

BMJ Open Spin in RCTs of anxiety medication with a positive primary outcome: a comparison of concerns expressed by the US FDA and in the published literature

Lian Beijers,¹ Bertus F Jeronimus,^{1,2} Erick H Turner,^{3,4} Peter de Jonge,^{1,2} Annelieke M Roest¹

To cite: Beijers L, Jeronimus BF, Turner EH, *et al.* Spin in RCTs of anxiety medication with a positive primary outcome: a comparison of concerns expressed by the US FDA and in the published literature. *BMJ Open* 2017;**7**:e012886. doi:10.1136/bmjopen-2016-012886

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-012886>).

Received 1 June 2016
Revised 28 November 2016
Accepted 6 February 2017



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Bertus F Jeronimus;
b.f.jeronimus@rug.nl

ABSTRACT

Objectives: This study aimed to determine the presence of spin in papers on positive randomised clinical trials (RCTs) of antidepressant medication for anxiety disorders by comparing concerns expressed in the Food and Drug Administration (FDA) reviews with those expressed in the published paper.

Methods: For every positive anxiety medication trial with a matching publication (n=41), two independent reviewers identified the concerns raised in the US FDA reviews and those in the published literature. Spin was identified when concerns or limitations were expressed by the FDA (about the efficacy of the study drug) but not in the corresponding published paper. Concerns mentioned in the papers but not by the FDA were scored as 'non-FDA' concerns.

Findings: Only six out of 35 (17%) of the FDA concerns pertaining to drug efficacy were reported in the papers. Two papers mentioned a concern that fit the FDA categories, but was not mentioned in the corresponding FDA review. Eighty-seven non-FDA concerns were counted, which often reflected general concerns or concerns related to the study design.

Conclusions: Results indicate the presence of substantial spin in the clinical trial literature on drugs for anxiety disorders. In papers describing RCTs on anxiety medication, the concerns raised by the authors differed from those raised by the FDA. Published papers mentioned a large number of generic concerns about RCTs, such as a lack of long-term research and limited generalisability, while they mentioned few concerns about drug efficacy. These results warrant the promotion of independent statistical review, reporting of patient-level data, more study of spin, and an increased expectation that authors report FDA concerns.

INTRODUCTION

It is becoming increasingly apparent that the results of scientific research are not always

Strengths and limitations of this study

- Novel study of spin regarding efficacy-related concerns in drug randomised clinical trials with positive primary outcomes.
- Food and Drug Administration (FDA) reviews are used as an objective source to compare with the published literature.
- The estimate may be conservative because FDA concerns required agreement between FDA reviewers, which was not always the case.
- No other types of spin were examined.

accurately reported. Results are frequently presented more positively than they actually are.^{1 2} For example, trials are more likely to be published if they are positive, and authors may selectively report outcome measures in such a way that negative trials appear positive.^{3–7} Reporting strategies, intentional or unintentional, which mislead readers in their evaluation of the beneficial effects of experimental interventions or their safety, are called 'spin'.^{8 9} Examples include the overinterpretation of results¹⁰ and exaggerated claims in press releases and media coverage of studies.¹¹

Although awareness of spin is increasing, most studies of spin in the scientific literature have focused on randomised clinical trials (RCTs) whose primary outcomes are negative.^{7 8 12–16} However, the fact that a primary outcome is positive does not preclude spin.¹¹ Researchers might, for example, emphasise the most favourable-looking variable or manipulate figures.^{3 5 16} In papers about RCTs with positive primary outcomes, the most prominent types of spin are arguably the exaggeration of drug efficacy and concerns regarding trial outcome reliability.

Examples include when the study drug had no effect on secondary end points or when drug efficacy was shown only for men or only for women. Treatment decisions are influenced by the way such concerns are presented.

