Editorial

The FDA's failure to address the lack of generalisability of antidepressant efficacy trials in product labelling

Mark Zimmerman

Summary

According to the US Food and Drug Administration's (FDA's) regulations, the criteria used to select patients into registration studies should be addressed in a product's label. The FDA's labelling guidelines, which specifically indicate that the routine exclusion of patients of a certain level of severity should be noted in the label, has been uniformly ignored.

Declaration of interest

None.

Mark Zimmerman is the Director of Outpatient Psychiatry and the Partial Hospital program at Rhode Island Hospital and principal investigator of the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project.

Antidepressants have modest efficacy

Despite questions about the robustness of their efficacy,[1] antidepressants are one of the most commonly prescribed classes of medication. Half of the studies submitted to the US Food and Drug Administration (FDA) for approved antidepressants failed to show significant differences between medication and placebo.[2] The effect size of antidepressants has been considered to be modest, and related to the level of severity of depression.[3] Because of concerns about the number of studies that failed to demonstrate a difference between active drug and placebo, efforts have been undertaken to understand those aspects of trial design that impact on drug–placebo differences,[4] and suggestions have been made to modify trial design to enhance drug–placebo differences.[5]

Inclusion/exclusion criteria in antidepressant efficacy trials

One aspect of the methodology of antidepressant efficacy trials (AETs) that has changed over the past 20 years is the narrowness of the criteria used to screen patients into a study. More than a decade ago our clinical research group applied the inclusion/exclusion criteria typically used in an AET to patients presenting for treatment in our out-patient practice and found that most patients would not qualify for the trial.[6] This finding was independently replicated multiple times.[7] These consistent results raised questions about the appropriateness of prescribing antidepressants to many, perhaps most, patients with major depressive disorder (MDD) because antidepressant efficacy in a narrow subgroup of patients with depression does not ensure efficacy in all, or most, patients diagnosed with MDD.

We followed up our initial study of how many patients in routine clinical practice would qualify for an AET with a limited review of the psychiatric inclusion/exclusion criteria used in AETs. We examined the criteria of 39 AETs published between 1994 and 2000 in 5 journals and found variability in the inclusion/exclusion criteria used across studies.[8]

More recently, we conducted a comprehensive review of 170 placebo-controlled AETs published from 1995 to 2014, to determine whether there have been any changes in the inclusion/exclusion criteria subsequent to the publications that highlighted the unrepresentativeness of the samples studied in AETs.[9] We speculated that the concerns raised a decade earlier would result in a broadening of the inclusion/exclusion criteria thereby enhancing the generalisability of AETs. In fact, we found that a significant change has occurred. In a comparison of the inclusion/exclusion criteria of studies published during the past 5 years with those of studies published during the prior 15 years, we found that AETs have become more restrictive in the criteria used to select patients into the trials. The more recently published studies were significantly more likely to exclude patients with depression with any comorbid psychiatric disorder, and significantly more likely to require a minimum symptom duration that is longer than the DSM-5 2-week threshold. Moreover, the severity score cut-off on measures such as the Hamilton Rating Scale for Depression (HRSD) and Montgomery–Åsberg Depression Rating Scale for inclusion was significantly higher in the more recent studies.

Importance of symptom severity

AETs are expensive to conduct. Given their cost, and the difficulty in demonstrating a significant difference in outcome between active drug and placebo, it is easy to understand why a pharmaceutical company would be reluctant to fund studies of their antidepressant with more liberal inclusion/exclusion criteria if their product may not be effective for large subgroups of patients with depression. For example, the rising symptom severity threshold for inclusion likely reflects industry's response to research suggesting that the difference in efficacy between antidepressants and placebo is greater in more severely ill patients.[3] However, research on the relationship between baseline severity and drug–placebo differences has produced inconsistent results. The widely cited meta-analysis of the FDA database by Kirsch et al[3] found that clinically significant drug–placebo differences were present only for studies with higher baseline severity scores. An earlier analysis of the FDA database similarly found that trials with higher baseline severity were more likely to demonstrate a

significant difference between active drug and placebo.[10] More recent studies examined patient-level data. Fournier and colleagues[11] collected individual patient data from six studies and found that drug–placebo differences increased with increasing severity of depression. They also suggested that the benefit of antidepressants were minimal for patients with mild or moderate depression. In contrast, Gibbons *et al*[12] collected patient-level data of published and unpublished placebo-controlled studies of fluoxetine and venlafaxine. They did not find a significant effect of severity on drug–placebo differences. However, Gibbons *et al* did not describe the minimum symptom severity thresholds required for participation in the studies they included in their analysis. Nor did they describe the mean baseline HRSD scores of the included studies. Thus, it is not possible to determine whether their non-significant findings might have been due, in part, to the failure to include patients with mild and/or moderate depression severity. The Gibbons *et al* study was also criticised for including studies which did not require that all patients were diagnosed with MDD.[13,14] Thus, in consideration of the inconsistent findings, the final chapter on the relationship between severity and drug–placebo differences has yet to be written.

