

Restoring the two pivotal fluoxetine trials in children and adolescents with depression

Peter C. Gøtzsche^{a,*} and David Healy^b

^a*Institute for Scientific Freedom, Hørsholm, Denmark*

^b*Department of Family Medicine, McMaster University, Hamilton, Canada*

Received 11 June 2021

Accepted 12 May 2022

Abstract.

BACKGROUND: Fluoxetine was approved for depression in children and adolescents based on two placebo-controlled trials, X065 and HCJE, with 96 and 219 participants, respectively.

OBJECTIVE: To review these trials, which appear to have been misreported.

METHODS: Systematic review of the clinical study reports and publications. The primary outcomes were the efficacy variables in the trial protocols, suicidal events, and precursors to suicidality or violence.

RESULTS: Essential information was missing and there were unexplained numerical inconsistencies. (1) The efficacy outcomes were biased in favour of fluoxetine by differential dropouts and missing data. The efficacy on the Children's Depression Rating Scale-Revised was 4% of the baseline score, which is not clinically relevant. Patient ratings did not find fluoxetine effective. (2) Suicidal events were missing in the publications and the study reports. Precursors to suicidality or violence occurred more often on fluoxetine than on placebo. For trial HCJE, the number needed to harm was 6 for nervous system events, 7 for moderate or severe harm, and 10 for severe harm. Fluoxetine reduced height and weight over 19 weeks by 1.0 cm and 1.1 kg, respectively, and prolonged the QT interval.

CONCLUSIONS: Our reanalysis of the two pivotal trials showed that fluoxetine is unsafe and ineffective.

Keywords: Fluoxetine, depression, children, placebo-controlled trials

1. Introduction

Fluoxetine was approved for depression in children and adolescents in the United States in 2002 based on two placebo-controlled clinical trials, even though a statistical review for the Food and Drug Administration (FDA) had noted there was not a statistically significant benefit for the drug on the primary outcome in either trial [1].

In 2004, the FDA issued a black box warning that all antidepressants may increase the risk of suicidal thinking and behaviour in children and adolescents. An increase in suicidal events is also seen in adults. A meta-analysis of placebo-controlled trials in healthy adult volunteers using precursor events defined by the FDA found that SSRIs and SNRIs double the risk of harms related to suicidality and violence, and the number needed to treat to harm one healthy person was only 16 (95% confidence interval 8 to 100) [2].

About half of the suicides are missing in published trials of psychiatric drugs [3], and suicidal events are often called something else, e.g. emotional lability, hospital admission or depression [4–6].

* Address for correspondence: Peter C. Gøtzsche, Institute for Scientific Freedom, DK-2970 Hørsholm, Denmark. E-mail: pcg@scientificfreedom.dk.

Neither of the peer-reviewed publications [7,8], described the suicidal events that were mentioned in the Clinical Study Reports (CSRs), X065 and HCJE, that Eli Lilly submitted to the drug regulators for marketing approval [9,10]. Furthermore, Lilly concluded that fluoxetine was effective, even though the FDA found that both studies were negative on their primary endpoint.

We therefore set out to review and restore the public record of the trials [4–6].

2. Methods

We obtained the CSRs for the two trials from the UK Medicines and Healthcare products Regulatory Agency. As per RIAT methodology, we asked Eli Lilly and Graham Emslie, the primary investigator, if they wanted to restore the trials [11]. We did not hear from Emslie. Lilly did not believe that “any additional analyses are needed at this time”.

The CSRs for X065 and HCJE are 1008 and 2549 pages, respectively [9,10]. There were redactions, mostly of names of people, funders, and institutions. Many sections were empty even though indexes suggested otherwise. The empty sections should have included psychiatric histories, efficacy data, adverse events and electrocardiogram data, but they consisted of one page with the text: “Please see SAS Transport file located in Item 11 of this submission,” which was not present.

The X065 report had 159 pages with individual patient data for abnormal laboratory values. The HCJE report had 535 pages with laboratory data and individual patient data for blood concentrations of fluoxetine and norfluoxetine, with details on weight, age, height, body mass index, ethnic origin, and smoking, alcohol, and caffeine intake.

Many pages were missing. The index in X065 had entries for pages 833 and 869 but failed to note that pages 841 to 868 did not exist. In HCJE, a secondary index on page 1471 did not suggest anything was missing because there were no page numbers, only section headings, but 470 pages were missing.

The missing materials mean that a full RIAT restoration is not possible.

The primary outcomes were the efficacy variables listed as primary in the trial protocols, as well as suicidal events and precursors to suicide or violence. We compared patient reported with investigator reported outcomes. We focused on blinding, withdrawal effects in patients who had an antidepressant drug discontinued before randomisation, and concomitant use of drugs with sedative properties that might obscure harms like agitation on fluoxetine [4,5].

One investigator (PCG) extracted terms for all adverse events data, sorted them alphabetically, and presented them to the other investigator (DH) who decided blindly, without knowing if they had occurred on fluoxetine or placebo or how many there were, whether they could be considered precursors to suicidal or violent events. The response options were no, possibly, and yes, with a comments field.

We used Fisher’s exact test for proportions. We did not perform meta-analyses because the efficacy data were biased.

3. Results

3.1. Results for trial X065

This was a single-centre investigator-initiated trial conducted in Texas from 10 April 1991 to 28 February 1995 and published in 1997 [7]. The study report from 2000 contains the investigator’s original protocol and a protocol revised by Lilly [9]. Lilly’s statistical analysis plan was written *a posteriori*.

The trial included 96 outpatients (44 females) aged 7 to 18 years (mean 12.8) with major depressive disorder. The average duration of the current episode was 14 weeks. Patients had to have a score above 40 on the Children's Depression Rating Scale-Revised (CDRS-R), a 17-item scale with a scoring range from 17 to 113. The patients were treated for 8 weeks with fluoxetine 20 mg once daily or placebo.

3.1.1. Randomisation and blinding

Randomisation was stratified by gender and age [9]. There was no information on block size. Treatment assignment was carried out by a local pharmacist using a list prepared by the biostatistician. A study site nurse verified the assignment was correct by comparing dispensed medication to the list. This suggests the allocation was not concealed.

The pharmacy initially blinded the drugs by emptying Prozac capsules for the placebo group and refilling them with lactose powder [9]. From August 1993, Lilly supplied blinded drugs in white capsules.

The hospital pharmacy informed the chemistry labs about treatment assignment "to avoid running unnecessary blood levels [9]". A study site nurse who checked the study medication liaised between the clinical site and the pharmacy and provided psychiatric ratings for two patients. Two patients who attempted suicide on fluoxetine "may have had their treatment assignments revealed," and there were "a few instances ... (less than ten)" where study records indicated physicians were unblinded prior to study completion. For patients who dropped out or became depressed, the treating clinician had access to their treatment assignment [9].

3.1.2. Outcomes

There were two primary psychiatrist rated outcomes: improvement on the Clinical Global Impressions (CGI) scale and on the CDRS-R [7]. These were reduced in the published article to binary outcomes, very much or much improved on the CGI scale, and a CDRS-R ≤ 28 . The CSR had only one primary outcome, at least a 30% reduction from baseline on the CDRS-R, but there were three success criteria: remission (CDRS-R ≤ 28); response (CGI-Improvement score of 1 or 2); and recovery (both criteria) [9].

Secondary outcomes reported in the CSR or published article included:

- (1) Survival analysis of remission on the CGI scale [7]
- (2) Analyses of variance (ANOVA) on weekly CDRS-R scores with the last observation carried forward (LOCF) [7]
- (3) Analyses of observed cases [9]
- (4) Change in CDRS-R scores using linear regression on available data [7]
- (5) Analysis of covariance using available data [7,9]
- (6) CGI-Improvement [9]
- (7) CGI-Severity [9]
- (8) Brief Psychiatric Rating Scale for Children (BPRS-C) [9]
- (9) Children's Depression Inventory (CDI) for those below 13 years of age and Beck Depression Inventory (BDI) for those aged 13 and above [9]
- (10) Weinberg Screening Affective Scale (WSAS) [7]
- (11) Bellevue Index of Depression (BID), parent and patient versions [9]
- (12) Children's Global Assessment Scale (CGAS) [9]
- (13) Family Global Assessment Scale (FGAS) [9].

Subgroup analyses included:

- (1) Analyses by gender, age and origin of trial drugs [9]
- (2) 50% reduction in CDRS-R [9]
- (3) 30% and 50% reductions for patients completing 4 weeks of treatment [9]
- (4) Analyses of all 21 items on BPRS-C [9]
- (5) Analyses of all 17 items on CDRS-R [9]
- (6) Analyses of sums of 3-6 items on CDRS-R called Mood Subtotal, Somatic Subtotal, Subjective Subtotal and Behaviour Subtotal (not listed in the protocol) [9]
- (7) A repeated measures ANOVA where the dependent variables were the baseline and postbaseline CDRS-R scores [9].

Adverse events were collected by asking the patients if they had any of 32 symptoms on a Side-Effects Checklist (described as 30 but there were 32), or any of 30 symptoms on the Fluoxetine Side-Effects Checklist (starting in January 1993), and by collecting non-solicited adverse events [9]. Adverse events were coded according to COSTART (Coding Symbols for a Thesaurus of Adverse Reaction Terms) by blinded Lilly personnel. Subgroup analyses of adverse events were performed for age, gender, and provider of trial drugs.