One way to detect spin in publications on positive trials is to compare the concerns raised about a medication trial in the reporting paper with those reported by an objective, independent third party. The US Food and Drug Administration (FDA) can play such a role. Before a drug can be marketed in the USA, it must be approved by the FDA. Before starting any US trial, pharmaceutical companies must notify the FDA of its details, such as the primary end point and statistical analytic method. When the trial is completed, FDA medical and statistical reviewers inspect the sponsor's results, trying to replicate them using the patient-level data submitted by the sponsor. The reviewers then write a summary report on each of the trials, together with a recommendation as to whether the drug meets the FDA's criteria for marketing approval. These reports are examined by higher level reviewers (team leader, division director, sometimes office director), culminating in the decision whether to approve the drug.¹⁷ These reports are combined into an FDA drug approval package (hereafter called FDA review), which is made publicly available.¹⁸ The existence of multiple reviews increases the odds that any concerns regarding drug efficacy results will be expressed. Spin can be measured by comparing concerns mentioned in the FDA drug approval packages with those mentioned in corresponding trial publications.

The objective of this study is to examine spin in the reporting of positive RCTs of antidepressant medication for anxiety disorders, by comparing concerns expressed in the FDA reviews with those expressed in the published paper.

METHOD

FDA reviews

In a previous paper, Roest *et al*⁷ identified a cohort of phase 2 and 3 double-blind placebo-controlled clinical trials that had been registered with the FDA in pursuit of marketing approval; the drugs were second-generation antidepressants for the treatment of five anxiety disorders (generalised anxiety disorder, panic disorder, social anxiety disorder, post-traumatic stress disorder and obsessive-compulsive disorder). Nine drugs were approved for these indications by the FDA: seven selective serotonin reuptake inhibitors (SSRIs): paroxetine, paroxetine controlled release, sertraline, fluoxetine, fluvoxamine, fluvoxamine controlled release and escitalopram, and two serotonin norepinephrine reuptake inhibitors (SNRIs): venlafaxine extended release and duloxetine. FDA reviews were downloaded from the FDA's website (<http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm>), or, if these reviews were not available for download, requested from the FDA's Freedom of Information Office (<http://www.accessdata.fda.gov/scripts/foi/FOIRequest/requestinfo.cfm>). The reader may access these reviews at <http://doi.org/10.6083.M4H9949P>.

In that previous paper, the FDA's regulatory decisions were previously classified as (1) positive (clearly supporting efficacy) or (2) not positive.⁷ The current study included only positive trials ($\kappa=41$), that is, trials with statistically significant results on the primary end point (s). Since the FDA review for fluoxetine in the treatment of panic disorder became available after the publication of Roest *et al*, this study was able to include one additional positive trial. These 42 trials were covered in 21 FDA reviews (one for each drug indication combination, each including 1–4 trials) (see figure 1). Further details on the number of trials included in each FDA review can be found in supplementary Table 1 of Roest *et al*.⁷

Published papers

Matching journal publications for the 41 trials were identified, as reported in Roest *et al*,⁷ by systematically searching PubMed, EMBASE and the Cochrane Central Register of Controlled Trials without language restrictions, with a search cut-off date of 19 December 2012. The title field was searched for the name of the drug and the type of anxiety disorder and any field was searched for the word *placebo*. The matching paper for the additional (fluoxetine) trial included in this study was found by hand search. One trial was not published. The other 41 trials were described in 38 papers because six trials were published in pooled analyses (see figure 1).

Concerns

For each trial with a matching publication ($\kappa=41$), we scored concerns in that publication and in the FDA review. A concern was defined as an expressed limitation

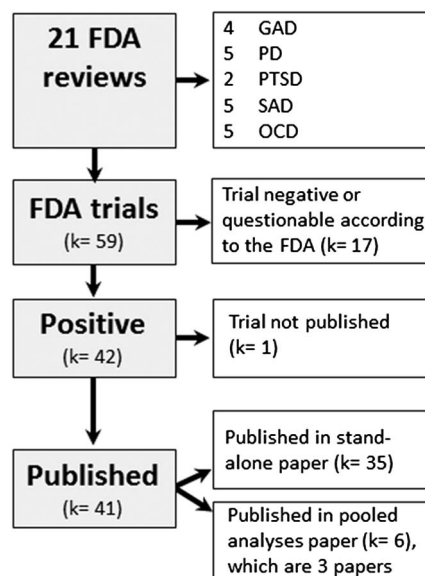


Figure 1 Flow chart of trial and paper selection.

regarding a specific trial. Two independent researchers (AMR and LB) first scored concerns expressed in the FDA reviews. Subsequently, the concerns in the papers were scored independently by researchers LB and BFJ. Discrepancies were resolved by consensus (LB, BFJ, AMR). The inter-rater agreement was expressed in terms of Cohen's κ , which, following Landis and Koch,¹⁹ is regarded as moderate from 0.41 to 0.60, substantial from 0.61 to 0.80 and almost perfect above 0.81.

