

Science, rhetoric and the causality of adverse events

David Healy*

Department of Psychological Medicine, Cardiff University, Hergest Unit, Wales, UK

Abstract. This paper outlines aspects of the interface between law and science in the domain of treatment induced injury, using examples drawn from litigation on SSRIs. In the face of claims that randomized controlled trials provide uniquely appropriate evidence and that there should be a statistically significant doubling of the risks on treatment, the examples used demonstrate that the methods of assessing the possibility of causal links between treatment and injuries must necessarily be tailored to both the injury and the treatment.

keywords: Paroxetine, suicide, dependence, birth defects, controlled trials, case studies

1. Introduction

A recent paper in this journal laid out a template for exploring the confusions surrounding the use of medical evidence in legal cases and argued for the establishment of forensic pharmacovigilance as a discrete field of expertise [1].

This paper uses recent cases involving the Selective Serotonin Reuptake Inhibitors (SSRIs) and their propensities to trigger suicides, birth defects and dependence as a case study to explore further the appropriate use of evidence in medico-legal settings. The complexity of “causality” assessments outlined supports the need for a specialized approach to forensic pharmacovigilance.

2. The medico-legal background

The SSRI group of drugs were launched in the mid to late 1980s, with Prozac reaching celebrity status in 1989/1990. This was a time when debates about the nature of the appropriate scientific evidence in mass tort cases had begun to take shape. A series of publications such as *Science on Trial* [2] and *Galileo's Revenge* [3] had helped to frame that debate.

As a consequence of cases involving Bendectin and birth defects (and later breast implants and connective tissue disorders), where claims for causality based on individual case studies were seen to be inappropriate in the wake of findings from dedicated epidemiological studies, arguments were put forward in the *Daubert v Merrell Dow* case that the legal system in general should require mass tort cases to be based on solid epidemiological foundations [4]. As is argued further below, this kind of evidence has a merit in cases involving some birth defects or device induced diseases that it does not necessarily

*Address for correspondence: David Healy, Department of Psychological Medicine, Cardiff University, Hergest Unit, Bangor LL57 2PW, Wales, UK. E-mail: David.Healy54@googlemail.com.

29 have for other treatment induced injuries. When not in fact needed, an insistence that such studies are
30 necessary plays into drug company hands as these are costly and difficult studies to mount.

31 A further notion emerged from these cases, namely that epidemiological studies should demonstrate
32 a relative risk of injury from an offending agent of 2.0 or greater. Unless the evidence of injury reached
33 such a threshold, litigation it was argued would be inappropriate.

34 A relative risk of 2.0 or greater in an epidemiological study appears to correspond with legal thinking
35 that appropriate regards actions as appropriate when it is more probable than not that an agent has caused
36 a particular defect. This legal standard is sometimes referred to in terms of reasonable medical certainty
37 or a balance of probabilities, and traditionally this has been taken to mean greater than 50% probability.
38 There is an apparent intersection here between legal and scientific thinking in that in epidemiology what
39 is termed the etiologic fraction is the relative risk (RR) minus 1 divided by the relative risk. If the relative
40 risk is 2, the etiologic fraction therefore equals 50 per cent.

41 Despite appearances, there is no philosophical, scientific, or legal justification to assume these notions
42 from different domains have any common basis. Legal thinking in this domain arose from a consideration
43 of the likely role a treatment might have played in an injury sustained in an individual case (see below)
44 and notions of relative risk, which depend on repetitive sampling, cannot apply to such judgements. The
45 confusion arises because both sets of judgements for different reasons are cast in terms of a balance of
46 probabilities, and such an index is bureaucratically convenient for making decisions about cases. But it
47 offers greater comfort to pharmaceutical companies than anyone else.

48 Within European settings, for reasons developed below, one of the greatest absurdities in applying such
49 a rule without thought arose in cases involved thrombo-embolism on third generation oral contraceptive
50 drugs, where the judge although finding that the risk had been increased by 70% refused to allow the
51 action to proceed [5]. This was quite different to the approach of American courts in letting actions against
52 Seroxat for birth defects proceed, even though the relative risk was not of this order. Several hundred
53 such actions have now been resolved with media reports claiming a possible outlay by GlaxoSmithKline
54 of over \$1 billion.

55 **3. SSRIs & suicide: From case studies to controlled trials**

56 The first legal actions involving SSRIs and suicide appealed to traditional formulations regarding
57 possible causal links between treatments and effects (in this case injuries) - as outlined by Robert Koch in
58 the 1880s and elaborated by Bradford Hill and others in the 1960s [6–9]. These are taken to hold that for
59 a causal relationship to be present the problem should appear after exposure to the drug; the effect of the
60 drug should ideally clear up after treatment is stopped (de-challenge), and may re-appear on re-exposure
61 to the drug (re-challenge); there should be some evidence of a dose response relationship between drug
62 and effect; the effect would ideally be reversed by antidote; and the proposed effect should be biologically
63 plausible.

64 There are several formulations of these points. One of them, the Naranjo algorithm, classifies events
65 in terms of virtual certainty, probable and possible based on the number of items endorsed and a numeric
66 scoring of these items. This should not be understood in terms of a statement of frequencies or the precision
67 of our knowledge but more in terms of the quality of fit between a hypothesis and the available data.
68 Many injuries do not fit this template and as outlined below withdrawal syndromes invert the algorithm.