Irrespective of whether drug–placebo differences are related to depression severity, it is clear that all AETs require a minimum score on a measure of depression severity, and the thresholds used on these scales have increased in more recently conducted studies.[9] The increasing selectivity and reduced generalisability of the samples studied in AETs is where the FDA comes in.

The FDA guidelines on labelling

Back in 1977 the FDA issued guidelines for the clinical evaluation of antidepressant drugs.[15] The preface to the 1977 monograph anticipated that the guidelines would be updated approximately every 2 years. It has been nearly 40 years since their first issuance and they have yet to be updated.

One section of the FDA guideline for industry on the evaluation of antidepressants discussed the issue of sample selection, but the description of inclusion and exclusion criteria did not address in detail several of the methods used in contemporary studies. For example, there was little discussion of using symptom severity scales to select patients into studies, the exclusion of patients who express suicidal thoughts, or the exclusion of patients with comorbid substance use, non-depressive psychiatric disorders or medical illnesses.

The inclusion/exclusion criteria used to select patients is supposed to be addressed in a product's label. The 1977 FDA antidepressant guideline did not discuss the issue of labelling medications receiving FDA approval.

The FDA's Code of Federal Regulations on the labelling of medications (21 CFR 201.57) notes that a product's label should identify the subgroups of patients for which a medication is effective if the evidence of its effectiveness is limited to select subgroups.[16] A separate FDA industry guidance monograph on the labelling of prescription drugs indicated that the clinical studies section of a product's label should identify the important limitations of the empirical evidence supporting a product's effectiveness.[17] The FDA's guideline states that a label's description of the study population 'should identify those characteristics that are important for understanding how to interpret and apply the study results. The description thus should identify important inclusion and exclusion criteria [. . .] For example, the description should discuss enrollment factors that exclude subjects prone to adverse effects, the age distribution of the study

population, *a baseline value that results in a study population that is more or less sick than usual . . .*' [italics added] (Part III B.4.).

The inclusion criteria of every AET published during the past 20 years have required a minimum score on a depression symptom severity scale.[9] The symptom severity inclusion criterion has the greatest impact on the number of patients in routine clinical practice that would not qualify for a study.[6] Yet, no label for an FDA-approved antidepressant includes the caveat that the medication was only found to be effective in patients scoring above a symptom severity cut-off. No label for an FDA-approved antidepressant includes the caveat that the medication was not found to be effective in patients with mild MDD. Thus, the FDA labelling guidelines, which specifically indicate that the routine exclusion of patients of a certain level of severity should be noted in the product's label, has been uniformly ignored.

Another subchapter of the FDA's Code of Federal Regulations on the labelling requirements of prescription drugs (21 CFR 201.56) indicates that a product's label 'must be updated when new information becomes available that causes the labelling to become inaccurate, false, or misleading'.[16] The results of our literature review, which found that every AET of the past 20 years excluded patients who scored too low on a symptom severity measure, could be considered such 'new information' warranting a change in antidepressant product labels to note that the medications are indicated for patients with MDD of certain levels of severity. This is particularly true for the most recently approved medications in which the symptom severity inclusion thresholds are even higher than the cut-offs used in older studies. To be sure, the exclusion of patients with insufficient symptom severity (despite meeting DSM criteria for MDD) is not the only significant inclusion/exclusion criterion limiting the generalisability of AETs. The FDA should also consider requiring a label modification to denote the limits to generalisability due to the exclusion of patients with comorbid psychiatric disorders since most recent studies limit sample selection in this manner as well.[9]

Limits of effectiveness studies

Some might argue that the problem of limited generalisability of AETs is largely overcome in effectiveness studies such as the STAR*D trial which use minimal inclusion and exclusion criteria to recruit patients.[18] In effectiveness studies, in contrast to AETs, generalisability is prioritised and response to treatment is examined in a sample that is more representative of patients treated in routine clinical practice. Specifically, patients with comorbid disorders and with lower levels of depression severity are included. A limitation of effectiveness studies such as STAR*D, however, is the failure to include a placebo control group. Thus, conclusions cannot be drawn as to whether medication 'worked' in this more broadly representative group of patients.

Lack of clarity in EMA guidelines

To be sure, the problem of inappropriate labelling of antidepressants is not unique to the FDA. The European Medicines Agency (EMA) guideline for investigating antidepressant medications, updated in 2013,[19] specifically addressed the issue of symptom severity and labelling, although the EMA guideline discussed the labelling issue in a curious and inconsistent manner. In section 4.1.3, under the heading 'Extrapolations', the EMA guideline noted that 'Clinical trials will usually recruit patients, who are moderately ill, as it is difficult to demonstrate an effect in mildly ill patients'. Despite acknowledging that antidepressants may not be effective in patients with mild depression, who