Concerns in the FDA reviews

Concerns in the FDA reviews were counted when they were expressed by at least two FDA parties (eg, medical reviewer, statistical reviewer, team leader, division director). Only in exceptional cases were concerns counted when they were mentioned only by the medical reviewer: (1) when the statistical review was missing and the concern was also reported in other FDA reviews, or (2) when no detailed evaluation on a specific trial was present for the statistical reviewer and team leader. The concerns found in the FDA reviews were sorted into 11 categories (figure 2).

Concerns in the papers

The FDA concerns were counted as present or absent in the abstract or discussion section of each paper. Additionally, since FDA concerns might not be the only relevant ones, concerns mentioned in the papers but not by the FDA were scored as 'non-FDA' concerns and grouped into categories (see figure 3). Concerns in the papers were counted when they were either mentioned in the limitations section or clearly formulated as a concern. This included any formulation that contained signal words like 'unfortunately' or 'cause for concern', or was framed as a future research question.

Spin

Spin was identified when concerns about drug efficacy expressed by the FDA were absent from the published reports.

RESULTS

FDA concerns

A total of 38 concerns were identified in the 41 trials (89% directly, the rest after debate). Inter-rater reliability was substantial ($\kappa=0.73$). Eleven of these concerns pertained to the lack of evidence for a dose-response relationship. The second and third largest categories of concerns pertained to the lack of statistical significance on the primary end point(s) according to the observed cases (OC) and last observation carried forward (LOCF) analyses, respectively. The other categories occurred only one to three times (see figure 2), but these were more serious concerns, including lack of efficacy in one of the sexes, lack of significant findings for all dose groups and a lower efficacy of the study drug compared with previously approved drugs.

Paper concerns

In our review of the 38 papers, we identified a total of 86 concerns (62%) directly; the rest were debated (inter-rater reliability substantial, $\kappa=0.64$). Almost all (99%) concerns were mentioned in the discussion paragraph of the paper. Four of these concerns were also mentioned in the abstract. One concern occurred in the abstract but not in the discussion section of the paper.

Spin in the papers

In three instances, one FDA concern was present for both trials published in pooled analyses. To keep the comparison between FDA and paper concerns

