regarded as non-specific. In 1958, Khun presented imipramine as an antidepressant although its biological foundations were not established. He noted that "best responses were obtained in cases of endogenous depression showing the typical symptoms of mental and motor retardation, fatigue, feeling of heaviness, hopelessness, guilt, and despair" and that this "condition is furthermore characterized by the aggravation of symptoms in the morning with a tendency to improvement during the day". Promptly, the monoamine theory of depression emerged with the work by Sigg who demonstrated that imipramine can potentiate the effects of noradrenaline, by Burn and Rand who described the uptake of noradrenaline by adrenergic nerves, by Marshall et al. who reported that imipramine blocked the uptake of serotonin by platelets, by Axelrod et al. who described the uptake of labelled noradrenaline by adrenergic nerves which could be blocked by imipramine and by Denger et al. who reported similar data regarding noradrenaline uptake by brain tissue. Arvid Carlsson developed zimelidine, a new treatment blocking the uptake of serotonin without blocking the uptake of catecholamines. While zimelidine had a very favourable safety profile, within a year and a half of its introduction, some case reports of Guillain-Barré syndrome emerged, apparently caused by the drug, prompting its manufacturer to withdraw it from the market. After its withdrawal, it was succeeded by fluoxetine and the other serotonin reuptake inhibitors (SRIs) which were considered as selective drugs with fewer adverse events.

The idea of an antidepressant, and the discoveries about their putative biological properties, reshaped the concept of depression. A debate emerged concerning whether there was any value in distinguishing "endogenous depression" and milder conditions in relation with stressful events known as neurotic depression (the Khun perspective) and treating them differently, or whether there was no basis for separate categories of depression since they all lie on a continuum of severity, as proposed by Akiskal and Mc Kinney. In 1980, the Diagnostic and Statistical Manual of Mental Disorders (DSM) III was retained the latter view by combining the two entities under the label of major depressive disorder (MDD).

Non-scientific reasons have probably also contributed to the wide acceptation of the concepts of antidepressant and MDD. Concerning the ideological conflict of interest, these concepts were not in favour of the psychiatric profession's desire to integrate with general medicine and to counter attacks from the antipsychiatry movement. Concerning the financial conflict of interest, the pharmaceutical industry also had an interest in promoting these concepts.

3. RCTs became inescapable in the evaluation of antidepressants

Alongside these conceptual changes randomized controlled trials (RCTs) developed in the evaluation of medication. The Medical Research Council (MRC) ran the first RCT versus placebo in 1948 to explore the efficacy of streptomycin in tuberculosis. Previous non-randomized studies had established that streptomycin worked in the short term treatment of tuberculosis, but an a posteriori interpretation of this trial is that it probably proved the "efficacy of RCTs" rather than the efficacy of streptomycin. In the years following this trial, many RCTs were funded by national public bodies, for example the MRC evaluation of imipramine versus phenelzine, electroconvulsive therapy and placebo in the relief of depressive illness. These trials were often concerned with broad questions regarding classes of treatments, rather than specific compounds.

After the thalidomide crisis in 1962, the Kefauver-Harris drug amendments were passed to ensure drug efficacy and greater drug safety. It was because medications entailed a risk that evidence of efficacy was sought and, for the first time, drug manufacturers were required to prove to the Food and Drug Administration (FDA) the efficacy of their products before marketing them. Gradually, the situation changed, public funding declined and the vast majority of clinical trials on drug treatments in psychiatry...
were sponsored and conducted by the pharmaceutical industry, the number of trials increased dramatically, trials concerned single patented compounds and were designed to meet the requirements of the regulatory bodies. While for a large proportion of medical interventions, few or no clinical trials are ever conducted, for antidepressants there are probably now well over a thousand.