69 In 1990s SSRI cases, courts initially accepted arguments based upon “traditional assessments of causal-
70 ity”. In the case of fluoxetine (Prozac), several compelling case series drawn from a number of independent

71 and distinguished academic centres showing a typical duration between first intake of the drug and the
72 emergence of the problem, relief of the problem on discontinuation, re-emergence of the problem on
73 reinstatement of treatment, dose-responsiveness, reversal by antidote and some knowledge of possible
74 mechanisms through which such effects might be mediated [10].

75 Companies responded by claiming the disease caused the problem. In addition, in the Miller case, which
76 involved the suicide of a 13 year old boy after a week's exposure to sertraline (Zoloft), the manufacturer
77 argued that unless plaintiffs had clinical trial evidence that sertraline could trigger suicidality the plaintiffs
78 did not have a case owing to the unreliability of case reports and of individual clinical judgements.

79 In the Miller case, the plaintiff's expert's argument had been the standard argument based on compelling
80 case reports that had succeeded in Daubert hearings in Prozac cases. In a Daubert hearing in the Miller
81 case, two independent court appointed experts, John Davis and John Concato, agreed with the basic
82 position that there needed to be some clinical trial or epidemiological evidence indicating that there was
83 a problem. This position rested on the preference of the two experts rather than argument.

84 When challenged on the need for a relative risk of 2.0 or greater before an effect could be said to be
85 medically or legally significant, both Davis and Concato denied that there was any such standard.

86 In the course of the Miller Daubert hearing, it became clear that a new premium had been put on clinical
87 trial evidence. Plaintiffs offered data from Pfizer clinical trials indicating that the relative risk of suicidal
88 acts on sertraline was not only greater than on placebo but greater with a relative risk in excess of 2.0.
89 The judge denied the admissibility of this evidence arguing that the only material that the independent
90 experts were to review was the original expert report. Based on the views of these experts about what
91 reports should look like the plaintiff's case was dismissed. However in dismissing the case, the judge
92 also noted that the argument being made by Pfizer was unbelievable and self serving [11].

93 Following this development, in further legal cases against GlaxoSmithKline and Pfizer for suicidal
94 and homicidal behaviour on paroxetine and sertraline there was a new premium on clinical trial data.
95 This insistence on clinical trial data led in the course of these actions to legal discoveries outlined below
96 regarding the information pertinent to any attempt to make "causality" assessments in the medico-legal
97 domain using epidemiological or clinical trial data. (Comparable discoveries have since been made in
98 Hormone Replacement Therapy (HRT), Vioxx and Avandia studies).

99 First, it became clear that a large number of the clinical trials that had been undertaken remained unpub-
100 lished. It also became clear that of the trials that were published almost all had been ghost written. Some
101 of these ghostwritten articles portrayed a positive result for the drug when in fact the findings had been
102 negative and in particular the adverse events of drugs were often omitted or misleadingly coded [12, 13].

103 Against this background, it is difficult to sustain an argument that the data from clinical trials offer more
104 useful or reliable information on the adverse effects of treatment than good case studies written up by
105 competent observers and reproduced across a number of sites. Dedicated studies designed to investigate
106 a particular adverse effect from which all the data are made available may provide a useful quantification
107 in some instances over good case studies, but even this cannot be assumed to apply in all cases.

108 Second, it became clear that in addition to non-publication of studies, data on significant adverse events
109 such as suicidal acts or completed suicides in the case of the SSRI drugs and heart attacks in the case
110 of Vioxx were omitted from published effectiveness studies or were treated in a misleading fashion so
111 that disinterested observers would come to false conclusions about the numbers of adverse events. In the
112 case of paroxetine, GlaxoSmithKline coded suicidal events under a heading of emotional lability [14],
113 and breached regulations in listing certain events under the heading of placebo [15].

114 Third, the new medico-legal focus drew attention to company publications indicating a higher fre-
115 quency of serious adverse events such as death on active drugs compared with placebo. The data in these

116 publications were typically not statistically significant and companies argued that results that are not
117 statistically significant in effect do not point to an increase in risk [16]. The fact that this insupportable
118 argument has traction in medico-legal cases points to failures in both legal and medical assessments of
119 evidence.

120 Fourth, the question of suicide on SSRIs raises issues about the use of the etiologic fraction in medico-
121 legal cases as a scientific criterion for proceeding with a case. In the case of an SSRI healthy volunteer
122 study, it is likely that the rate of suicidal acts or completed suicides on an active antidepressant agent would
123 yield a relative risk for active treatment compared to placebo in excess of 10.0. In contrast, were studies
124 of potent antidepressant agents, such as the tricyclics or electroconvulsive therapy (ECT) undertaken in
125 patients with severe depressive disorders at high risk of suicide it is possible that an active treatment
126 might lead to lower relative risk than placebo - below 1.0.

127 In fact the trials that brought the SSRIs to market were undertaken in patients with mild to moderate
128 disorders and against this background the relative risk of suicidal acts on active treatment hovers around
129 2.0. This illustrates the fact that epidemiological studies including controlled trials provide assay systems
130 or datasets and parameters such as relative risk and confidence intervals that describe those datasets rather
131 than offer definitive statements about objective reality. In the case of suicidal acts, these SSRI studies
132 point to the ratio between the risks and benefits of these drugs in particular populations rather than the
133 potential of the drugs to trigger suicidality.