3.1.3. Washout phase

The patients were randomised after a one-week washout period on single-blind placebo [9]. As 5 fluoxetine versus 9 placebo patients were on tricyclic antidepressants [9], the risk of withdrawal symptoms was greatest for patients randomised to placebo.

3.1.4. Effect of fluoxetine on depression

The efficacy analyses favoured fluoxetine. After 4 weeks, 6 patients had discontinued on fluoxetine and 12 on placebo [9]. As most analyses used the last observation carried forward (LOCF) method, more patients on placebo than on fluoxetine had high depression scores carried forward.

After 8 weeks, the difference from baseline in CDRS-R was 9.7 larger on fluoxetine than on placebo using the LOCF method. It was 2.5 using observed cases on one graph and 6.0 on another graph on the same page with no explanation for this discrepancy.

The published article left out results for BID and FGAS. The CSR left out both of these results and those for WSAS and CGAS, stating the self-report scales were not collected consistently (but providing no evidence for this) and because “the data came from relatively unvalidated scales”.

Lilly did not explain discrepant results for BPRS-C in the CSR and the publication, e.g. on fluoxetine, the average score after treatment was 38.9 and 18.0 respectively, and the *P*-values were also markedly different, *P* = 0.32 (calculated by us) versus *P* = 0.09 (Table 1).

The published article combined the results for the CDI and BDI scales, but all averages were smaller than the weighted averages for the CDI and BDI scales in the CSR (see Table 1), which is a mathematical impossibility.

Despite the bias introduced by using exit values, there were no significant differences in BPRS-C (*P* = 0.32), CGAS (*P* = 0.18), or in outcomes reported by the patients, CDI/BDI (*P* = 0.58) and WSAS (*P* = 0.17) (Table 1) [7]. Emslie et al. stated that, “given the wide variability of initial child self-reports, these findings are difficult to interpret [7]”. However, the coefficients of variation (standard deviations divided by the means) were *smaller* for BPRS-C and CGAS than for CDRS-R (Table 1).

Table 1

Discrepancies between results reported in tables in the published report [7] and in the study report [9] for trial X065. *P*-values for published data calculated by us based on final scores; other *P*-values calculated by Lilly based on changes from baseline (SD)

Scale	Fluoxetine		Placebo		<i>P</i> -value
	Baseline	Final	Baseline	Final	
CDRS-R [7]	58.5 (10.5)	38.4 (14.8)	57.6 (10.4)	47.1 (17.0)	0.009
CDRS-R [9]	59.9 (10.4)	38.7 (14.6)	57.5 (10.4)	47.0 (16.9)	0.002
CDI/BDI [7]	15.8 (10.6)	9.9 (12.0)	15.3 (11.9)	11.2 (10.8)	0.58
CDI [9]	16.1 (10.5)	10.3 (10.4)	17.8 (13.2)	13.4 (8.6)	0.59
BDI [9]	17.6 (12.0)	11.7 (13.6)	14.8 (9.6)	9.6 (10.3)	0.80
WSAS [7]	20.6 (11.8)	13.1 (12.0)	20.6 (12.8)	16.7 (13.5)	0.17
WSAS [9]	Data missing				
BPRS-C [7]	47.3 (7.7)	38.9 (10.0)	46.2 (8.9)	41.0 (10.4)	0.32
BPRS-C [9]	26.3 (7.9)	18.0 (9.9)	24.7 (8.5)	20.0 (10.1)	0.09
CGAS [7]	47.9 (8.3)	63.9 (12.9)	48.4 (7.8)	60.1 (14.8)	0.18
CGAS [9]	Data missing				

The CSR provided *P*-values for all individual 21 items on the BPRS-C and emphasized that fluoxetine increased hyperactivity significantly ($P < 0.001$) [9]. Hyperactivity is a warning signal for suicide and violence [2], but Lilly introduced the opposite concept: “A decrease of hypoactivity... which is associated with the patients returning to normal functioning at greater rates than placebo-treated patients” transforming a potentially harmful result into a benefit. There were no data in support of this interpretation and Emslie et al. stated: “whether long-term treatment would result in the amelioration of school, general functioning, or concurrent comorbidities is unknown [7]”.

No statistical adjustments can substitute missing data reliably. We therefore focussed on those patients with minimal symptoms after 8 weeks. Only 15 versus 11 patients had minimal symptoms ($P = 0.49$, our calculation), defined by Lilly as CDRS-R ≤ 28 , and only 14 versus 9 had recovered ($P = 0.34$), defined as CDRS-R ≤ 28 and a CGI Improvement score of 1 or 2 [9]. Even these results may be biased in favour of fluoxetine, as spontaneous remission is common and more data were missing for patients on placebo than for patients on fluoxetine (23 versus 15 had dropped out by week 8).

3.1.5. Serious adverse events and discontinuations

The information on discontinuations and suicide attempts was inconsistent. In the published report [7], 36 patients discontinued, in the CSR 38 [9]. The numbers, reasons, and timing were the same in only 5 of 28 cells with information (Table 2). There was no explanation for these discrepancies [9].

No patients died. Two patients attempted suicide on fluoxetine, after 12 and 15 days, respectively [9]. These suicide attempts were left out from the published report [7]. One of them led to study discontinuation after 15 days, and it seems that Emslie et al. called this a “protocol violation” (see Table 2). Adverse events definitely or possibly predisposing to suicide in this patient according to our criteria were manic reaction, insomnia, and nervousness (possibly akathisia) [9]. Predisposing events for the other patient were anxiety, depression, neurosis, thinking abnormal, asthenia, and also hyperkinesia (features of akathisia).

Four additional patients discontinued fluoxetine because of adverse events called “minimal” in the published report, even though three of them developed manic symptoms and the fourth had a severe

Table 2
Discontinued patients in the published article versus the CSR (numbers in square brackets) for trial X065

Reason for discontinuation	Week since randomisation								Total	[Total]
	1	2	3	4	5	6	7	8		
Fluoxetine										
Lack of efficacy				[1]	1 [2]	3 [2]	2	1 [1]	7	[6]
Adverse event			[2]	1 [2]	2	[1]	1		4	[5]
Protocol violation			1	[1]	1 [2]	1			3	[3]
Physician decision							[1]		0	[1]
Total									14	[15]
Placebo										
Lack of efficacy	1	[2]	1 [1]	1 [5]	5 [5]	5 [5]	5 [1]	1	19	[19]
Adverse event				1					1	[0]
Protocol violation		1 [1]	[1]	1					2	[2]
Patient decision	[2]								0	[2]
Total									22	[23]

rash [7]. The CSR summary described two patients with hypomania, one with increased impulsivity, and one with rash [9]. The narratives on these patients were extremely brief, between 17 and 35 words [9]. One also had an anxiety attack and the other had hyperkinesia. The patient with increased impulsivity had hyperkinesia, anxiety and depression but was coded as personality disorder.

The CSR stated that “No placebo-treated patients discontinued due to adverse events” while the published article mentioned one such patient but not the event that caused discontinuation. In 2004, Emslie et al. published additional data, mentioning these two suicide attempts but ascribing one of them to the placebo group, referenced as “Emslie, personal communication [12]”. They also claimed that another patient on placebo discontinued because of mania even though this patient had received fluoxetine.

The CSR had more patients dropping out on placebo than on fluoxetine due to lack of effect (19 versus 6, $P = 0.005$) but fewer patients dropping out due to adverse events (0 versus 5, $P = 0.056$). Only one cause was assigned for each patient [7,9]. Adding both reasons, 19 versus 11 patients dropped out ($P = 0.12$, our calculation).

Concomitant drugs with sedative properties, often used to manage adverse events linked to suicide and violence, were mentioned for 11 patients on fluoxetine and for 3 patients on placebo.

3.1.6. Other adverse events

One patient on fluoxetine and three on placebo did not experience any of the 32 adverse events listed in the Side-Effects Checklist [9].

The data on non-solicited adverse events and events on the Fluoxetine Side-Effects Checklist were not included in the main text of the CSR but came after “Discussion and Overall Conclusions” under the heading “Tables, Figures, and Graphics Not Included in Other Sections [9]”.

In the table of non-solicited adverse events, Lilly provided 29 P -values and wrote that there were no statistically significant differences but without noting that more patients on fluoxetine than on placebo experienced one or more events ($P = 0.051$, Lilly’s P -value) [9].

In contrast, a table showing 28 of the 30 items listed in the Fluoxetine Side-Effects Checklist did not present a single P -value even though there were conspicuous differences. For example, 32 fluoxetine versus 18 placebo patients experienced at least one adverse event ($P = 0.008$), 19 versus 6 experienced restlessness ($P = 0.005$), 9 versus 1 had nightmares ($P = 0.02$), and 7 versus 4 felt tense inside. Restlessness, including feeling tense inside, and nightmares increase the risk of suicide and violence [4,5]. Lilly argued that this checklist was not used consistently throughout the study and was not administered to all randomised patients but did not provide information on the number of missing values. There were no data on the severity of the adverse events even though it was assessed [9].

The published article did not present any data obtained in one of the three ways just described. Only 49 of the 1099 words (4%) in the Results section were related to safety and they were only about discontinued patients [7].

The published report stated electrocardiograms were taken at baseline and after 4 and 8 weeks, but according to the CSR only at baseline and “were not analyzed statistically”. The data were not reported in either the published article or the CSR.