4. The mutual influence of psychopharmacology and trial methodology

Nonetheless, it should be noted that antidepressant efficacy, if not proved, was accepted before antidepressant RCTs were run, and that no antidepressant in the RCT era was proved to be superior to imipramine in terms of efficacy. Thus, being thought-provocative, one can say that antidepressants have made advances in methodology possible, rather than stating that methodology has enabled major advances in psychopharmacology for MDD. Indeed, when RCTs became a gold standard, it became necessary for them to take into account the particular features of psychopharmacology, and especially those relating to MDD, for instance paying particular attention to inclusion criteria and outcomes. Concerning inclusion criteria, as it became necessary to accept a common definition of MDD, the DSM viewpoint was reinforced as a standard. It also became necessary to adopt measurable, relevant and consensual outcomes providing a sensitive and accurate estimate of change occurring with antidepressants. The Hamilton Depression Rating Scale (HDRS), developed in 1960, was progressively imposed as a standard, and was subsequently challenged by the Montgomery and Asberg Depression Rating Scale (MADRS), a scale developed to be particularly sensitive to treatment effects. It is nonetheless interesting that a scale that is to be used to assess the difference between a treatment and a placebo was developed to be particularly sensitive to specific changes occurring under treatment. The Clinical Global Impression (CGI) which rates severity on a scale of 1 to 7, was retained as a reference for global assessment and some self-administered questionnaires like the Beck Depression Inventory (BDI) among others were popularised by the wide development of RCTs in MDD. Binary outcomes also had to be adopted, such as response and remission, which have meaning for clinicians. Despite the fact that they are intuitive, their definition is not straightforward and a consensus emerged to derive these outcomes from continuous rating scales by calculating the proportion of people who fall below predefined threshold scores, which tend to be validated merely by convention and tradition. Since 1991 remission is defined as a score ≤ 7 on the 17 items of the Hamilton Depression Rating Scale (HDRS-17) and response is usually defined as a reduction of 50% on the HDRS-17.

5. Statistically significant versus clinically meaningful results

While these methodological tools enable the measurement of outcomes and test whether antidepressants provide statistically significant benefits as compared with a placebo, there is a considerable debate concerning the real meaning of the difference in term of its clinical significance. Indeed, the identification of a minimal clinically relevant difference on a scale is not straightforward. In 2004, the National Institute of Clinical Excellence stated that a Hamilton score difference of three points across groups could be considered as clinically significant. This threshold was consistent with previous research but a recent linking analysis provided new insight by suggesting that a slight reduction on the HAMD-17 of up to 3 points corresponds to a rating of “no change” as measured with the CGI. A change close to 10 points was linked to the “much improved” category defined by the CGI. But these considerations on an individual level are not totally transposable to group level. On the other hand, this study also suggested that the commonly used measures for response (1) and remission (2) in MDD trials could reasonably be considered valid because they were coherent with the CGI definitions “much improved” (1) and “not at all” or “borderline mentally ill” (2), respectively. Bearing in mind that the CGI is not a perfect gold standard, these results are very interesting.

6. RCTs and the dilution of efficacy

To cope with the questions of variability and randomness, randomised controlled trials (RCT) “tell stories” about average patients, and the statistical inferences underpinning RCT conclusions concern expected values of random variables. This type of paradigm implies that sufficient efficacy in a subgroup of patients can induce an impression of efficacy for the whole group, providing the study is adequately powered. This “dilution” of efficacy can occur especially in the case of heterogeneous categories such as MDD. Recent meta-analyses have indeed shed new light on this debate. Meta-analyses on aggregated data by Khan et al. and Kirsh et al. suggested that the baseline severity of depressive symptoms is related to clinical trial outcomes. These two meta-analyses were based on FDA data (i.e. an exhaustive set of studies) but were prone to an ecological fallacy since they were based on aggregated data. Nonetheless, their results were reproduced by Fournier et al. within the framework of an individual data meta-analysis. This study addressed the limitations
of aggregated data meta-analyses, but since personal data are difficult to collect, it was prone to publication bias. Nevertheless, these three meta-analyses concluded consistently that the distinction between antidepressants and placebo is clinically meaningful (using the National Institute for Clinical Excellence threshold for clinical significance) only for severe and very severe patients.

Interestingly, Gibbons et al. addressed the limitations of the preceding studies by reanalysing all intent-to-treat individual longitudinal data during the first 6 weeks of treatment for major depressive disorder from all sponsored randomized controlled trials on fluoxetine and venlafaxine. In this meta-analysis, average differences at 6 weeks were small and not clinically meaningful (2.5 HAM-D units) and baseline severity was not shown to affect symptom reduction. But these small overall mean differences translated into clinically significant differences in response rates (estimated response rates were 58.4% for drug versus 39.9% for placebo) and remission rates (59.1% for drug versus 41.9% for placebo, relative risk = 1.5, number needed to treat = 5). This finding seems surprising. Intuitively, the two methods of assessing outcome should produce similar conclusions, since they are derived from the same data. However, this result can be explained by an artefact inherent in the transformation of continuous data into categorical data, which can magnify small differences. But on the other hand, transformation of continuous outcomes into categorical outcomes implies a misclassification bias, and measures of association such as relative risk are likely to be biased towards 1. An alternative explanation is that “efficacy dilution” is at play here.