134 A judgement needs to be made as to whether findings from assay systems like these can be generalized.
135 Assay systems are designed to produce a relative risk assessment and to quantify the likely role of chance
136 (random error) in producing the data found in the assay system alone, and not necessarily the role of
137 other chance factors that might arise in clinical settings. In particular such assays do not tackle the role
138 of systematic error stemming from our lack of knowledge. It may take several years and much debate to
139 make judgements about the role of systematic error in producing the observed data.

140 The SSRI trials for instance were never designed to look at the issue of suicide or suicidal acts on
141 treatment. It was and is possible to design such a trial. In the case of Prozac, in co-operation with the US
142 regulator, Eli Lilly, the makers of Prozac, designed a blinded rechallenge-dechallenge study of treatment
143 linked suicidality [17].¹ The study was however never undertaken. The relative risk assessment and
144 confidence intervals that would have emerged from such a study would have been far more pertinent to
145 legal actions than any estimates derived from trials conducted for other purposes.

146 It is now clear that the controlled trial data in the case of SSRIs and suicide are consistent with the case
147 studies that demonstrate suicide induction on treatment but it is easy to conceive of a situation in which
148 controlled trial data and the case studies might have stood in apparent opposition. In such circumstances,
149 appeals to a hierarchy of evidence that considers controlled trial data as more reliable than case studies
150 needs to be modified. (A possible modification is proposed in Appendix 1).

151 As must be clear from the argument above it would simply not be right to say a relative risk of less
152 than 1.0 means that a drug cannot have caused a problem. Equally even if the relative risk were 10.0 for
153 a problem like suicide, it would still be necessary to establish whether the drug had in fact caused the
154 problem in an individual case under consideration. The reasoning that would be applied in this case is
155 just the same as the reasoning that needs to be applied in a case where the relative risk is less than 1.0 –
156 namely whether in the individual case there are features consistent with a reduction in the problem on
157 dechallenge, relief by dose reduction, amelioration by antidote and other such considerations.

¹ Protocol available from the author.

158 Taking a Bayesian approach to causality, Caster and Edwards [18] have suggested that good case
159 studies that meet the traditional canons of causality should from the point of view of prior probabilities
160 be assigned a notional starting value of 50% as a hypothesis, to be modified by additional evidence as it
161 comes along.

162 This seems reasonable in that it is difficult to point to any evidence that is stronger than a case series
163 demonstrating challenge-dechallenge and rechallenge effects. However, the SSRI and suicide case series
164 contained cases that were close to conclusive using traditional assessment criteria, but not everyone was
165 persuaded, even though all clinical trial data from the very start pointed in exactly the same direction.
166 Few it seems were prepared to assign a value of 50% to the prior probability that treatment could trigger
167 this problem. One complicating factor in this case was that the disorder being treated was also a source of
168 the problem attributed to the drug. But in contrast, as outlined below, despite confounding by indication,
169 physicians clearly found reports of dependence on paroxetine highly probable.

170 It may be that many of us find certain effects of treatment, perhaps behavioural effects in particular,
171 inherently improbable. One merit of healthy volunteer trials or case studies is that in removing the element
172 of confounding by indication they make treatment induced effects more salient. Phase 1 healthy volunteer
173 studies of SSRIs demonstrate convincing evidence of treatment related suicide and dependence from the
174 early 1980s.

175 In 2004, three years after the Miller case, FDA put black box warnings on sertraline for adolescents in
176 2004, on the basis of data generated prior to the Daubert hearing in this case.

177 **4. SSRIs and suicide: National suicide data**

178 As it became clear that the clinical trial data did not help the defendants in SSRI and suicide cases,
179 the defence argument shifted to data from national suicide rates and other cohort data. It was argued
180 that were SSRIs causing suicide then national suicide rates would increase but that in fact since the
181 early 1990s with increasing consumption of SSRIs national suicide rates have fallen [19]. The data put
182 forward resembled the data put forward in arguments about smoking and lung cancer pointing to rising life
183 expectancy in western countries in parallel with increased consumption of cigarettes. Such correlational
184 data clearly create doubt – as was intended in the case of tobacco and cancer. Data on selected windows of
185 antidepressant consumption and apparent national suicide rates can do likewise and have almost certainly
186 been used for this purpose.

187 The data in such studies have universally omitted national suicide rates from the 1960s, 1970s and
188 early 1980s along with data for antidepressant consumption during these periods. This was a time when
189 antidepressant consumption was increasing and these drugs if beneficial were being given to patients
190 who were more severely depressed than those treated during the 1990s and hence it was more likely that
191 these antidepressants would lead to lower suicide rates but in fact national suicide rates in most countries
192 in which these drugs were used increased during this period.

193 Another factor can help account for rising suicide rates through the 60s and 70s and falling suicide
194 rates from the 1990s onward. Autopsy rates rose during the 60s and began to fall from the early 1980s
195 onwards when the decline in suicide rates was first noted – some years before SSRIs came into use.
196 Several studies have since shown a closer correlation between autopsy and suicide rates than between
197 antidepressant consumption and suicide rates – when less autopsies are done, less suicides are diagnosed
198 [19, 20].