3.2. Results for trial HCJE

HCJE was a multicentre trial conducted by Eli Lilly in the United States from 27 April 1998 to 16 December 1999 [10]. The trial included 219 outpatients (108 females) aged 8 to 17 years (mean 12.7) with major depressive disorder with an average episode duration of 61 weeks, and a score above 40 on the CDRS-R plus a rating of at least moderate on the CGI-Severity scale.

The patients were treated for one week with fluoxetine 10 mg once daily (109 patients) or placebo (110 patients), followed by 8 weeks on 20 mg or placebo (acute treatment phase) and by another 10 weeks where non-responders to 20 mg fluoxetine were randomised to 20 mg or 40 mg fluoxetine (which could be increased to 60 mg 4 weeks later) while the remaining patients continued on 20 mg fluoxetine or placebo. The terminology was inconsistent. The 19 weeks were called the “subchronic treatment phase” even though it included the acute phase, and the protocol mentioned an acute 19-week treatment phase and a subsequent 32-week relapse prevention phase where patients with a CDRS-R ≤ 28 at 19 weeks could be re-randomised to continue current treatment or be switched to placebo.

3.2.1. Randomisation and blinding

Randomisation was stratified by gender and age using a computer-generated randomisation sequence. There was no information on block size. Sealed codes were available, which risks violating allocation concealment.

The medication was packaged in blister cards with a package number. All patients received three capsules daily containing 10 or 20 mg fluoxetine or matched placebos. Patients who did not tolerate fluoxetine could have their dose reduced from 20 to 10 mg or from 40 to 20 mg.

It is not clear if the blinding was maintained after the first 9 weeks when patients on placebo and responders on 20 mg fluoxetine continued with their drug while non-responders on fluoxetine were re-randomised to 20 or 40 mg fluoxetine.

“A minimal number of Lilly personnel” would see the randomisation table and codes before the study was complete. This minimum included statisticians, regulatory scientists, systems analysts, people working with pharmacokinetics, clinical laboratory medicine personnel, and medical writers, all of whom had unblinded access to data during the study. Patients dropping out could receive rescue therapy from

a physician who was not blinded but “agreed to maintain the study blind from study personnel”. Lilly reported to the FDA that in their review of the source records it was not uncommon to see notations defining the patient’s blinded treatment, or in some cases to find fluoxetine plasma concentration results [13].

3.2.2. Outcomes

All study objectives centered on efficacy; none on safety. The primary outcome, CDRS-R, was dichotomized into at least a 30% reduction from baseline.

Numerous secondary objectives, outcomes and comparisons were described and what was reported was not always what was planned.

In the protocol, the secondary objectives “are as follows” whereas in the CSR, they “included the following,” suggesting there could have been more than those listed, which was indeed the case. CDRS-R remission rates and CGI response rates were exploratory “additional analyses” in the protocol and would be “conducted as deemed appropriate,” but in the CSR they were “also analyzed,” which term Lilly used for pre-planned analyses. Thus, protocol defined exploratory analyses acquired a status of secondary outcomes in the CSR.

Rating scales and analyses were changed. Mean score in the protocol became mean change for seven outcomes in the CSR and in another place in the protocol, and subgroup analyses by age became subgroup analyses by age, gender, and family history of depression.

Efficacy outcomes were:

- (1) Mean change in CDRS-R
- (2) “Evaluation of CDRS-R”
- (3) CGI-Severity
- (4) CGI-Improvement
- (5) Montgomery-Asberg Depression Rating Scale (MADRS)
- (6) BDI
- (7) CDI
- (8) Global Assessment of Functioning (GAF)
- (9) Current Functioning
- (10) CGI-Efficacy Index
- (11) Hamilton Anxiety Rating Scale (not mentioned in the amended protocol)
- (12) Kiddie Schedule for Affective Disorders and Schizophrenia–Present and Lifetime (K-SADSPL) (Affective Disorders module only), which was Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) in the protocol.

Clinician rated outcomes were recorded at every single visit after randomisation (11 times) whereas the patient rated outcome (CDI or BDI, depending on age) was only recorded at 5 visits.

Non-solicited adverse events were collected by questioning the patients about the presence of adverse events “in a non-directed manner,” with no further instructions. A form for registering all adverse events that had occurred since the last visit and all clinically relevant abnormalities found on the physical exam or ECG took up one page. It had a 4.5 cm² box for each narrative, and three severity codes (mild, moderate, and severe) that were not defined anywhere. The forms for benefits took up 10 pages. There was an optional one-page form for comments on benefits or aberrant laboratory values, but not adverse events.

Solicited adverse events were captured with the Side-Effects Checklist. Verbatim terms were coded as COSTART terms by blinded Lilly personnel. The Fluoxetine Side-Effects Checklist developed by Lilly and used in trial X065 was not used.

Blood pressure, heart rate, height and weight were measured. The text noted that ECGs were taken at various times during the study. A schedule of events noted that they were taken only at the first visit and after 19 weeks, but a table of missing measurements showed that only 41 recordings had been missed, at various numbered visits throughout the whole trial.

3.2.3. Statistical analyses

A large number of analyses were carried out that were not prespecified in the protocol or in the statistical analysis plan, even though it was updated on 7 February 2000, two months after the 19 weeks trial was over, when a lot of Lilly staff had seen the unblinded values.

We found problematic uses of statistics and contradictory information. It was stated that all patients with at least one post-randomisation visit would be included in the efficacy analyses, but this did not apply to the primary efficacy analyses for which two visits were needed, including “at least 1 week taking fluoxetine 20 mg/day”.

Contrary to the company’s claim to have conducted intention-to-treat analyses, some patients with post-randomisation visits were excluded and the patients who were discontinued were not assessed at the end of the trial. As it was stated patients would stay in the groups to which they had been randomly assigned even if they did not follow the protocol, it made no sense to discontinue patients, the only effect of which was to bias the study results in favour of fluoxetine.

CSR analyses included:

- (1) 10%, 20%, 30%, 40%, 50%, 60%, and 70% reductions in CDRS-R (only 30% was prespecified in the protocol)
- (2) Change in CDRS-R from baseline to each subsequent visit
- (3) CDRS-R remission rate
- (4) Analyses of observed cases
- (5) ANOVA on CDRS-R
- (6) Change in CDRS-R subtotal
- (7) Change in CGI-Severity
- (8) Change in BDI
- (9) Change in CDI
- (10) Change in MADRS
- (11) Change in HAMA
- (12) Change in GAF
- (13) CGI-Improvement
- (14) ANOVA on endpoint values
- (15) ANOVA on changes
- (16) ANOVA with treatment, investigator, gender, age category, and treatment-by-investigator interaction
- (17) ANOVA using a mixed-model approach with the dependent variables being baseline and postbaseline CDRS-R and independent factors being treatment, investigator, treatment-by-investigator interaction, visit, and treatment-by-visit interaction
- (18) CGI-defined response rate
- (19) Recovery

- (20) Percentage of patients with a diagnosis of depression using the K-SADS-PL at endpoint
- (21) Analyses on both the original and rank-transformed data
- (22) Logistic regression models with treatment, investigator, gender, age category, and treatment-by-investigator interaction for comparing percentages.

Later in the CSR, there were additional analyses, e.g. for 4 CDRS-R subtotals and for each of the 17 individual items. Subgroup analyses were performed on CDRS-R response and mean change; CGI-Improvement; non-solicited and solicited adverse events, for children and adolescents separately, males and females, and for patients with and without a family history of depression. There were also results, with *P*-values, for each of 15 investigator sites for CDRS-R, CGI-Improvement, and MADRS.

Furthermore, “After reviewing some of the results of this study, additional exploratory analyses were performed. These included additional subgroup analyses and “completers” analysis of laboratory, ECG, and vital signs data... A second set of blinded ECG readings performed by a pediatric cardiologist supplemented the original readings”.

Even though only 34% of the patients completed all 19 weeks, the analyses after 19 weeks were called interim analyses. All the analyses performed after 9 weeks were repeated after 19 weeks along with additional ones, e.g. including those patients who received “at least 4 weeks of fluoxetine 20 mg/day”. As nothing was said about the placebo group, it is not clear if the same criteria applied to the placebo group, or if all placebo patients were included, regardless of placebo intake.

A two-page appendix on “patients excluded from the efficacy analysis” appeared 90 pages after the statistical analysis section ended, on page 2356. It was a table with no explanatory text listing 9 patients on placebo with their patient numbers and CDRSTL17 values.

CDRSTL17 values were likely CDRS-R scores at baseline, since the visit number corresponded to the randomisation visit for all 9 patients. But the table indicated that all 9 patients were excluded at the randomisation visit, which was true for only one of them. Scattered around in the report, information showed that 3 of the patients were discontinued after 7–10 days. The remaining 5 patients were not discontinued. According to the protocol and CSR, only 4 patients on placebo should have been excluded from the primary efficacy analysis, but another 5 were excluded, with no explanation why, while no fluoxetine patients were excluded. The difference between 0 and 9 exclusions is highly unlikely to have occurred by chance ($P = 0.003$, our calculation).

3.2.4. Washout phase

There was a two-week diagnostic evaluation period starting at week -3 (also called a “no drug” period in a figure), but it was not explained if previous antidepressant treatment was discontinued or if new drugs were not allowed. In a table of protocol violators 2405 pages later, a patient received the last dose of sertraline 12 days before the first visit, which should have been 14 days before. Only in this table could we see that it must have been a requirement to discontinue previous antidepressants 5 weeks before randomisation.