nonetheless are diagnosed with MDD, the EMA guideline states that the demonstration of efficacy 'in moderately ill patients will be considered sufficient for a registration package to get a general license for "Treatment of Episodes of Major Depression"'. Unfortunately, the guideline does not explain the logic behind this declaration. It is perplexing how the guideline can simultaneously acknowledge that antidepressants may not work for a large number of patients with MDD, allow patients with mild depression to be routinely excluded from studies of the product's efficacy, but then approve the product for a broad indication that includes this unstudied subgroup. Moreover, in the very next sentence of the Extrapolations section of the EMA guideline, a limit on a product's approval is noted. That is, the treatment of a major depressive episode in the context of bipolar disorder requires a separate product development effort to warrant approval. Thus, an antidepressant would not be approved for the treatment of bipolar depression in the absence of data demonstrating the efficacy and safety of medication for this subgroup. It is unclear why this same approach would not also apply to the treatment of MDD of a severity level that falls below the cut-off scores on the scales that are routinely used to recruit patients into AETs. In fact, the EMA guidelines explicitly express concerns about generalisability. In section 4.2.4.1, the guideline states 'Though some of the earlier studies may be done in hospitalised patients, the majority of the database should be in out-patients for better generalisability of study results'. Should not concerns about generalisability extend beyond whether or not a patient is admitted to hospital for their depression?

Concern for the future in the era of personalised medicine

The focus of this editorial has been on the FDA because their explicit guidelines have been ignored. (In contrast, the EMA guidelines seem to lack an internally coherent, consistent framework in the discussion of generalisability and labelling.) Unless the FDA enforces their labelling guidelines, it is likely that the pharmaceutical industry will continue to limit the samples studied in AETs to those patients who are believed to be the most likely to demonstrate a difference between active drug and placebo. Of relevance to the future is what FDA labelling guidelines will be enforced in the era of personalised medicine when medications may be found to be effective only for subsets of patients with specific genetic or biological markers? If the FDA does not act now in the face of clear data demonstrating the limited generalisability of AETs, how confident can we be that it will appropriately label antidepressants in the future? If the current restrictive inclusion/exclusion criteria of AETs are not sufficient grounds for narrowing an antidepressant's label, then can officials at the FDA detail the conditions under which a label would be narrowed to the spectrum of patients who are included in AETs and for whom the medication has been found to be effective?

Mark Zimmerman, MD, Department of Psychiatry and Human Behavior, Brown University School of Medicine, Rhode Island Hospital, Providence, RI 0294, USA. Email: mzimmerman@lifespan.org

References

1 Kirsch I. *The Emperor's New Drugs: Exploding the Antidepressant Myth*. Basic Books, 2010.

2 Kahn A, Warner H, Brown W. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials. An analysis of the Food and Drug Administration Database. *Arch Gen Psychiatry* 2000; **57**: 311–7.

3 Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008; **5**: e45.

4 Gelenberg AJ, Thase ME, Meyer RE, Goodwin FK, Katz MM, Kraemer HC et al. The history and current state of antidepressant clinical trial design: a call to action for proof-of-concept studies. *J Clin Psychiatry* 2008; **69**: 1513–28.

5 Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* 2003; **72**: 115–27.

6 Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice. *Am J Psychiatry* 2002; **159**: 469–73.

7 Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, et al. Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR*D report. *Am J Psychiatry* 2009; **166**: 599–607.

8 Zimmerman M, Chelminski I, Posternak M. Exclusion criteria used in antidepressant efficacy trials: consistency across studies and representativeness of samples included. *J Nerv Ment Dis* 2004; **192**: 87–94.

9 Zimmerman M, Clark HL, Multach MD, Walsh E, Rosenstein LK, Gazarian D. Have treatment studies of depression become even less generalizable? A review of the inclusion and exclusion criteria in placebo conrolled antidepressant efficacy trials published during the past 20 years. *Mayo Clinic Proc* 2015; **90**: 1180–6.

10 Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 2002; **22**: 40–5.

11 Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010; **303**: 47–53.

12 Gibbons R, Hur K, Brown C, Davis J, Mann J. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry* 2012; **69**: 572–9.

13 Carroll BJ. Suicide risk and efficacy of antidepressant drugs. *JAMA Psychiatry* 2013; **70**: 123–5.

14 Spielmans GI, Jureidini J, Healy D, Purssey R. Inappropriate data and measures lead to questionable conclusions. *JAMA Psychiatry* 2013; **70**: 121–3.

15 US Department of Health, Education, and Welfare. *Guidance for Industry: Guidelines for the Clinical Evaluation of Antidepressant Drugs*. US Food and Drug Administration, 1977.

16 Department of Health and Human Services. Labelling requirements for prescription drugs and/or insulin. In: *Code of Federal Regulations, Title 21: Food and Drugs* (available at http://www.ecfr.gov/cgi-bin/text-idx?SID=f1f7838579f99e0fde1cc5cacd481299&mc=true&node=sp21.4.201.b&rgn=div6).

17 Center for Drug Evaluation and Research. *Guidance for Industry: Clinical Studies Section of Labeling for Human Prescription Drug and Biological Products – Content And Format*. US Food and Drug Administration, 2006.

18 Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D implications for clinical practice. *Am J Psychiatry* 2006; **163**: 28–40.

19 Committee for Medicinal Products for Human Use. *Guideline on Clinical Investigation of Medicinal Products in the Treatment of Depression*. European Medicines Agency, 2013: 1–22. Avaialable at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/05/WC500143770.pdf (accessed 17 Feb 2016).