199 Studies of drug consumption bring out further factors likely to apply to all epidemiological studies of
200 drug use and adverse events. In the case of any drug, there will be a shifting relationship between the

201 amount of drug consumed in any one year and the number of people newly exposed to the drug. In the
202 case of the SSRIs, for instance, early consumption came from patients exposed to the drug for the first
203 time but in later years those taking the drugs chronically come to account for an ever larger proportion
204 of the amount of the drug consumed in any one year. If the drug causes a problem acutely rather than
205 chronically one can predict that levels of an adverse event such as suicide would level out in the face of
206 increasing consumption and might even fall depending on the suicide risk of the patients being treated –
207 a switch from depressive to anxiety disorder patients for instance [21].

208 **5. Paroxetine and birth defects: The role of epidemiology**

209 Unlike cases involving suicidal acts, in which subjects can be challenged, and dechallenged, doses can
210 be changed, and antidotes introduced, the risk that a drug might trigger birth defects cannot be investigated
211 in this way. All that can be done is to design observational studies that attempt to control for confounders.
212 It was recognition of these difficulties after claims of Bendectin induced injuries that led to the design of
213 a series of such studies that appeared at the time to indicate the drug did not cause the injuries imputed
214 to it.

215 In the case of paroxetine, a meta-analysis of all epidemiological studies designed to look at the rate
216 of birth defects in women taking this drug during pregnancy, shows a 1.5 fold increase in the risk of
217 major birth defects and of cardiac defects on paroxetine compared with non-treatment [22]. These data
218 are consistent with the possibility that paroxetine can cause birth defects and legal actions in the United
219 States have proceeded on this basis.

220 This helps bring out two inter-related issues. First, no question has been raised in court about the
221 necessity for the relative risk of defects on paroxetine to be 2.0 or greater. Furthermore on the basis of
222 data pointing to a relative risk of 1.5, FDA has categorized paroxetine as pregnancy category D. This
223 means that de facto paroxetine is regarded as “causing” birth defects – with the word cause used in the
224 labelling.

225 It is a moot point however whether this should be regarded as a scientific statement. It might be better
226 regarded as a policy statement. The scientific data are not inconsistent with the statement but as outlined
227 above epidemiological and controlled trial data cannot determine what is actually happening or has
228 happened in an individual case. The reasons for this are developed further below.

229 Were the scientific data inconsistent with the claim that the treatment causes birth defects, it would be
230 difficult to take a case, but once it is clear that the evidence is not inconsistent with the possibility of a
231 problem, a case could proceed in the United States.

232 The next step is to determine whether there are grounds to think that the possibility of injury was realised
233 in the individual plaintiff. Here a clinical judgement is called for. It is in this area that formulations of
234 reasonable medical certainty or on the balance of medical probabilities first arose. These formulations
235 refer to efforts on the part of clinical experts to come to clinical judgements in individual cases. Are there
236 other factors that could have led to this injury in the index patient? Given that judgements have to be made
237 about an individual case, this exercise cannot provide mathematical basis for relative risk assessments or
238 assignment of probabilities.

239 In the case of injuries such as birth defects, where there is ordinarily no possibility of making an assess-
240 ment in the individual case based on observed prodromal changes in response to challenge, dechallenge
241 or rechallenge, or informed by evidence from differing doses of the drug, such judgements are based
242 primarily on exclusions of other possible causes.

243 But American legal cases against pharmaceutical companies (as opposed to prescribers) do not proceed
244 as a scientific debate. The courts are not called to make a judgement as to whether for instance paroxetine
245 causes birth defects. The cases taken are actions regarding a company's failure to warn about a hazard or
246 negligence in failing to design studies to establish the existence of a hazard that could have been expected
247 on the basis of other pertinent studies. The focus is on company behaviour rather than on the drug.

248 In contrast, under the United Kingdom Consumer Protection Act stemming from a European Directive,²
249 in actions against a pharmaceutical company that concern the effects of a drug, the focus is on the defects
250 of the drug. As all prescription drugs are risky, this raises the question as to when an action can be taken
251 on the basis of a defect in the drug.

252 Within the United Kingdom, the only case precedent is *A and Others v National Blood Authority and*
253 *Others*; this offers a basis for actions in the case of drugs that are worst in their class [23]. Such a basis is
254 clearly acceptable in the case of a treatment induced injury linked to a defective batch of medication. In
255 such a case, there will clearly be no epidemiological or other such evidence brought to bear on the issue
256 as a scientific study would never include a drug with defects of this kind.

257 The issues become more complicated in legal actions for injuries that stem from a source other than
258 a defective batch of the drug. These issues have recently come to the fore in legal actions regarding
259 paroxetine and dependence in Europe.

260 6. Paroxetine and dependence: The role of epidemiology

261 In contrast to birth defects, where exploring a linkage with treatment by challenge and dechallenge
262 is impossible so that investigators have little option but to turn to epidemiological studies, dependence
263 on a drug can only be established by dechallenge and rechallenge. In the case of paroxetine, there is a
264 classic time from stopping exposure to appearance of the problem, relief on reinstitution of treatment,
265 dose sensitivity, relief by an appropriate antidote and plausible biological mechanisms to underpin some
266 of the manifestations of the problem.

