THE COCHRANE COLLABORATION®

Preparing, maintaining and disseminating
systematic reviews of the effects of health care

The Nordic Cochrane Centre
Rigshospitalet, Dept. 7811
Blegdamsvej 9
2100 Copenhagen Ø, Denmark
Tel: +45 35 45 71 12
Fax: +45 35 45 70 07
E-mail: general@cochrane.dk

5 Sept 2014

Dear Professor Erkki Isometsä,

During the discussion at the meeting arranged by the Finnish Psychological Association in Helsinki 1 Sept, I mentioned that the meta-analysis by Fournier in JAMA (1) found that antidepressants worked only for very severe depression, and not for mild, moderate or severe depression.

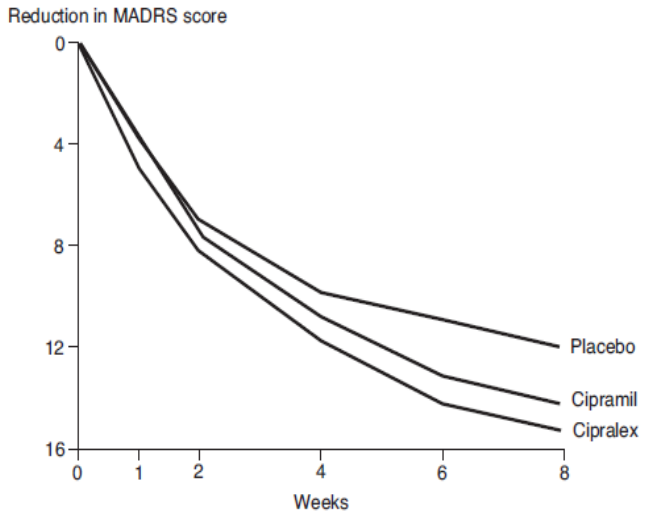
It is actually worse than what I said at the meeting. In patients with very severe depression, Cohen's d was 0.47 (1). The Hamilton 17 scale has an SD of around 3.5 at baseline for patients in depression trials. We have found that the SD is larger after treatment, however. The median SD after treatment was 7.5 in 11 placebo-controlled trials with 37 arms (clinical study reports of SSRIs in our possession, which we obtained from the European Medicines Agency).

Cohen's d of 0.47 = d/7.5, which means that d = 3.5. Thus, the effect of antidepressants in very severe depression corresponds to 3.5 on the Hamilton 17 scale.

NICE in the UK was too generous when it arbitrarily decided that 3 was the least clinically important difference. Stefan Leucht et al. (2) found that the least clinically important difference on the Hamilton 17 scale is 5-6. If we accept this, it means that Fournier showed that antidepressants are ineffective for all severities of depression.

You mentioned at the meeting a larger meta-analysis that found larger effects than Fournier et al. and which, like Fournier's analysis, was also based on individual patient-level data. You kindly gave me the reference to this meta-analysis, which was performed by Gibbons et al. (3). I have the following nine reservations:

1. Linear regression was used although we all know that the course of improvement is far from being linear. It is a curve, and the Hamilton score drops most in the beginning. See, for example, the graph below, which appears in my book on p.225 (4):



This graph shows a Lundbeck meta-analysis of three of the company's own trials. There seems to be a difference between Cipramil and Cipralex, which is also what Lundbeck claims, but this is spurious. In actual fact, the two drugs give the same effect (4,5), which is expected, as the active molecule is the same in both drugs.

2. It is clear that the assumptions for linear regression have been seriously violated in this case. The data points should be contained in an ellipse like pattern, but there are some gross outliers in the figure that must have carried undue weight in the regression analysis. This is elementary knowledge for anyone with a superficial knowledge of statistics.

3. The effects are clinically insignificant. In adult and geriatric trials, the difference on the Hamilton scale was only 2.6. In children, the difference was 4.6 on fluoxetine. This is below the 5-6 that Leucht et al. consider clinically relevant. Furthermore, I believe that SSRIs should never be used in children and adolescents. It makes no sense to use drugs that double the risk of suicide (4,6,7) when what we want more than anything else is to reduce the suicide risk in depressed children and adolescents.

4. A huge amount of data seems to be missing, e.g. "Among all subjects, 52.2% had a score during week 6 and 21.4% had a score on day 42" (3). It is not clear how the authors dealt with this problem, and I cannot see anywhere in their paper any mentioning of multiple imputation, which seems to be the least bias-prone method for missing values. Furthermore, I could not find any information about how patients who dropped out were treated in the analyses, or how many they were, which is crucial for antidepressant trials, as shown by statistician Hans Melander at the Swedish Drug Agency (8).