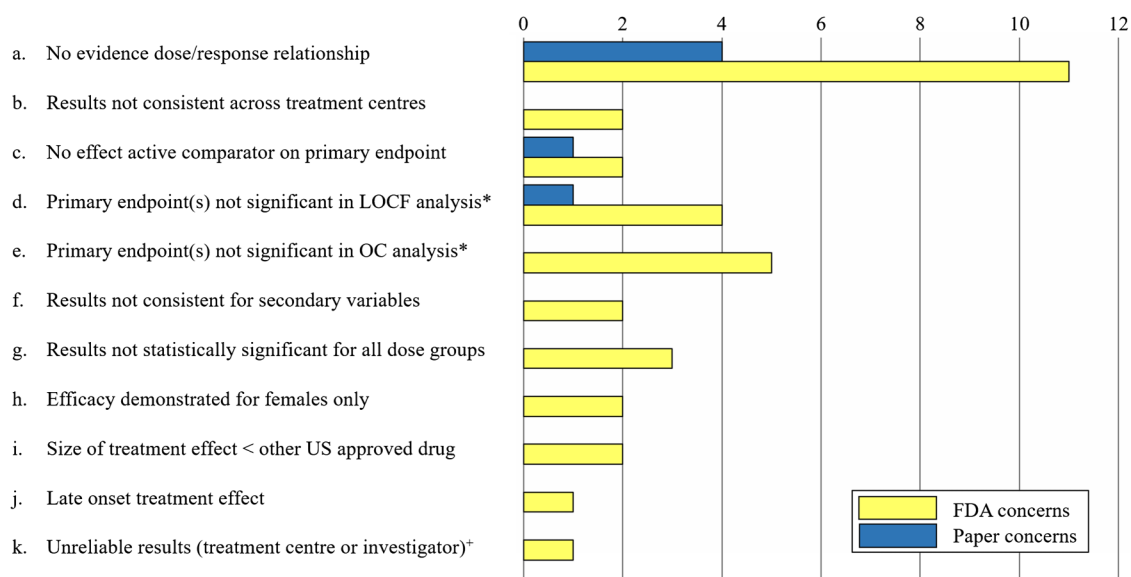


Figure 2 FDA concerns in FDA review packages and related trial papers. *In specific cases, the FDA allowed for multiple primary end points. +Such that the FDA excluded results from analysis. FDA, Food and Drug Administration.

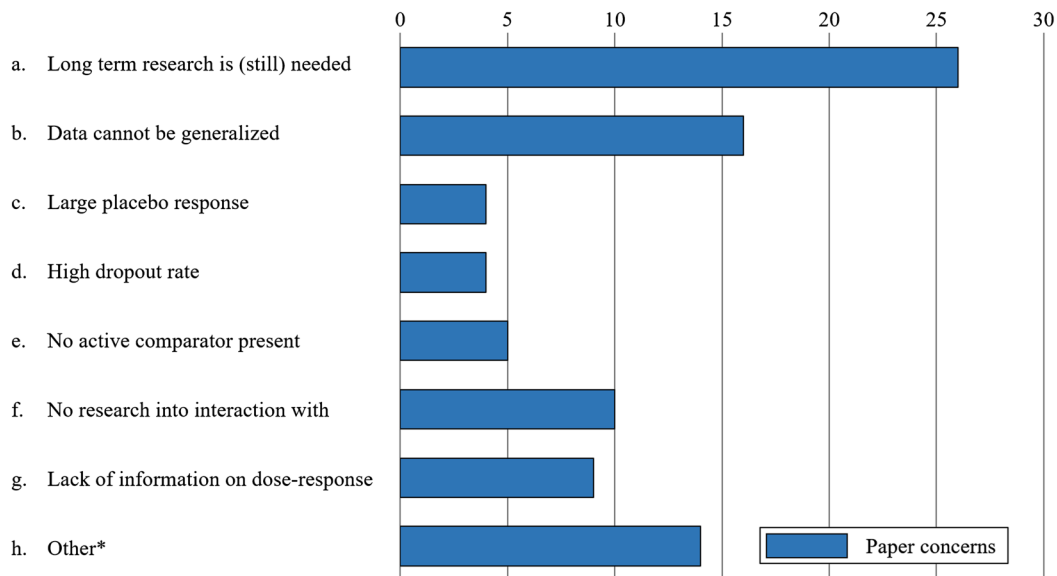


Figure 3 Non-FDA concerns in trial papers. *Concerns that occurred only once were classified as 'Other. FDA, Food and Drug Administration.

conservative, such a pair was counted as a single FDA concern, leaving 35 unique concerns. Only six of these 35 (17%) FDA concerns were found in the trial papers (covering three categories), as outlined in figure 2. The concern category most frequently present in the papers was the lack of a dose–response relationship (four cases in papers vs 11 in FDA reviews). Two papers mentioned a concern that fit the FDA categories, but was not mentioned in the corresponding FDA review (not shown in figure 2). These concerns pertained to the lack of a dose–response relationship and the lack of significance in OC analyses.

Non-FDA concerns in the papers

Six of the concerns mentioned in the papers were also scored by the FDA (7%), and two concerns from the papers fell into the FDA categories. This left 78 additional concerns not mentioned by the FDA, yielding an average of two concerns per reviewed paper. As can be seen in figure 3, the most abundant of the concerns not mentioned by the FDA pertained to the lack of long-term research (68% of papers). The second most common concern focused on the lack of generalisability, usually due to study exclusion criteria. Fourteen concerns occurred only once and were scored as 'other' concerns.

DISCUSSION

This paper demonstrates, within a cohort of positive RCTs for anxiety disorders, the existence of spin. The FDA commonly expressed concerns about drug efficacy, but only a fraction of these (17%) were conveyed in the corresponding journal publications. The published papers mentioned 12 times more non-FDA concerns (n=78) than FDA concerns (n=6), but the former

consisted largely of generic concerns applicable to virtually any RCT, such as a lack of long-term research and limited trial generalisability. Other non-FDA concerns often reflected design choices, such as lack of an active comparator. Publication of such generic concerns, when coupled with a lack of stated concerns about drug efficacy, could itself be interpreted as a type of spin.