A one-week single blind placebo washout period started at week -1, but week -3 was the baseline for assessing the so-called placebo response at week 0. This did not make sense because placebo was not given before week -1.

Previous antidepressants were listed 6 times for fluoxetine and 6 for placebo, but as patients, drugs and adverse events were not linked, it was impossible to know if any of the serious adverse events in the introductory three-weeks period were iatrogenic withdrawal effects. Serious adverse events occurred in

Table 3
Patients discontinued from the trials according to reasons

	Trial X065 (8 weeks)		Trial HCJE (9 weeks)	
	Fluoxetine	Placebo	Fluoxetine	Placebo
Adverse event	5	0	5	9
Lack of efficacy	6	19	5	12
Patient decision	0	2	3	11
Physician decision	1	0	1	0
Protocol requirement	3	2	4	3
Lost to follow-up	0	0	1	7
Total	15	23	19	42

three patients who were not randomised: At week -3, one patient had explosive aggression, and another had psychotic symptoms; the third patient had suicidal ideation at week -1 coded as depression.

3.2.5. Concomitant medication

Judged by the clinical report forms, patients were not routinely asked which drugs they were currently on. After 447 pages of blank report forms, there was a form about concomitant medication investigators could use “at entry and during the study,” but as it was not obligatory to use it, information on other drugs used during the study was not reliable.

Concomitant medication was reported to have been used by more patients on fluoxetine than on placebo, 82% versus 66% during the first 9 weeks ($P = 0.01$), and 84% versus 72% during all 19 weeks ($P = 0.03$). Paracetamol was used more often ($P = 0.02$ and 0.04 , respectively).

For drugs with sedative properties, the occurrences were: antihistamines 38 versus 31; sedatives/hypnotics 2 versus 5; antipsychotics 1 versus 0.

3.2.6. Effect of fluoxetine on depression

As in trial X065, all efficacy analyses were biased in favour of fluoxetine because the degree of depression after 9 and 19 weeks was unknown for discontinued patients. After two weeks, none had dropped out on fluoxetine versus 10 on placebo. Most analyses used the LOCF method, but Lilly did not alert its readers to the bias this caused.

FDA’s medical reviewer noted the considerably more dropouts on placebo than on fluoxetine and that the pattern was “rather unusual” because there were more dropouts on placebo than on fluoxetine for adverse events (9 versus 5), patient decision (11 versus 3) and lost to follow up (7 versus 1) [13]. Trial X065 had 0 dropouts for adverse events on placebo versus 5 on fluoxetine, and there were no losses to follow-up (see Table 3). FDA noted that Lilly had provided statements to the effect that their clinical investigators did not receive payment in return for particular results [13].

Many tables and graphs were confusing, with the type of analysis left unstated, and the terminology unclear, e.g. “from baseline to endpoint” could mean observed cases or LOCF.

The primary outcome result was described as: “A strong numerical trend was observed, with 71 (65%) fluoxetine-treated patients meeting response criteria compared with 54 (54%) placebo-treated patients;

however, the difference between treatment groups was not statistically significant ($P = 0.093$) for this 9-week treatment period". This analysis had no results for those 28% of the patients who were discontinued.

The difference in change scores on CDRS-R after 9 weeks was 5.9 in a graph whereas it was 7.2 in a table. In another graph, the difference was 7.2 using LOCF, which suggests that Lilly used LOCF also for the table without saying so.

Despite the flaws, placebo tended to be better than fluoxetine as evaluated by the patients: "Placebo-treated patients exhibited greater numerical reductions in the change from baseline for CDI and BDI total scores compared with fluoxetine-treated patients. The difference between treatment groups was not statistically significant".

In the published article, the lack of an effect of fluoxetine on CDI and BDI was mentioned in these terms: "Given the high percentage of patients in this study who had comorbid ADHD, it may not be surprising that results of the clinician-rated measures were not reflected by the results of the patient-rated scales [8]". However, only 14% of the children had an ADHD diagnosis, and it is not likely that such a diagnosis renders the children's assessment of drug effects unreliable.

The report ended on page 224 but was succeeded by another 2325 pages with additional data. The psychiatrists used a CGI-Efficacy Index with 8 categories to "rate overall therapeutic effect in conjunction with side effects for each patient". They assessed for each patient if the improvement in the depression outweighed any drug harms in terms of their interference with daily activities (which was not defined). Lilly claimed the results indicated that therapeutic effects outweighed any side effects because 58% versus 40% had a favourable score. We combined the data from the 8 categories by subtracting bad outcomes from good outcomes and found that 59% versus 55% had a good outcome ($P = 0.58$).

Fluoxetine was ineffective after 19 weeks for CDRS-R when observed cases were used (the difference to placebo was 2.5, $P = 0.36$). For the LOCF analysis, the difference was significant for all visits, even after the first week when patients on fluoxetine had only received 10 mg, but after 19 weeks, the difference was only 4.8 ($P = 0.02$). In an ANOVA of changes, the difference was 1.7 ($P = 0.13$).

In several cases, we could not understand how additional analyses differed from previous ones and why patient numbers differed when they should have been the same. In contrast to the 9-week results in the report summary (see below), several results were no longer statistically significant after 19 weeks, e.g. for MADRS, CGI-Improvement, response based on CGI-Improvement, recovery based on CDRS-R and CGI-Improvement scores, and CDRS-R mood subtotal. As an example, for CGI-Improvement, $P = 0.03$ after 9 weeks and $P = 0.32$ after 19 weeks.

3.2.7. Serious adverse events and discontinuations

There were narratives for 10% of the patients (6 in X065 and 24 in HCJE for post-randomisation events). As we explain below, these were very brief, unclear and incomplete. There were important errors, and two sets of tables separated by 668 pages needed to be combined.

The CSRs used two sets of terms interchangeably, the verbatim term reported by investigators and the coded term added by the company. Lilly did not specify which terms were used in three very different sets of tables. In one table, the column header was "Event Classification Term," but what was reported was the verbatim term.

No patients died. Two patients had serious adverse events on fluoxetine and four on placebo, but the reporting was brief, opaque, and contradictory, and the terms were not the same across tables describing the same patients. Combining three tables and a figure it turned out that a patient on fluoxetine developed suicidal ideation 89 days before the first visit and was discontinued 70 days after randomisation when the

suicidality had become serious, and the patient was hospitalised. The investigator considered the event possibly related to study drug, but Lilly coded it as depression. The suicidality became serious two weeks after some patients had been re-randomised to 40 mg of fluoxetine, but we could not assess its possible relation to the dose increase, as there was no information on dose. One of the tables indicated that this patient was withdrawn at randomisation and that fluoxetine could therefore not have caused the suicidality.

The other patient on fluoxetine had swollen tonsils coded as pharyngitis and underwent a tonsillectomy. The narrative revealed that the patient suffered from much else including fatigue and irritability.

The four serious adverse events on placebo were kidney infection, abdominal pain/appendicitis, aggressive behaviour coded as hostility, and self-mutilatory behaviour coded as intentional injury. The narrative for the patient with appendicitis used the term viral infection coded as infection, which was confusing, as these data were from the relapse prevention phase of the trial even though the company specified in numerous places that data from that phase would not be included in the CSR.

The patient with aggressive behaviour and the patient with self-mutilatory behaviour had symptoms at the first visit and were discontinued 10 and 37 days after randomisation, respectively, when the symptoms had become serious, and they were hospitalised. We had no information allowing us to judge if the events could be iatrogenic effects caused by withdrawal of an antidepressant before randomisation.

The information on the patient with aggressive behaviour appeared in a table, which mixed nonrandomised patients with patients with serious adverse events and patients withdrawn due to nonserious adverse events. The table header was nevertheless “All randomized patients”. As there were no dates, the patient may not have been randomised.

When self-mutilatory behaviour is coded as intentional injury, it conceals if it is violence towards self or others. A brief narrative of 65 words revealed that this patient had both suicidal ideation and homicidal ideation (coded as hostility) and was hospitalised for these reasons.

During the first 9 weeks, 19 versus 42 patients discontinued the study but Lilly’s report to the Data Monitoring Board two months after the study had been completed described 15 versus 33 discontinuations. These numbers were from week 7 without an explanation as to why the Board did not get the full results. This error was repeated for the 19 weeks results.

Eleven patients in each group discontinued due to adverse events, six of which were for nonserious psychiatric reasons on fluoxetine and one on placebo. The terms for fluoxetine were agitation (after 28 days), elevated mood coded as euphoria (65 days), physical aggression coded as hostility (98 days), hyperactivity coded as hyperkinesia (18 days), mania (32 days), and behavioural disinhibition coded as personality disorder (73 days).

According to the narratives, these patients were more severely affected than the tables suggested. The agitation was severe, and additional events included irritability, fatigue, decreased concentration, insomnia, anger, hearing voices, racing thoughts, mood swings and temper tantrums, which suggested drug induced psychosis and possibly akathisia.

The patient with euphoria was discontinued due to elevated mood, increased irritability, restlessness, increased hyperactivity and impulsivity, pressured speech, tangentiality, flight of ideas, belligerence, and mild looseness of associations. These events predispose to suicide and violence, which one would not suspect from the investigator term elevated mood or the coded term euphoria.

The patient with hyperactivity had an ADHD diagnosis. After 22 days on fluoxetine, the ADHD symptoms had become extreme, which may be drug induced akathisia.

The patient with mania also had irritability, agitation, insomnia, pressured speech, and delusions.