267 The profile of changes was so clear that after paroxetine was launched clinically, its use led to a greater
268 volume of published case studies outlining dependence on and a withdrawal syndrome on discontinuation
269 from it than existed for other antidepressants. Despite confounding by indication, it would appear that
270 clinicians were happy that the notional prior probabilities pointed to a probability substantially in excess
271 of 50%, even though none of these reporters were likely to have known that the propensity of paroxetine
272 to cause dependence had in fact first been noted in healthy volunteer studies before marketing.

273 These publications are a convincing manifestation of both the frequency and severity of the problem.
274 The volume of reports from doctors to regulators on paroxetine and dependence was significantly greater
275 than for other antidepressants [24]. It should also be noted doctors are encouraged only to report to
276 regulators problems that are serious – that is problems that lead to significant morbidity or mortality.

277 By analogy, were there to be reports of people surviving falls of 18,000 feet by using a parachute, no
278 one would call for controlled trials of parachutes or evidence of a relative risk of survival using parachutes
279 of greater than 2.0. Such a trial would be both uncalled for and unethical.

280 The situation was so stark that a competing pharmaceutical company making Prozac could take the
281 usual step of running adverts and supporting symposia at meetings portraying their drug as preferable to
282 paroxetine as it was less likely to cause withdrawal.

² The EC Council Directive on liability for defective products, 85374/EEC dated 25 July 1985.

283 In the United States actions were filed against GlaxoSmithKline for dependence on paroxetine. These
284 actions led to the discovery of a pattern of company behaviour that including the ghostwriting of articles
285 denying the problem and instructions to sales staff on how to deflect concerns about dependence. The
286 actions were quickly resolved. The labelling of the drug changed to reflect the risks of dependence.

287 In the United Kingdom, a legal action regarding paroxetine was filed initially on the basis that paroxetine
288 was liable to cause suicide, and violence, in addition to dependence. The cases recruited following suicides
289 or suicide attempts or violent episodes were subsequently dropped as the evidence was not convincing
290 that the effects of paroxetine were any worse in this regard than other SSRIs. The action proceeded for
291 those claimants who alleged physical dependence.

292 Given current legal constraints and the dominance of the “epidemiological paradigm” within medicine,
293 epidemiological evidence was seen to be of paramount importance. There was also a premium on data
294 that might demonstrate a doubling of the relative risk on paroxetine compared to other drugs in its class
295 in order to establish that paroxetine was worst in its class.

296 There are multiple incoherencies here. First as outlined above, epidemiological evidence should only
297 have an *a priori* premium in a subset of cases such as birth defects or delayed injuries from devices or
298 toxins or chemicals. To hold that a case cannot otherwise go forward in the absence of epidemiological
299 or controlled trial evidence is simply wrong. Second epidemiological evidence is never unequivocal – a
300 judgement has to be made about its generalizability and notions such as a relative risk of 2.0 should be
301 abandoned as determinants of the strength of the evidence. Such rules substitute inappropriate formulae
302 for thought.

303 Even if the position were not incoherent, in the case of pharmaceutical agents a full set of studies
304 comparing one drug with another in its class almost by definition will not exist. Expecting such data to
305 exist demonstrates a fundamental ignorance of modern marketing.

306 The fact that data does exist demonstrating that Seroxat/Paxil has an increased propensity to cause
307 dependence is quite extraordinarily, but this data exists only because already existing case series made it
308 clear that were a controlled trial undertaken the results could only come out one way. It is the pre-existing
309 publications of case studies that constitute the core scientific data in this case, in addition to a greater
310 number of reports of dependence to regulators worldwide for paroxetine than for other drugs. In the
311 patient domain, there were double the number of participants in and posts to paroxetine linked websites
312 than there are for other antidepressants and these posts have centred primarily on the risks it posed of
313 dependence. This is convincing evidence that paroxetine is worst in its class.

314 Given greater evidence for dependence on paroxetine than for other agents, with a consequent likelihood
315 that more women of child-bearing years will be taking it during the critical first trimester of pregnancy
316 than will be taking other drugs, and the fact that it is labelled as causing birth defects where other SSRIs
317 or other antidepressants are not, there is clear evidence for the severity of the consequent problems
318 dependence on this drug can cause compared to others in its class.

319 Finally in addition an investigation of studies undertaken by GlaxoSmithKline reveals that paroxetine
320 doubles the rates of suicidal and aggressive behaviours during the withdrawal period; there are no compa-
321 rable data for any other drug (Tables 1, 2).³ These studies also reveal that suicidal and related behaviours
322 are more common on withdrawal from paroxetine than for other SSRIs in general (Table 3).