7. Antidepressant alibis

In all events, beyond any fuzziness concerning the interpretation of antidepressant efficacy in MDD, a large number of RCTs turn out negative. It is frequently suggested that this is due to a marked placebo response in antidepressant trials, which could result from many different factors, such as spontaneous improvement, statistical regression to the mean, low level of severity at inclusion, co-interventions, and other biases in addition to the so-called placebo effect. For example, spontaneous improvement is common in clinical practice and the number of follow-up assessments is related to a significant therapeutic effect. From a naive point of view, one might have expected that in MDD, since it is a “mental disorder”, the placebo effect (with its psychological component) might be greater than in other conditions and, as a consequence, the resulting true “pharmacological” effect would be weaker than in general medicine. However the distinction is probably more subtle. In a meta-analysis, Hrobjartsson et al. identified no statistically significant effect of placebo interventions in depression, while a meta-analysis by Kirsh et al. suggested that placebo effects were considerable. But the RCTs included in Hrobjartsson’s meta-analysis were not designed (they were underpowered) to study the placebo effect adequately. Similarly, in Kirsch’s meta-analysis, which comprises no “untreated group” or waiting list, we cannot determine the size of the placebo effect. There is thus considerable debate about the size, the nature and the mechanism of the placebo effect in depression. For example, it has been proposed that the apparent antidepressant effect could be in part an active placebo effect, or result from bias, since side effects like sexual effects of antidepressants could reveal the identity of the medication to participants or investigators.

Nonetheless, while some general medical drugs have very high effect sizes, the effect sizes obtained by psychiatric drugs are in the same range as most general medical pharmaceuticals. Although it is difficult to compare effect sizes of drugs in different conditions, indications and outcomes, this finding puts the small effect sizes observed with antidepressants into perspective.

8. Overestimation and distortion of efficacy

Antidepressants efficacy is nonetheless certainly overestimated in the published literature by selective publication and selective outcome reporting. To explore this phenomenon, Turner et al. performed an analysis of 74 studies that were submitted to FDA for the approval of 12 antidepressant drugs. Among these studies, the FDA considered that 38 (51%) were “positive” (with a statistically significant result on the principal outcome), 12 (16%) “indeterminate” and 24 (33%) “negative” (with no statistically significant result on the principal outcome). Among the “positive” studies, 37 (97%) were published and only one (3%) was not published. Among the “indeterminate” studies, 6 (50%) were published as positive and 6 (50%) were unpublished. Finally, of the “negative” studies, 3 (12%) were published as “negative”, in agreement with the opinion of the FDA, 5 (21%) were published as “positive”, in disagreement with the opinion of FDA and 16 (67%) were not published. The effect size measured by performing a meta-analysis on the basis of published results is 0.41 with a 95% confidence interval of [0.36-0.45], whereas it is estimated to be 0.31 with a 95% confidence interval of [0.27-0.35] based on all studies reported to FDA.

The best-documented case of selective outcome reporting is probably study 329. It was a large study of 275 depressed adolescents conducted by SmithKline Beecham in the US from 1993-1996. Its results failed to show any statistically significant difference between paroxetine and placebo for the two primary outcomes. A GSK internal document stated that the results of study 329 indicated paroxetine was no more effective than...
placebo, and provided guidance on how to manage these disappointingly results by recommending they should "effectively manage the dissemination of these data in order to minimize any potential negative commercial impact." It also stated that "it would be commercially unacceptable to include a statement that efficacy had not been demonstrated, as this would undermine the profile of paroxetine."[51] Subsequently, an article was written (or more precisely ghostwritten) with positive results concerning new secondary outcome measures that had been introduced. It was concluded that paroxetine is "generally well tolerated and effective for major depression in adolescents."[52]

### 9. Paradoxes in comparative effectiveness assessments

As a result of a selective outcome reporting of this type, meta-analyses are likely to give misleading impressions about efficacy and comparative effectiveness of antidepressants.[53,54] There is the case of reboxetine, a selective norepinephrine reuptake inhibitor used in the treatment of depression. The previously favorable risk-benefit profile of reboxetine shown in published trials[55] was reversed by the addition of unpublished data.[56] In a network meta-analysis performed by Cipriani et al., reboxetine was consistently shown to be worse than 11 other antidepressants,[57] including paroxetine which was however found in another meta-analysis by the same team to not have any superiority over placebo.[58] All in all, these meta-analyses appear paradoxical, giving the impression that paroxetine is not superior to placebo, while it does better than reboxetine, which has itself been shown not to be superior to placebo. Additionally, although the Cipriani study found differences between antidepressants, this was not the case for another network meta-analysis performed by Gartlehner et al.[59]