A seventh patient discontinued fluoxetine after 84 days due to endometrial hyperplasia. The 167-word narrative was only about this unusual diagnosis in a 15-year-old girl but the list of symptoms included akathisia.

An eighth patient discontinued fluoxetine after 120 days due to an increase in the severity of migraine. Bad dreams and aggression occurred 1-2 days before randomisation during the placebo washout. During treatment, sleep disturbance, shakiness, bad dreams, and intermittent restlessness coded as akathisia occurred.

A patient on placebo had anxiety at randomisation and discontinued 7 days later. Two days after randomisation this patient developed asthenia, and after seven days, he was “agitated and yelling at his mother. Patient stated that the study medication made him nervous, gave him a headache, and made him sick to his stomach. Patient demanded to be withdrawn from the study”. This patient could have suffered akathisia from withdrawal of prior medication.

In total, 7 fluoxetine versus 3 placebo patients experienced psychiatric adverse events leading to discontinuation, which become 9 versus 3 patients with significant psychiatric events if we add the two patients with akathisia on fluoxetine.

The information in a published paper from 2004, which was only about safety, disagreed markedly with the data in the CSR even though all five authors were from Lilly [14]. The paper stated that 49 versus 47 patients “completed 19 weeks of treatment” but the correct numbers were 40 versus 35. The paper noted that 4 patients on fluoxetine reported an event related to suicide or self-harm, but we found only one such patient in the CSR (who was hospitalized because of suicidal ideation). The three other patients, two with suicidal ideation and one with self-mutilation, were not described.

The paper also noted that 4 patients on placebo reported an event related to suicide or self-harm, but we found only one (who was hospitalized for suicidality and self-mutilation). One patient took 9 placebo capsules instead of 3, one made comments about wanting to die, and the third patient reported suicidal ideation.

These discrepancies suggest the company has access to information related to suicidality left out of the CSR, which stated that there was “no difference in the rate of reported suicide-related events between fluoxetine and placebo [14]”. This contradicted our findings (see Supplementary Table S1).

3.2.8. Other adverse events

Almost all patients experienced a solicited adverse event, 105 on fluoxetine versus 100 on placebo after 9 weeks ($P = 0.17$). The numbers after 19 weeks were 107 versus 101 ($P = 0.06$) in one analysis and 108 versus 102 ($P = 0.04$) in another.

For non-solicited adverse events, there were 94 on fluoxetine versus 80 on placebo after 9 weeks ($P = 0.02$) and 101 versus 87 after 19 weeks ($P = 0.006$).

After 9 weeks, more patients had experienced nervous system events on fluoxetine than on placebo, 35 versus 24 ($P = 0.095$). Lilly did not comment on this although, the $P = 0.093$ for the primary efficacy outcome was called a “strong numerical trend”. After 19 weeks, the difference was significant, 42 versus 28 patients ($P = 0.01$, number needed to harm 6). There were also more events related to the respiratory system, 61 versus 42 ($P = 0.01$) and to special senses, 16 versus 4 ($P = 0.005$).

There were no severity data in the main part of the CSR and Lilly did not define the three severity grades used. In drug trials, the usual definitions are:

- (1) Mild: awareness of sign or symptom, but easily tolerated.
- (2) Moderate: discomfort enough to cause interference with usual activities.

(3) Severe: incapacitating with inability to work or do usual activity.

For drugs that only have symptomatic effects, harms that interfere with usual activities are relevant for an assessment of the balance between benefits and harms. If only one comparison is chosen, it should therefore not be for severe events, as Lilly did, but for events classified as moderate or severe.

Lilly reported 19 versus 15 patients with one or more severe adverse events after 9 weeks ($P = 0.46$), whereas we found 68 versus 56 with moderate or severe events ($P = 0.10$). After 19 weeks, Lilly mentioned that there were not more patients with severe events on fluoxetine than on placebo, 22 versus 18, ignoring that 78 versus 64 patients had moderate or severe adverse events ($P = 0.047$, our calculation, number needed to harm 7).

Lilly downplayed even more the solicited adverse events remarking that “There were no clinically relevant differences in maximum intensity for any treatment-emergent solicited adverse event”. However, using Lilly’s terminology, there was a “strong” trend for severe adverse events after 9 weeks, 64 versus 52 ($P = 0.105$), which became 71 versus 57 after 19 weeks ($P = 0.055$). After 9 weeks, significantly more patients on fluoxetine than on placebo were feeling sleepy, 19 versus 7 ($P = 0.01$), having trouble getting along with parents, 19 versus 9 ($P = 0.045$), and trouble paying attention, 18 versus 7 ($P = 0.02$). Lilly stated that the differences were small and that the adverse events did not lead to discontinuations. However, the differences were about 10%, which means that for every 10 patients treated with fluoxetine, one was severely harmed. Nine versus 5 patients had severe problems with sitting still, which Lilly did not comment on although it could mean akathisia.

Fluoxetine reduced the increases in height and weight over 19 weeks by 1.0 cm and 1.1 kg, respectively ($P = 0.008$ for both). There were no data about these harms in the published article [8], and no comment on them in the CSR when they were presented [10].

The company concluded that “fluoxetine 20 to 60 mg/day is safe”. Discussing the 19-weeks data for all the patients, the CSR stated that the significance of the effect on height was uncertain, as previous studies had not found this. Lilly repeated this statement in their published paper on safety and added that the “Confidence in the interpretation of this result is limited, given that height in this study was not collected in a standardized fashion. Moreover, since measurements were recorded and rounded to the nearest inch, a small imprecision in measurement could potentially result in a 1-inch difference in recorded height [14]”. However, if recordings are rounded, this affects both groups. Furthermore, the concern is at odds with the fact that height for each patient was reported with two decimals (e.g. 160.02 cm) and weight with eight decimals (e.g. 47.62719885 kg) [14].

Fluoxetine increased the QTc interval by 6.95 msec ($P = 0.02$ for the difference to placebo). After numerous analyses and manipulations over six pages, e.g. showing that the P -value was not statistically significant for children with out-of-range ECG intervals and was not present in subgroup analyses by age, Lilly concluded that, “Two independent, blinded analyses of ECG interval changes did not reveal clinically significant changes in any ECG parameter”.

Finally, fluoxetine increased serum cholesterol (difference to placebo 0.2 mmol/L, $P = 0.01$).

3.2.9. Additional analyses of adverse events

In a “Safety Data Summary” of 1160 pages, for changes in height and weight, Lilly introduced a new interval, all randomised patients “who did not discontinue prior to Visit 11”. This visit was two weeks into the last 10-week period. There was no explanation why this interval was chosen. Data on ECGs were also presented using the new cut-off of 11 weeks. As before, the harmful effect of fluoxetine was removed by dichotomising the data and looking at out-of-range values.

3.3. Lilly's conclusions

The CSRs for both studies started with a 3-page summary, which included a table without data but with 9 and 13 *P*-values, respectively, for cherry-picked efficacy outcomes, which were all statistically significant [8,10]. There was no indication that LOCF had been used.

For trial HCJE, 7 of the 13 outcomes were not prespecified in the protocol: 20% and 40% reductions, recovery, and four subtotal scores on the CDRS-R, which we have only seen in studies of fluoxetine. The published paper did not mention that LOCF was used or that some of the outcomes were not prespecified [8]. Six of the paper's eight authors were employees of Lilly and "may own stock in that company". The two remaining authors were paid consultants for Lilly.

A claim that doses of 40 and 60 mg daily were more effective than 20 mg in those who had not responded to 20 mg was not correct.

A later 2.5-page "Discussion and Overall Conclusions" section praised fluoxetine's benefits and lack of harms, with no mention that the children did not find fluoxetine effective or that the number needed to harm was only 6 for nervous system events and 7 for moderate or severe adverse events. Contradicting the data, it claimed there were no "clinically significant changes in any ECG parameter," and that the evaluation of adverse events, vital signs (which included height and weight), and laboratory data demonstrated the safety of fluoxetine 20 to 60 mg daily [10].

3.4. Adverse events predisposing to violence against self or others

As we did not have access to individual patient data, we counted events even though some patients had more than one predisposing event (see Supplementary Table S1).

For solicited adverse events, 11 of 32 events were considered definite or possible precursors to suicide or violence, or both, by the blinded assessor (DH), and 9 were considered definite precursors. There were 131 versus 124 definite or possible precursors in trial X065; 396 versus 396 in trial HCJE after 9 weeks; and 450 versus 440 after 19 weeks. For definite precursors, the numbers were 100 versus 100, 314 versus 320 and 358 versus 353, respectively.

For non-solicited adverse events, the data were very different. There were 84 versus 72 definite or possible precursors in trial X065; 72 versus 42 in trial HCJE after 9 weeks; and 102 versus 49 after 19 weeks. There were 70 versus 61, 40 versus 18 and 58 versus 22 definite precursors, respectively.

For the Fluoxetine Side-Effects Checklist, which was used only in trial X065, there were 58 versus 23 definite or possible precursors and 39 versus 13 definite precursors.

Taking the two studies together, the occurrence of adverse events definitely predisposing to violence against self or others leading to discontinuation was 11 versus 3.

One of the strongest precursors for violence against self or others is akathisia. In an exploratory analysis, we included akathisia and other potentially related symptoms (see Supplementary Table S2). For non-solicited adverse events, there were 37 versus 32 such adverse events in trial X065; 38 versus 16 in trial HCJE after 9 weeks; and 51 versus 24 after all 19 weeks. We included nervousness because Lilly had coded restlessness as nervousness in a patient narrative. For the Fluoxetine Side-Effects Checklist, there were 30 versus 12 potentially extrapyramidal drug harms.