5. When testing whether the effect was related to baseline severity, Gibbons et al. dichotomized baseline severity according to whether the Hamilton score was above 20 or the CDRS-R score was above 60. This is appallingly poor science. These cutoffs are totally arbitray and we are not told why they were chosen. Further, it would have been much more statistically powerful to use the scores as they were, i.e. "continuous" data, without dividing them in two groups. When Gibbons et al. do not find a relation between the effect and baseline severity, in contrast to researchers who have a more honest approach to their data, it is a finding that should be ignored.

6. The first author, Gibbons, has previously published studies that some commentators consider scientifically dishonest. He reported that in the wake of the black box warning (and a similar warning by European regulatory authorities), the prescribing of SSRIs to children and adolescents decreased in the U.S. and Europe, and that when this happened, there was a dramatic increase in suicides in the two countries he studied, the U.S. and the Netherlands (9). The black box warnings, he concluded, apparently led to an increase in paediatric suicides. In the Netherlands, Dutch academics were incensed with Gibbons and his statistical antics. In the Dutch Drug Bulletin, they noted that the increase in suicides in the Netherlands was so small that it was "not statistically significant."

7. In his meta-analysis, Gibbons cites the STAR*D trial for having found a 67% remission rate for patients "who finished all 4 phases of the trial." The STAR*D trial is considered fraudulent (10,11). The trial was biased by design but even so, its results were so bad that the STAR*D investigators did all sorts of machinations, and there is a whistleblower's lawsuit charging this. For this reason alone it should not be cited without serious reservations, but Gibbons had none. What really happened in

this trial was that only 38% of the patients properly enrolled remitted during one of the four stages of drug treatment and that only 3% of the patients remitted and then stayed well throughout the 12-month follow-up (108 of 4,041 patients) (11). The remaining patients either failed to remit, relapsed during the follow-up, or dropped out. Thus, Gibbons citation is so misleading that it can hardly by worse than this.

8. Gibbons et al. claim in their meta-analysis that what they found was "an enormous difference." They also say that "it seems unlikely that reliance on industry-sponsored studies produced biased results" (3). This contradicts everything we know about drug industry supported antidepressant trials (4,6). I believe such statements speak volumes about the author team's a priori biases, which also so clearly influenced the substandard way they analysed their data.

9. The true effect of antidepressants is much less than what is measured in placebo controlled drug trials. The lack of effective blinding due to the drugs' side effects can explain what is being measured and it likely means that they have no true effect at all (4,6,7). Which means that we don't even need to consider what difference on the Hamilton scale is clinically relevant.

References

1. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity. A patient-level meta-analysis, JAMA 2010;303:47–53.

2. Leucht S, Fennema H, Engel R, Kaspers–Janssen M, Lepping P, Szegedi A. What does the HAMD mean? Journal of Affective Disorders 2013;148:243–8.

3. Gibbons RD, Hur K, Brown H, Davis JM, Mann JJ. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. Arch Gen Psychiatry. Published online March 5, 2012. doi:10.1001/archgenpsychiatry.2011.2044

4. Gøtzsche PC. Deadly medicines and organised crime: how big pharma has corrupted health care. London: Radcliffe Publishing, 2013.

5. Alkhafaji AA, Trinquart L, Baron G, Desvarieux M, Ravaud P. Impact of evergreening on patients and health insurance: a meta analysis and reimbursement cost analysis of citalopram/escitalopram antidepressants. BMC Medicine 2012;10:142.

6. Healy D. Let Them Eat Prozac. New York: New York University Press; 2004.

7. Gøtzsche PC. Why I think antidepressants cause more harm than good. Lancet Psychiatry 2014;1:104-6.

8. Melander H, Ahlqvist-Rastad J, Meijer G, et al. Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 2003;326:1171–3.

9. Whitaker R. The triumph of bad science. http://www.madinamerica.com/2012/07/the-triumph-of-bad-science/. 11 July 2012 (accessed 12 July 2012).

10. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants: current status of research. Psychother Psychosom 2010;79:267-79.

11. Whitaker R. The STAR*D scandal: a new paper sums it all up. http://www.madinamerica.com/2010/08/%EF%BB%BFthe-stard-scandal-a-new-paper-sums-it-all-up/. 27 Aug 2010 (accessed 5 Sept 2014).

Best wishes

Peter C Gøtzsche
Director, The Nordic Cochrane Centre, and
Professor in Clinical Research Design and Analysis, the University of Copenhagen