Another paradox has been shown in a recent paper comparing citalopram with its "me-too", escitalopram, which found an inconsistency between direct evidence (showing a superiority of escitalopram) and indirect evidence (which did not find any significant difference).[60]

### 10. Poor transposability of RCT results

Beyond these issues RCTs are often criticised for their lack of external validity. Indeed, the vast majority of patients with clinical depression are catered for in primary care, and most RCTs have involved secondary care patients.[61] These patients probably differ from primary care patients.[62,63] in terms of severity (primary care patients are less severely depressed, milder course of illness) and in terms of complaints (fatigue and somatic symptoms).[64] Additionally, antidepressant RCTs use numerous non-inclusion criteria (for example suicidal ideations)[65-67] and excluded patients are a more chronically ill group with more numerous previous episodes, greater psychosocial impairment, and more frequent personality disorders. Finally, the vast majority of RCTs last no more than 8 weeks, whereas it is recommended that an antidepressant treatment be continued for at least 6 months after remission of the episode.[68]

There is debate as to whether these issues can be translated into different outcomes between RCTs and a "real life" setting.[69-72]

### 11. Conclusion

While meta-analyses should be reproducible, in 2013, a meta-analysis of published and unpublished studies on agomelatine found "evidence suggesting that a clinically important difference between agomelatine and placebo in patients with unipolar major depression was unlikely";[73] in 2014 a meta-analysis of published and unpublished studies on agomelatine found that it "was an effective antidepressant with similar efficacy to standard antidepressants".[74] This particular paradox sums up the fuzziness of antidepressant literature. We suggest that, instead of providing easy answers, the application of the RCT paradigm to MDD raises many questions. This is probably in the nature of all scientific studies, but it can be in contradiction with clinicians' expectations: what they want is to be sure that the treatment will work for individual patients (or to know if it will not). At the same time, their clinical experience is biased by many other parameters, including placebo response. This is precisely where the debate arises.

Acknowledgements. We thank Angela Swaine Verdier for revising the English.

Details of contributors. Wrote the first draft of the paper: N.F. Revised the paper critically for important intellectual content: B.R., P.C., G.L., R. J.M., F.B. Final approval of the version to be published: N.F. B.R., P.C., G.L., R. J.M., F.B.

Conflict of interests. There are no conflicts of interest regarding this paper. All authors have completed the Unified Competing Interest form at http://www.icmje.org/doi_disclosure.pdf (available on request from the corresponding author) and declare that (1) No authors have support from any company for the submitted work; (2) N.F. has relationships (board membership or travel/accommodations expenses covered/reimbursed) with Servier, BMS, Lundbeck and Janssen who might have an interest in the work submitted in the previous 3 years; B.R., P.C. G.L. and R.J.M. have no relationships with any company that might...
have an interest in the submitted work in the previous 3 years; FB has relationships (board membership or consultancy or payment for manuscript preparation or Travel/accommodations expenses covered/reimbursed) with Sanofi-Aventis, Servier, Pierre-Fabre, MSD, Lilly, Janssen-Cilag, Otsuka, Lundbeck, Genzine, Roche, BMS who might have an interest in the work submitted in the previous 3 years (3) none of the authors’ spouses, partners, or children have any financial relationships that may be relevant to the submitted work; and (4) none of the authors has any non-financial interests that may be relevant to the submitted work.

Abbreviations. BDI: Beck Depression Inventory; CGI: Clinical Global Impression; DSM: Diagnostic And Statistical Manual Of Mental Disorders; FDA: Food and Drug Administration; HDRS: Hamilton depression rating scale; MDD: major depressive disorder; MRC: Medical Research Council; RCTs: randomised controlled trials; SRIs: serotonin reuptake inhibitors.

References

7. Sigg EB. Pharmacological studies with tofranil. Can Psychiatr Assoc J 1959; 4(Suppl): 75-85
25. Falissard B, Lukasiewicz M, Corruble E. The MDP75: a new approach in the determination of the minimal clinically meaningful difference in a scale or a questionnaire. J Clin Epidemiol 2003; 56(7): 618-21

2015 Société Française de Pharmacologie et de Thérapeutique Thérapie 2015
4. Suh T, Gallo J. Symptom profiles of depression among general medical service users compared with specialty mental health service users. Psychol Med 1997; 27(5): 1051-63
5. Suh T, Gallo J. Symptom profiles of depression among general medical service users compared with specialty mental health service users. Psychol Med 1997; 27(5): 1051-63
22. Correspondence and offprints: Florian Naudet, INSERM U3669, Maison de Solen, 97 boulevard de Port Royal, 75679 Paris cedex 14, France.
23. E-mail: floriani.naudet@gmail.com