3.5. Other HCJE papers

HCJE was a 19-week trial but when we searched on PubMed and on clinicaltrials.gov (where the trial was not registered) using Emslie's name, we only found spin-off publications of the 9-week data, in which

HCJE is described as a 9-week trial [15,16], which is also the case in the FDA approved package insert [17].

Two of the spin-off papers stated wrongly that 309 patients had been randomised [18,19]. A third paper mentioned 315 patients and that one patient in X065 and five in HCJE were excluded because they did not have a post-randomisation visit [15]. However, only four patients in HCJE did not have a post-randomisation visit [10].

After additional searching, we found that some of the 19-week results had been published, but without Emslie as author [14,20]. As noted above, the 19-week efficacy results were less positive for fluoxetine than the 9-week results, but they seem never to have been published in full. Only results for 29 non-responders re-randomised after 9 weeks to continue with 20 mg fluoxetine or to increase the dose to a maximum of 60 mg were published [20]. This is 13% of those originally randomised. The authors only mentioned four of the numerous efficacy outcomes and concluded that a “dose escalation may benefit some patients” even though there were no significant differences: CDRS-R response ($P = 0.13$), CDRS-R score ($P = 0.099$), CGI-Severity ($P = 0.40$), and CGI-Improvement ($P = 0.30$).

One patient receiving fluoxetine 60 mg/day reported self-mutilatory behaviour of mild severity and discontinued from the study shortly thereafter “owing to lack of efficacy [20]”. This event was not described in the 2549-page study report. The authors concluded that adverse events were similar in the two groups but only counted them if they were worse than in the preceding 9-week period. In the Discussion, they were more cautious and noted that the increased dosage “was not associated with a marked increase in the number or severity of adverse events”. However, they did not explain what they meant by a “marked increase” or describe the severity of the adverse events.

The last phase of HCJE, the 32-week relapse prevention study, was published by Emslie et al. in 2004 [21]. This time, Emslie called HCJE a 51-week study. The mean time to relapse was longer in 20 patients continuing on fluoxetine than in 20 patients switched abruptly to placebo ($P = 0.046$) [21]. The company already knew from a prior study that abrupt withdrawal of antidepressants could cause abstinence depressions in many patients [22].

A 2007 Lilly meta-analysis of violent events included all placebo-controlled studies of fluoxetine undertaken in children and adolescents (376 patients on fluoxetine and 255 on placebo) [23]. Potential aggression or hostility-related events were identified by a computerized text string search of all investigator-recorded adverse events, all coded adverse events, and narratives. The text strings consisted of 71 words or abbreviations, and the events were reviewed and categorized blindly by company staff. Given this comprehensive effort, it is totally implausible that aggression or hostility-related events were experienced by *fewer* children and adolescents treated with fluoxetine, 2.1%, than treated with placebo, 3.1% ($P = 0.59$).

These results contradicted our findings and FDA’s assessment of Lilly’s application for treatment of children and adolescents with fluoxetine. FDA created a table of discontinuations because of adverse events in X065, HCJE and HCJW, a trial of obsessive-compulsive disorder comparing fluoxetine 10–60 mg daily with placebo for 13 weeks in 71 versus 32 patients [13]. There were 14 versus 3 discontinuations ($P = 0.02$, our calculation) among the 228 versus 190 patients for reasons related to suicide and violence (suicide attempt, euphoria, manic reaction, agitation, hyperkinesia, nervousness, personality disorder, hostility, and depression). In these trials, there were 3 suicide attempts on fluoxetine and 1 on placebo, and another fluoxetine patient was hospitalized because of suicidality. Six patients (2.6%) on fluoxetine developed mania or hypomania versus none on placebo ($P = 0.03$) [13]. The FDA reviewer remarked that mania and hypomania appeared to be more common on fluoxetine in these trials than in adult clinical studies. A table of spontaneously reported adverse events in HCJW and HCJE (9 weeks data)

showed that more patients developed hyperkinesia on fluoxetine than on placebo, 12 versus 1 patients ($P = 0.008$, our calculation) [13].

4. Discussion

Both Eli Lilly and Emslie et al. concluded for both trials that fluoxetine is safe and effective for children and adolescents who are depressed [7–10]. As already noted, in the absence of individual patient level data as recorded by the investigators, it is not possible to estimate the full scale of fluoxetine's effects. However, even in the absence of the full data, we found that fluoxetine is unsafe and ineffective.

4.1. Comments on the results

Patient ratings did not find fluoxetine effective, and the effects Lilly reported were not clinically relevant. The effect on the CDSR-R relative to the baseline values (58.2 and 56.2 in X065 and HJCE, respectively), was 4% in both trials (16% versus 9% if LOCF is used). By comparison, the least recognizable effect on the equivalent adult scale, the Hamilton depression scale was found to be 5–6 [24], corresponding to 28% of a median baseline of 25.4 in 35 placebo-controlled trials [25]. Furthermore, there were only minor differences in numbers of patients who had minimal symptoms, or had recovered, or had a good outcome on the CGI-Efficacy Index, which compares the benefits and harms for each patient.

The company violated its trial protocols; essential information was scattered around in the reports; a lot was missing despite being indexed; there were many errors, inexplicable numerical inconsistencies, and unexplained exclusions of patients from analyses; and results that were inconsistent with the conclusion that fluoxetine is safe and effective were side-lined or explained away in a disturbing manner.

Statistical testing was done to an extreme. We found 5,910 significance tests with P -values in the CSRs. For efficacy outcomes, 39% were significant in favour of fluoxetine, concealing a failure to find significance on primary endpoints [1]. If fluoxetine had been as harmless as placebo, 229 (5%) of 4,575 tests for adverse events would have been statistically significant by chance. With an active drug, *more* results than 229 should have been significant but there were only 174 (4%). Many tests were run on events that occurred in only one or two patients.

The lack of efficacy in primary endpoints and significant safety hazards were noted by the FDA reviewer of the fluoxetine license application in 2002 [1]. FDA criticised Lilly for not having searched their database for signals of any unusual adverse events in children and adolescents. In response, Lilly provided a literature review for FDA, which was only about efficacy [13].

Lilly did not accept that a restoration of the two studies was needed, arguing in their letter to us that they had “adequately analysed the data” and that we had not reported a “significantly increased risk of suicidality or aggressive behaviour” in these trials in our 2016 meta-analysis of CSRs of antidepressants [26]. However, the increase in the risk of suicide and violence is a class effect that is difficult to detect in two small trials, and we documented it. The odds ratios for children and adolescents were 2.39 (95% confidence interval 1.31 to 4.33) for suicidality and 2.79 (1.62 to 4.81) for aggression. The summary trial reports on Lilly's website for fluoxetine and duloxetine are seriously misleading, as 90% of the suicide attempts, all suicidal ideation events, and most cases of aggression and akathisia are missing [26].

Lilly claimed that “depression is an organic disease that readily responds to treatment” and that “Introduction of effective antidepressant treatments earlier in the progression of the disease state has the potential to effectively treat and control the disease as well as improve daily functioning and overall

quality of life” [10]. There is no evidence that either of this is true [5,27]. For quality of life, there is an extreme degree of selective reporting not only in the published literature [28] but even *within* the CSRs of the placebo-controlled trials [29]. These drugs likely decrease quality of life, e.g. 12% more patients drop out on drugs than on placebo [30], and they disrupt the sex lives in about half of those treated [31], which may continue long after the treatment stopped [32].

This is a significant issue for young people going through puberty. The package insert for fluoxetine mentions that significant toxicity on muscle tissue, neurobehaviour, reproductive organs, and bone development has been observed in juvenile rats, with testicular degeneration and necrosis, epididymal vacuolation and hypospermia [17]. The findings indicate that the drug effects on reproductive organs are irreversible, and when animals were evaluated after a drug-free period (up to 11 weeks after cessation of dosing), fluoxetine was associated with neurobehavioural abnormalities with decreased reactivity at doses corresponding to only 10–20% of the maximum recommended human dose.

The package insert also notes that “there are no studies that directly evaluate the longer-term effects of fluoxetine on the growth, development and maturation of children and adolescent patients” [17]. If extrapolated from the trial data, the harm corresponds to an annual loss in height and weight increase of 2.7 cm and 3.0 kg, respectively. We do not know if fluoxetine also has deleterious effects on the developing brain. FDA requested that Lilly conducted a one-year study of the effect of fluoxetine on growth, which the company declined to do [13].

There was a significant increase in the QT interval. Lilly stated that only one patient had a QT interval that increased by more than 60 msec and that “this patient experienced no adverse events that were of clinical concern” [14]. FDA requested Lilly to search their adverse events database, which, as of 2001, in the 6-17-year-old group, produced 7 reports of prolonged QT interval, 3 reports of cardiac arrest, 1 sudden unexplained death, and 1 ventricular fibrillation [13].

Informed consent was also problematic. Many harms were listed in the parent consent form for trial HJCE but there was nothing about the increased risk of suicide and violence. Furthermore, even though some patients were switched abruptly to placebo, there was no warning about withdrawal symptoms, nor that withdrawal can lead to suicide and violence. About risks, it was explained that the skin and eyes could turn yellow, but not that this is a sign of liver damage.

For trial X065, 34 words were used to describe that a needle prick might result in a small bruise that should cause little or no discomfort while only 20 words were used to mention concrete adverse effects of fluoxetine, with no hint that that they could be serious.