323 The fact that a legal case like this might fail, when the defect in the drug is so apparent that it would be
324 unethical to undertake controlled trials points to difficulties in the way European law is framed but also a

³ These tables are assembled by the author from GlaxoSmithKline’s Article 31 submission to the Medicines’ and Healthcare Regulatory Agency, available on the company website

Table 1

Incidence of possibly suicide-related events in patients:
All placebo controlled trials 30 days post-taper

	Paroxetine %	Placebo %	Odds ratio
Overall	33/9219 0.36	8/6455 0.124	2.90 (1.34, 6.25)
Depression	22/3799 0.579	3/2402 0.125	4.63
Non-depression	11/5420 0.203	5/4053 0.123	1.65

Table 2

Incidence of suicide-related & hostility events in patients: All placebo controlled trials 30 days post-taper

	Paroxetine %	Placebo %	Odds ratio
Overall	42/9219 0.456	8/6455 0.124	3.68 (1.73, 7.82)
Depression	24/3769* 0.637	3/2402 0.142	4.45
Non-depression	18/5450 0.330	5/4053 0.123	2.68

Table 3

Incidence of suicide-related & hostility events in patients:
All adult comparator controlled trials 30 days post-taper

	Paroxetine %	Comparator %	Odds ratio
Overall	28/6522 0.429	15/4969 0.302	1.42
Tricyclic	13/2953 0.440	10/2714 0.368	1.20
SSRI	7/1200 0.583	2/1218 0.164	3.55

325 failure on the part of both the legal and scientific communities to understand how adverse events should
326 be assessed.

327 7. The role of randomized controlled trials (RCTs) in adverse event research

328 After a drug is launched few of the adverse events later included in a drug's label appear there on the
329 basis of controlled trial or epidemiological evidence. Few drugs are withdrawn or end up carrying warn-

ings on the basis of controlled trial or epidemiological research. In contrast drugs have been withdrawn on the basis of as few as three or four reported cases, pointing to the greater importance of analyzing key events to simply assembling a quantity of events and supposing these quantities will substitute for thought. (The latest drug withdrawn was efalizumab, Raptiva, following 4 cases of progressive multifocal leukoencephalopathy). More generally studies of clinician descriptions of adverse events have demonstrated that they are likely to be correct in anything between 80–100% of instances [25].

While RCTs are trumpeted as the Gold Standard in other areas of therapeutics, they are of lesser help in the area of adverse events. When interventions, such as parachutes, clearly have an effect RCTs are not needed. Many adverse events – such as the physical dependence on paroxetine – are close to equally unequivocal. This led doctors and others to report such effects in the clinical trials literature. Indeed given the consensus that rapidly developed about this effect, controlled trials would likely have been unethical.

It is when there is doubt about whether an effect should be linked to treatment that RCTs may be of help. In terms of adverse effects, RCTs relate more closely to observational studies of the type undertaken to establish whether a drug might cause birth defects, or a device might lead to longer term injuries. They may help to bring effects into view that cannot readily be studied with challenge-dechallenge and related procedures.

Furthermore, many agents, particularly those active on the central nervous system, whether SSRIs or dopamine agonists, have a track record in producing dual effects. Thus many of these drugs will sedate some and alert others, cause anorexia in some and increase appetite in others, cause loss of libido in some and increase libido in others, cause dyskinesias in some or alleviate dyskinesias in others, cause anxiety in some or be anxiolytic in others. These dual effects have left some claiming on the basis of controlled trial data that these agents are no more stimulant than, or have no more effect on weight than placebo, with the implication that these treatments are neutral in their effects across a variety of functions when this is not the case.

In the case of treatment induced suicidality, if antidepressants are helpful to some and cause problems to others then simply consulting the data on outcomes may give a misleading picture. It is quite possible that antidepressant treatment might lead to a reduction in suicidal events compared to placebo in some while still triggering suicidality in others. If the design of the RCT has not taken this into account, the data may mislead. This possibility raises further problems in interpreting relative risk, in that a relative risk of 1.0 might stem from a study in which some have been protected by treatment in which case the risk for a subset of the study population must be greater than 1.0 – but these data give no clue by how much.

The issue of suicidality brings out questions about the appropriate choice of outcome measure. In this case, the possible outcome measures include completed suicides, suicidal acts and suicidal ideation. It appears from the data that there is a clearer link to treatment when completed suicides or lethal and non-lethal suicidal acts are the outcome measure rather than suicidal ideation. Using a composite measure of suicidality might seem reasonable in the case of uncommon events as it will lead to an increased frequency of events under consideration but the results in the case of the antidepressants and suicidality suggest that the strategy may mislead.

A further issue concerns the choice of study population. This is especially true of events that may be confounded by indication. Consider the example of studies 057 and 106, two placebo controlled trials of paroxetine in “intermittent brief depression” – also termed borderline personality disorder.

In the late 1980s, Eli Lilly, faced with an excess of suicidal behaviors in fluoxetine trials, undertook a placebo-controlled trial of fluoxetine in a group of borderline personality disorder patients. This trial was stopped; placebo was sweepingly statistically superior to fluoxetine [10]. SmithKline then undertook a

Table 4
Data on suicides & suicidal acts from paroxetine placebo controlled trials

	Paroxetine suicides/ no. subjects	Placebo suicides/ no. subjects	Paroxetine suicidal acts/ no. subjects	Placebo suicidal acts/ no. subjects
All Paroxetine studies minus 057 & 106	1/8172	0/5391	16/8172	4/5391
Major depression	1/2943	0/1671	10/2943	0/1671
Studies 057 & 106	0/147	0/151	32/147	35/151
Total	1/8,319	0/5,542	48/8,319	39/5,542

trial in the same patient group, study 106, using the same lead investigator. This study also terminated early showing a doubling of suicidal acts on paroxetine, and remained unpublished. Despite these results in this group of patients, SmithKline undertook a further placebo controlled paroxetine trial in this patient group, study 057.