Strengths and limitations

The results of this study should be interpreted in the light of the following strengths and limitations. Our definition of spin excluded other types of spin, for example, linguistic spin⁵ and graphical misrepresentation of data,³ so this study arguably presents an overly conservative picture. Nonetheless, we feel that misrepresentation of concerns about drug efficacy, in the context of RCTs with positive results, may be a more important type of spin because of its more direct relevance to clinical practice. Failure to transparently convey efficacy concerns might distort the drug's apparent risk–benefit ratio, leading to undue clinician enthusiasm and potentially inappropriate treatment.

The strength of the study lies in our having undertaken a straightforward comparison using an objective third party. However, we cannot be sure that the FDA reviews included mention of all relevant concerns about drug efficacy, hence our search for concerns in published papers. Also, while the FDA reviewers are objective third-party evaluators of drug efficacy, they did not always report the same concerns. For example, high dropout rates, an issue sometimes mentioned by statistical reviewers, was never raised by medical reviewers or team leaders. Our requirement that specific FDA concerns be expressed in two or more reviews might have led to an underestimation of the total number of FDA

concerns. Finally, a minority of the publications preceded (rather than followed) the FDA approval date, leaving open the possibility that the FDA had not yet made the sponsor aware of its efficacy concerns, which could explain the sponsor's not mentioning them in the paper. However, our aim was not to ascertain whether authors conveyed concerns expressed to them by the FDA, but rather whether they had thoughtfully reflected on their trial's limitations and conscientiously reported them.

Implications

We observed that most FDA concerns about drug efficacy were not mentioned in trial papers, whereas most non-FDA concerns were vague and could apply to virtually any clinical trial. While we agree that the lack of long-term research is a problem, a plethora of non-efficacy-related concerns might serve to obfuscate or distract from issues of greater relevance to clinicians. This raises ethical questions, considering that problems associated with prescription drug use (eg, poisoning) are third on the list of causes of death in Europeans and Americans above age 65, after heart disease and cancer.^{1 20 21} Anxiety disorders have an estimated year prevalence of 12% and antidepressants are the primary pharmacological treatment for anxiety.²¹ About 10% of Europeans currently use SSRIs or SNRIs on a daily basis.^{22–24}

Previous studies have shown that bias and spin occur in many scientific fields.^{3 4 7 8 12–15 17} To further determine the extent to which concerns about drug efficacy are misrepresented in published papers, the method employed here can be applied to other fields. If misrepresentation of concerns proves to be widespread, a number of measures can be taken. One possible approach is to have patient-level data analysed and reviewed by an independent party, as has been carried out for at least one controversial trial.^{25 26} Another is for journal reviewers and editors to require authors of publications to mention FDA concerns, though this could be carried out only for studies published after FDA approval.

CONCLUSION

This study indicates that, in the context of anxiety medication trials with positive primary outcomes, published papers substantially misrepresent concerns about drug efficacy. The majority of concerns raised by the FDA, an independent third party, were not reported in matching publications. The concerns raised in the papers by sponsors, who have a financial interest in the outcome of the trial, were often vague and applicable to clinical drug trials in general. Such concerns may be overrepresented at the expense of issues of greater importance to clinicians, such as (problems with) drug efficacy. Patient-level data reporting, independent review and increased expectation that authors convey FDA concerns could help reduce spin. Before such measures are

implemented, however, this phenomenon should be investigated in related fields.

Author affiliations

¹University of Groningen, University Medical Center Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), Groningen, The Netherlands.

²University of Groningen, Department of Developmental Psychology, Groningen, The Netherlands

³Behavioral Health and Neurosciences Division, VA Portland Health Care System, Portland, Oregon, USA

⁴Departments of Psychiatry and Pharmacology, Oregon Health & Science University, Portland, Oregon, USA

Contributors The study was conceived by AMR, EHT and PdJ. The FDA reports were reviewed by LB and AMR. The papers were reviewed by LB and BFJ. The data were analysed by LB. The initial draft of the manuscript was written by LB and BFJ. EHT, PdJ and AMR edited the manuscript. All authors contributed significantly to the study and approved the final version of the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement FDA reviews used in this study can be downloaded from the following website <http://doi.org/10.6083.M4H9949P>.