Antidepressants double suicide attempts in both children and adults [4,5,27,33] and therefore likely also suicides. The 2016 package insert for fluoxetine underlines how dangerous these drugs are [17]. A meta-analysis of 24 placebo-controlled trials in over 4400 children and adolescents showed that for every 1000 patients treated with drug instead of placebo, there were 14 additional cases of suicidality (with the highest incidence in depression). Number needed to harm is therefore only 71.

Many leading professors of psychiatry and spokespersons for general practitioners supporting antidepressant use claim that they *protect* children and adolescents against suicide [5,34]. Websites are also misleading. A 2018 analysis showed that 25 (64%) of 39 popular websites from 10 countries stated that antidepressants may cause suicidal ideation, but 23 (92%) of them contained incorrect and sometimes outright dangerous information [35].

There are four main reasons why the misinformation continues. First, many suicidal events have been omitted or concealed in published trial reports [4–6,26,36]. Second, some investigators did not look for them [36]. Despite these obstacles, a 2005 meta-analysis of published trials that included all ages reported double as many suicide attempts on drug than on placebo (odds ratio 2.28, 1.14 to 4.55) [36]. Third, events

occurring shortly after active treatment was stopped were not included until recently [33]. Fourth, there is a commercial interest in sustaining the misinformation.

Fortunately, something can be done. The usage of antidepressants in children and adolescents increased by 59% in Denmark from 2006 to 2010, but in the following six years, when one of us constantly made clinicians and the general public in Denmark aware of the suicide risk of antidepressants, in interviews and articles, the usage dropped by 41% while it increased by 40% in Norway and 82% in Sweden in the same period [34].

When X065 and HCJE were run, there was an expectation that many unhappy children could be helped with psychotherapy and other types of support. This is still the case, and a meta-analysis showed that cognitive behavioural therapy halved the risk of future suicide attempts in young people admitted after a suicide attempt [37].

4.2. *Comments on the context*

Given FDA's recognition of a lack of efficacy on primary endpoints [1] and significant safety hazards of fluoxetine, the results of our paper raise as many questions about the company and the regulatory situation around 2002, as they do about fluoxetine use for minors. The context in which these trials took place may shed a light on these issues.

Prior to the development of the SSRIs, there had been 15 randomised trials of tricyclic and related antidepressants in children and adolescents, all negative [38]. An initial trial of fluoxetine was also negative [39]. None of these were high quality trials. There was hope that a well-done trial might demonstrate a benefit.

There was also, however, a clinical consensus and literature that children did not get endogenous depression. They might be miserable and unhappy, but this was situational and would respond to supportive interventions. Linked to this, there were almost no child psychiatrists with expertise in psychopharmacology.

A further feature of SSRIs at that time is that they were not effective in any age group for endogenous depression (melancholia). They had an anxiolytic, or serenic, action. The SSRIs became antidepressants in part to skirt around clinical concerns that any new anxiolytic would necessarily produce dependence as the benzodiazepines had [40]. The randomised trials done in children support the point that the drugs are essentially anxiolytic rather than antidepressant. A meta-analysis of published trials showed significantly larger effect sizes for anxiety ($g = 0.56$) and obsessive-compulsive disorder ($g = 0.39$) than for depression ($g = 0.20$) [41].

In 2008, Erick Turner and colleagues noted that 31% of adult trials done as part of a licensing application for SSRIs and related antidepressants viewed by FDA as negative or questionable were published as positive, and the effect size in the published articles was 32% higher than in the FDA reviews [42]. FDA made no comment about these findings.

In 1997, the year X065 completed, Congress offered half a year of patent extension to companies who submitted trials done in children to FDA. The studies did not need to be positive, or to be published; the stated intention was to help establish the safety profile in children [43].

Two positive trials are needed for a license and Lilly immediately began study HCJE. Both studies were submitted to FDA in support of a patent extension. FDA supported this and also licensed claims that fluoxetine could be used to treat depressed children, as did other regulators that year.

After licensing the fluoxetine claim, based on studies negative on their primary endpoint, FDA issued an approvable letter in October 2002 for paroxetine in the treatment of children and adolescents who were

depressed. The letter agreed with GlaxoSmithKline (GSK) that all three trials they submitted (protocols 329, 377 and 701) in the application were negative as regards efficacy (letter available on study329.org). FDA also noted: “Given the fact that negative trials are frequently seen, even for antidepressant drugs that we know are effective, we agree that it would not be useful to describe these negative trials in labelling”.

In the initial 2001 publication of study 329, a trial of paroxetine in depressed minors, GSK claimed paroxetine was safe and effective [44]. The study was not dissimilar to the fluoxetine studies in terms of safety and efficacy. An internal document from 1998, however, made it clear that GSK knew the study demonstrated its drug to be ineffective but that it would be commercially unacceptable to publish this [45,46]. The document states that the “good bits of the study would be published” (available on study329.org).

Based on this information, New York State’s Attorney General lodged a fraud action against GSK. The settlement of this action made it possible to access data on study 329 and restore it in a manner that demonstrated paroxetine’s lack of efficacy and a doubling of suicidal events acts compared with the original publication [6].

In response to concerns about the suicide risk of antidepressants in this age group, FDA convened a Psychopharmacologic Drugs Advisory Committee meeting in February 2004. FDA claimed none of the drugs (bar fluoxetine) demonstrated efficacy. FDA deferred a strong safety signal, a doubling of the risk of suicidal events ($P = 0.00005$) [43], for action at a hearing in September [43]. The New York case against GSK lodged in March 2004 triggered a crisis, and at its September meeting, FDA accepted the need for a Black Box Warning, in part because of the accepted lack of efficacy for antidepressants in this age group. The licensing of paroxetine and other drugs was aborted but the approval of fluoxetine was not rolled back. Regulators, then and since, apart from the Australian one, which did not approve fluoxetine in minors, continued to claim that fluoxetine is effective in this age group.

There has been one independent trial of fluoxetine, the National Institutes of Health’s Treatment of Adolescent Depression Study (TADS). It claimed efficacy and safety for fluoxetine, but the effect was not clinically relevant, and there were double as many suicidal events in patients randomised to fluoxetine than in patients randomised to placebo [47,48]. In none of 15 articles about this study between 2004 and 2010, published after the Black Box Warnings on these drugs, have these fluoxetine suicidality data been addressed. Duke University, where the trial data were lodged, has refused to hand over serious adverse event forms from the trial that might permit a restoration of this study [49].

The dataset from X065 and HCJE on QT intervals also has a regulatory context. When these trials were being reviewed by FDA, Lilly submitted a license application for R-fluoxetine, which was ultimately withdrawn in part because of QTc interval problems [4]. Such problems are clearly an issue with fluoxetine, as they are with all SSRIs. In response to FDA concerns about study HCJE, Lilly argued that the statistically significant increase in mean QTc found with the initial analysis was the product of random variability [13]. FDA’s reviewer responded that, with a P -value of 0.009, the result was, by definition, unlikely to be produced by random variability.

A further issue to consider stems from the Turner et al. article [42]. If FDA had stated that study 329 was negative, they might have opened GSK up to a fraud action and a large settlement fine, as later resulted. GSK and all companies are likely to have been aware of this when in discussions with FDA. It is not the job of drug regulators to police the medical literature but, if the medical literature limits a regulator’s freedom of movement, we have an uncomfortable nexus of issues.

The approval of fluoxetine for depression in children and adolescents and the publication of many articles since, often ghost written, claiming efficacy for a number of SSRIs swept away the idea of relying on psychotherapy and other forms of support. The usage of antidepressants in adolescents, particularly

females, is increasing markedly in most countries and these drugs appear to be among the most commonly used drugs by females in their later teenage years. The figures for usage increase in part because half of the patients find it difficult to withdraw from them [27]. This is a particular issue in women of child-bearing age given evidence that these drugs increase rates of birth defects, behavioural abnormalities in children born to mothers on them as well as miscarriages and voluntary terminations [50,51].

FDA's willingness to license claims that fluoxetine was an antidepressant in paediatric populations was a key step in the evolution of this public health problem, along with the great divide between what the academic literature on these drugs said before 2004 and what the data say when accessed. We need to understand how this situation arose if we are to prevent comparable predicaments in future.

Ethical approval

Not required.

Conflict of interest

Both authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that they have worked as expert witnesses for law firms in cases about suicide and homicide caused by antidepressants.

Funding

RIAT Support Center, Baltimore, Maryland. Award no. N210955-1.

Author contributions

PCG obtained funding, wrote the protocol, extracted data and wrote the first manuscript. DH judged blindly if adverse events were precursors of suicide or violence and contributed to the manuscript.

Data sharing

The clinical study reports are available on <https://www.scientificfreedom.dk/links/>.

Supplementary materials

The supplementary tables are available from <https://dx.doi.org/10.3233/JRS-210034>.

References

- [1] Mosholder AD. Application number: 18-936/SE5-064. Statistical review. FDA 2001; 20 July.
- [2] Bielefeldt AØ, Danborg PB, Gøtzsche PC. Precursors to suicidality and violence on antidepressants: Systematic review of trials in adult healthy volunteers. *J R Soc Med*. 2016;109:381–92.
- [3] Hughes S, Cohen D, Jaggi R. Differences in reporting serious adverse events in industry sponsored clinical trial registries and journal articles on antidepressant and antipsychotic drugs: A cross-sectional study. *BMJ Open*. 2014;4:e005535.
- [4] Healy D. *Let them Eat Prozac*. New York: New York University Press; 2004.
- [5] Gøtzsche PC. *Deadly Psychiatry and Organised Denial*. Copenhagen: People's Press; 2015.