In April 2006 GlaxoSmithKline reported on suicidal acts from their placebo controlled paroxetine trials – See Table 4.⁴

When the data are analysed without studies 057 and 106, they show a 2.8 fold (95% C.I., 0.94, 8.33; $p = 0.05$) increase in risk of a suicidal act on paroxetine compared to placebo. Studies 057 and 106 are clearly anomalous generating 67 suicidal acts in 298 patients compared to 21 suicidal acts in 13,563 patients from all other trials. Including them in the analysis would lead to a relative risk of a suicidal act on paroxetine of 0.86. These studies may well have been undertaken explicitly for this purpose.

8. The role of statistics in adverse event research

When looking at suicidal acts on fluoxetine or heart attacks on rofecoxib, in the face of a several fold excess of such acts on active treatment compared to placebo, companies have denied a relationship between treatment and the adverse effect when the results are not statistically significant. More generally unless clinical trials have demonstrated that adverse events occur at a greater rate than on placebo to a statistically significant extent, companies may choose not to declare them on the label of a drug [26]. At present this use of statistics may be the single most common “tool” used by companies to hide adverse events.

Tests of statistical significance were created to be used to test a pre-designated primary outcome in experiments designed specifically to test for that outcome. They should not be applied to other outcomes in the same study, determined retrospectively, or in studies that assemble the data from other studies. In these cases the data should be described in terms of confidence intervals, with retrospective groupings made transparent.

Even when described in terms of confidence intervals, companies and others keen to argue against a role of treatment in triggering adverse events will claim that if the confidence interval includes the

⁴ Update April 5th 2006, Paroxetine Adult Suicidality Analysis. gsk.com/media/paroxetine/briefing_doc.pdf (accessed 8th August 2007)

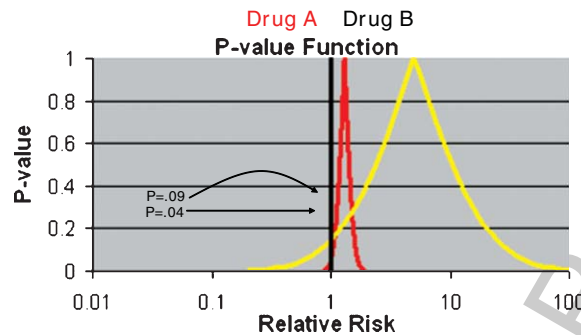


Fig. 1. Causing adverse events.

figure 1.0 that the results indicate that even an increased number of events on active treatment essentially constitutes no evidence for an increase in risk. This is unacceptable in situations where events (such as adverse reactions) are expected to be rare but possibly deadly.

Company interpretations can be tested readily using the example of two drugs, A and B, both of which produce a serious adverse event. In the case of drug A, the available data describes a doubling of risk on treatment and offers precise estimates of the effect so that the 95% confidence interval lies entirely to the right of 1.0. In the case of drug B, the data points to an 8-fold increase in risk, but perhaps because some aspect of the design of the experiment is less honed the distribution of the data is more scattered and the 95% confidence interval includes 1.0 (See Fig. 1).

Current company rhetoric would have courts believe that only drug A poses a risk, whereas in fact our best estimates are that drug B is roughly 4 times riskier than drug A.

A possible analogy is to offer anyone who disputes this interpretation the choice between playing Russian roulette with a gun where it is known for certain that there is one and only one bullet in one of the six chambers of the gun versus a gun where the best indications are that three or four of the chambers contain bullets but uncertainty remains. It is unlikely but possible that there are no bullets in the gun, but equally likely that all six barrels are filled. Companies and their experts argue that courts should only regard the gun with one confirmed bullet in it as posing any risk, whereas in real life if forced to choose any sensible person would choose the gun with a single bullet.

In a case taken by investors in pharmaceutical company stock, the United States Supreme Court recently offered a ruling pertinent to these issues [27]. In this case, Matrixx Pharmaceuticals had failed to declare anosmia as a possible side effect of their cold remedy, Zicam, on the basis that there were no studies showing a statistically significant increase in the risk of anosmia on Zicam. The court however noted that a plausible biological mechanism for such an effect had been demonstrated in studies on animals and on this basis that *investors* had a right to know about this risk.

9. Summary: The rhetoric of adverse events

This review of the history of legal actions surrounding the effects of SSRIs makes it clear that claims about the appropriate evidence to use in any debate about adverse events at present are likely to be rhetorical statements rather than statements that stem from a scientific evaluation of the problem that has led to agreement on the best method to study the phenomenon. It makes it clear that in principle there can be no “gold standard” method to tackle the issues raised by particular problems.

At present a combination of European law and both medical and legal approaches to evidence within European settings has made it virtually impossible to take a successful action against a pharmaceutical company, where actions for the same injuries succeed in North America. European lawyers commonly put their failures down to factors such as having to pitch to judges rather than juries, to the ability of American lawyers to embarrass company witnesses in depositions, to the existence of possible awards for punitive damages against companies, and to a no-loser pays system. They rarely acknowledge that depositions, Daubert and other procedures and jury trials give companies great scope to damage expert testimony for the plaintiff [28].

The confusions surrounding this series of cases and the evident need to weigh sources of evidence in a different manner in different cases strongly supports the establishment of forensic pharmaco-vigilance as a discrete field of expertise [1]. Simply accepting received wisdom is recipe for being duped by rhetoric.