Open access This is an Open Access paper distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build on this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Gotzsche PC, Smith R, Rennie D. *Deadly medicines and organised crime: how big pharma has corrupted healthcare*. London: Radcliffe, 2013.
2. Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
3. Al-Marzouki S, Roberts I, Marshall T, *et al*. The effect of scientific misconduct on the results of clinical trials: a Delphi survey. *Contemp Clin Trials* 2005;26:331–7.
4. Boutron I, Altman DG, Hopewell S, *et al*. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *J Clin Oncol* 2014;32:4120–6.
5. Marco CA, Larkin GL. Research ethics: ethical issues of data reporting and the quest for authenticity. *Acad Emerg Med* 2000;7:691–4.
6. Turner EH, Rosenthal R. Efficacy of antidepressants. *BMJ* 2008;336:516–17.
7. Roest AM, de Jonge P, Williams CD, *et al*. Reporting bias in clinical trials investigating the efficacy of second-generation antidepressants in the treatment of anxiety disorders: a report of 2 meta-analyses. *JAMA Psychiatry* 2015;72:500–10.
8. Boutron I, Dutton S, Ravaud P, *et al*. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64.
9. Lazarus C, Haneef R, Ravaud P, *et al*. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Med Res Methodol* 2015;15:85.
10. Ochodo EA, de Haan MC, Reitsma JB, *et al*. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of 'spin'. *Radiology* 2013;267:581–8.
11. Yavchitz A, Boutron I, Bafeta A, *et al*. Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. *PLoS Med* 2012;9:e1001308.
12. Gewandter JS, McKeown A, McDermott MP, *et al*. Data interpretation in analgesic clinical trials with statistically nonsignificant primary analyses: an ACTION systematic review. *J Pain* 2015;16:3–10.

13. Linden A. Identifying spin in health management evaluations. *J Eval Clin Pract* 2011;17:1223–30.
14. Lockyer S, Hodgson R, Dumville JC, *et al*. 'Spin' in wound care research: the reporting and interpretation of randomized controlled trials with statistically non-significant primary outcome results or unspecified primary outcomes. *Trials* 2013;14:371.
15. Patel S V, Chadi SA, Choi J, *et al*. The use of 'spin' in laparoscopic lower GI surgical trials with nonsignificant results: an assessment of reporting and interpretation of the primary outcomes. *Dis Colon Rectum* 2013;56:1388–94.
16. Fletcher RH, Black B. 'Spin' in scientific writing: scientific mischief and legal jeopardy. *Med Law* 2007;26:511–25.
17. Turner EH, Matthews AM, Linardatos E, *et al*. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252–60.
18. Turner EH. How to access and process FDA drug approval packages for use in research. *BMJ* 2013;347:f5992.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
20. Hall AJ, Logan JE, Toblin RL, *et al*. Patterns of abuse among unintentional pharmaceutical overdose fatalities. *JAMA* 2008;300:2613–20.
21. Kessler RC, Aguilar-Gaxiola S, Alonso J, *et al*. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psychiatr Soc* 2009;18:23–33.
22. Mojtabai R, Olfson M. National patterns in antidepressant treatment by psychiatrists and general medical providers: results from the national comorbidity survey replication. *J Clin Psychiatry* 2008;69:1064–74.
23. Organisation for Economic Co-operation and Development (OECD), Health Statistics 2016, pharmaceutical consumption. <http://www.oecd.org> (accessed 7 Oct 2016).
24. Bauer M, Monz BU, Montejo AL, *et al*. Prescribing patterns of antidepressants in Europe: results from the Factors Influencing Depression Endpoints Research (FINDER) study. *Eur Psychiatry* 2008;23:66–73.
25. Keller MB, Ryan ND, Strober M, *et al*. Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial. *J Am Acad Child Adolesc Psychiatry* 2001;40:762–72.
26. Le Noury J, Nardo JM, Healy D, *et al*. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320.