- [6] Le Noury J, Nardo JM, Healy D, Jureidini J, Raven M, Tufanaru C et al. Restoring Study 329: Efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ*. 2015;351:h4320.
- [7] Emslie GJ, Rush AJ, Weinberg WA, Kowatch RA, Hughes CW, Carmody T et al. A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Arch Gen Psychiatry*. 1997;54:1031–7.
- [8] Emslie GJ, Heiligenstein JH, Wagner KD, Hoog SL, Ernest DE, Brown E et al. Fluoxetine for acute treatment of depression in children and adolescents: A placebo-controlled, randomized clinical trial. *J Am Acad Child Adolesc Psychiatry*. 2002;41:1205–15.
- [9] Eli Lilly and Company. Protocol B1Y-MC-X065. Clinical study main report: Fluoxetine versus placebo in the acute treatment of major depressive disorder in children and adolescents. 2000.
- [10] Eli Lilly and Company. Protocol B1Y-MC-HCJE. Clinical study report: Fluoxetine versus placebo in childhood/adolescent depression. 2000.
- [11] Gøtzsche PC. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ*. 2019.
- [12] Cheung A, Emslie GJ, Maynes TL. Efficacy and safety of antidepressants in youth depression. *Can Child Adolesc Psychiatr Rev*. 2004;13:98–104.
- [13] Mosholder AD. Application number: 18-936/SE5-064. Medical review. FDA 2001.
- [14] Nilsson M, Joliat MJ, Miner CM, Brown EB, Heiligenstein JH. Safety of subchronic treatment with fluoxetine for major depressive disorder in children and adolescents. *J Child Adolesc Psychopharmacol*. 2004;14:412–7.
- [15] Mayes TL, Tao R, Rintelmann JW, Carmody T, Hughes CW, Kennard BD et al. Do children and adolescents have differential response rates in placebo-controlled trials of fluoxetine? *CNS Spectr*. 2007;12:147–54.
- [16] Jain S, Carmody TJ, Trivedi MH, Hughes C, Bernstein IH, Morris DW et al. A psychometric evaluation of the CDRS and MADRS in assessing depressive symptoms in children. *J Am Acad Child Adolesc Psychiatry*. 2007;46:1204–12.
- [17] FDA package insert for Prozac. October 2016. <https://pi.lilly.com/us/prozac.pdf> (accessed 14 March 2020).
- [18] Cheung A, Mayes T, Levitt A, Schaffer A, Michalak E, Kiss A et al. Anxiety as a predictor of treatment outcome in children and adolescents with depression. *J Child Adolesc Psychopharmacol*. 2010;20:211–216.
- [19] Emslie GJ, Kennard BD, Mayes TL, Nakonezny PA, Zhu L, Tao R et al. Insomnia moderates outcome of serotonin-selective reuptake inhibitor treatment in depressed youth. *J Child Adolesc Psychopharmacol*. 2012;22:21–8.
- [20] Heiligenstein JH, Hoog SL, Wagner KD, Findling RL, Galil N, Kaplan S et al. Fluoxetine 40–60 mg versus fluoxetine 20 mg in the treatment of children and adolescents with a less-than-complete response to nine-week treatment with fluoxetine 10–20 mg: a pilot study. *J Child Adolesc Psychopharmacol*. 2006;16:207–17.
- [21] Emslie GJ, Heiligenstein JH, Hoog SL, Wagner KD, Findling RL, McCracken JT et al. Fluoxetine treatment for prevention of relapse of depression in children and adolescents: a double-blind, placebo-controlled study. *J Am Acad Child Adolesc Psychiatry*. 2004;43:1397–405.
- [22] Rosenbaum JF, Fava M, Hoog SL et al. Selective serotonin reuptake inhibitor discontinuation syndrome: a randomised clinical trial. *Biol Psychiatry*. 1998;44:77–87.
- [23] Tauscher-Wisniewski S, Nilsson M, Caldwell C, Plewes J, Allen AJ. Meta-analysis of aggression and/or hostility-related events in children and adolescents treated with fluoxetine compared with placebo. *J Child Adolesc Psychopharmacol*. 2007;17:713–18.
- [24] Leucht S, Fennema H, Engel R et al. What does the HAMD mean? *J Affect Disord*. 2013;148:243–48.
- [25] 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.
- [26] Sharma T, Guski LS, Freund N, Gøtzsche PC. Suicidality and aggression during antidepressant treatment: Systematic review and meta-analyses based on clinical study reports. *BMJ*. 2016;352:i65.
- [27] Gøtzsche PC. *Mental Health Survival Kit and Withdrawal from Psychiatric Drugs*. Ann Arbor: L H Press; 2022.
- [28] Healy D. The assessment of outcomes in depression: Measures of social functioning. *Rev Contemp Pharmacother*. 2000;11:295–301.
- [29] Paludan-Müller AS, Sharma T, Rasmussen K, Gøtzsche PC. Extensive selective reporting of quality of life in clinical study reports and publications of placebo-controlled trials of antidepressants. *Int J Risk Saf Med*. 2021;32:87–99.
- [30] Sharma T, Guski LS, Freund N, Meng DM, Gøtzsche PC. Drop-out rates in placebo-controlled trials of antidepressant drugs: A systematic review and meta-analysis based on clinical study reports. *Int J Risk Saf Med*. 2019;30:217–232.
- [31] Montejo A, Llorca G, Izquierdo J et al. Incidence of sexual dysfunction associated with antidepressant agents: A prospective multicenter study of 1022 outpatients. Spanish Working Group for the study of psychotropic-related sexual dysfunction. *J Clin Psychiatry*. 2001;62(suppl 3):10–21.
- [32] Healy D. Antidepressants and sexual dysfunction: A history. *J R Soc Med*. 2020;113(35):133–135.
- [33] Hengartner MP, Plöderl M. Reply to the Letter to the Editor: “Newer-Generation Antidepressants and Suicide Risk: Thoughts on Hengartner and Plöderl’s Re-Analysis”. *Psychother Psychosom*. 2019;88(4):373–74.

- [34] Gøtzsche PC. Usage of depression pills almost halved among children in Denmark. *Mad in America* 2018; 4 May. <https://www.madinamerica.com/2018/05/usage-depression-pills-almost-halved-among-children-denmark/>.
- [35] Demasi M, Gøtzsche PC. Presentation of benefits and harms of antidepressants on websites: cross sectional study. *Int J Risk Saf Med*. 2020;31:53–65.
- [36] Fergusson D, Doucette S, Glass KC, Shapiro S, Healy D, Hebert P et al. Association between suicide attempts and selective serotonin reuptake inhibitors: Systematic review of randomised controlled trials. *BMJ*. 2005;330:396.
- [37] Gøtzsche PC, Gøtzsche PK. Cognitive behavioural therapy halves the risk of repeated suicide attempts: systematic review. *J R Soc Med*. 2017;110:404–10.
- [38] Jensen PS, Ryan ND, Prien R. Psychopharmacology of child and adolescent major depression: Present status and future directions. *J Child Adolesc Psychopharmacol*. 1992;2:31–45.
- [39] Simeon JG, Dinicola VF, Ferguson HB, Copping W. Adolescent depression: A placebo-controlled fluoxetine treatment study and follow-up. *Prog Neuropsychopharmacol Biol Psychiatry*. 1990;14:791–5.
- [40] Healy D. *The Antidepressant Era*. Cambridge: Harvard University Press; 1997.
- [41] Locher C, Koechlin H, Zion SR, Werner C, Pine DS, Kirsch I et al. Efficacy and safety of selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, and placebo for common psychiatric disorders among children and adolescents: A systematic review and meta-analysis. *JAMA Psychiatry*. 2017;74:1011–20.
- [42] Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358:252–60.
- [43] Newman TB. A black-box warning for antidepressants in children? *N Engl J Med*. 2004;351(8):1595–98.
- [44] Keller MB, Ryan ND, Strober M, Klein RG, Kutcher SP, Birmaher B 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.
- [45] Jureidini JN, McHenry LB. Conflicted medical journals and the failure of trust. *Accountability in Research*. 2001;18:45–54.
- [46] Healy D, LeNoury J, Wood J. *Children of the Cure*. Toronto: Samizdat Writers Co-operative; 2020.
- [47] March J, Silva S, Petrycki S et al. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA*. 2004;292:807–20.
- [48] Högberg G, Antonuccio DO, Healy D. Suicidal risk from TADS study was higher than it first appeared. *Int J Risk Saf Med*. 2015;27:85–91.
- [49] Aboustate N, Jureidini J. Barriers to access to clinical trial data: Obstruction of a RIAT reanalysis of the Treatment for Adolescents with Depression Study. *Int J Risk Saf Med*. 2021 (in press).
- [50] Healy D, Mangin D, Mintzes B. The ethics of randomized placebo controlled trials of antidepressants with pregnant women. *Int J Risk Saf Med*. 2010;22:7–16.
- [51] Healy D, Le Noury J, Mangin D. Links between serotonin reuptake inhibition during pregnancy and neurodevelopmental delay/spectrum disorders: A systematic review of epidemiological and physiological evidence. *Int J Risk Saf Med*. 2016;28:125–41.