Appendix 1: Hierarchy of evidence

In considerations of evidence based medicine it has been traditional to order the evidence in the hierarchy laid out below. A similar re-organisation of the evidence hierarchy has been proposed by Vandembroucke [29], who proposed a divide into evaluative or exploratory hierarchies. The difference proposed here is driven by a specific consideration of adverse events.

Table 5A

Traditional evidence hierarchy for evaluative purposes

Level 1	Systematic review of all RCT evidence
Level 2	Randomized controlled trials
Level 3	Prospective cohort & Case control studies
Level 4	Retrospective cohort & Case control studies
Level 5	Case series
Level 6	Case reports

Table 5B

Adverse event evidence hierarchy

Level 1	Systematic review of all publications on the event
Level 2	Case series meeting quality standards
Level 3	Case reports meeting quality standards
Level 4	Controlled trials undertaken for other purposes
Level 5	Other observational studies not specifically designed for this purpose

References

- [1] D. Body, I.R. Edwards, K. Hartigan Go, A. Herxheimer, J. Labadie and D. Healy, Is there a need for forensic pharmacovigilance as a specialty, *Journal of Risk & Safety in Medicine* **23** (2011), 31–42.
- [2] M. Angel, *Science on Trial*, Norton, New York, 1997.
- [3] P.W. Huber, *Galileo's and Revenge: Junk Science in courtroom*, Basic Books, New York, 1991.

- 452 [4] M. Green, *Bendectin and Birth Defects*, The challenges of mass toxic substances litigation, University of Pennsylvania
453 Press, Philadelphia, 1996.
- 454 [5] McKay Queen's Bench Division: High Court (London): XYZ and others (claimants) versus (1), Schering Health Care
455 Limited (2) Organon Laboratories Limited and (3) John Wyeth and Brother Limited. Judgement by the Hon. Mr Justice
456 Mackay, London, 29 July 2002, Case No. 0002638.
- 457 [6] A.B. Hill, Reflections on the controlled trial, *Annals of The Rheumatic Diseases* **25** (1966), 107–113.
- 458 [7] F.E. Karch and L. Lasagna, Towards the operational identification of adverse drug reactions, *Clinical Pharmacology &*
459 *Therapeutics* **21** (1977), 247–253.
- 460 [8] M.N.G. Dukes, M. Mildred and B. Swartz, *Responsibility for Drug-Induced Injury*, IOS Press, Amsterdam 1998.
- 461 [9] Federal Manual Federal Judicial Centre, *Reference Manual on Scientific Evidence*, Washington DC, 1994, 160–161.
- 462 [10] D. Healy, *Let Them Eat Prozac*, New York University Press, New York, 2004.
- 463 [11] K.H. Vratil, Order to Show Cause, August 18th, In Miller Vs Pfizer, 2000.
- 464 [12] E.H. Turner, A.M. Matthews, E. Linardatos, R.A. Tell and R. Rosenthal, Selective publication of antidepressant trials and
465 its influence on apparent efficacy, *New England Journal Medicine* **358** (2008), 252–260.
- 466 [13] D. Healy and D. Cattell, The Interface between authorship, industry and science in the domain of therapeutics, *British*
467 *Journal of Psychiatry* **182** (2003), 22–27.
- 468 [14] M. Newman, The rules of retraction, *BMJ* **341** (2010), 1246–1248.
- 469 [15] D. Healy, Did regulators fail over selective serotonin reuptake inhibitors, *BMJ* **333** (2006), 92–95.
- 470 [16] D. Healy, The Antidepressant Tale: Figures signifying nothing? *Advances in Psychiatric Treatment* **12** (2006), 320–328.
- 471 [17] C. Beasley, Draft Rechallenge Protocol by Dr Charles Beasley, 1990, in testimony of Dr G. Tollefson in Forsyth Vs Eli
472 Lilly, Hawaii March 24th 1999.
- 473 [18] O. Caster and I.R. Edwards, Reflections on attribution and decision in pharmacovigilance, *Drug Safety* **33** (2010), 257–260.
- 474 [19] S. Reseland, J. Le Noury, G. Aldred and D. Healy, National suicide rates 1961–2003: Further analysis of Nordic data for
475 suicide, autopsies and ill-defined death rates, *Psychotherapy and Psychosomatics* **77** (2008), 78–82.
- 476 [20] N.D. Kapusta, U.S. Tran, I.R. Rockett, D. De Leo, C.P. Naylor, T. Niederkrotenthaler, M. Voracek, E. Etzers-
477 dorfer, G. Sonneck, Declining autopsy rates and suicide misclassification. *Archives of General Psychiatry* 2011;
478 doi:10.1001/archgenpsychiatry.2011.66.
- 479 [21] D. Healy and G. Aldred, Antidepressant drug use and the risk of suicide, *International Review of Psychiatry* **17** (2005),
480 163–172.
- 481 [22] J.A. Cole, S.A. Ephross, I.S. Cosmatos and A.M. Walker, Paroxetine in the first trimester and the prevalence of congenital
482 malformations, *Pharmacoepidemiology and Drug Safety* **16** (2007), 1075–1085.
- 483 [23] A and Others v, National Blood Authority and Others, 26th March 2001, 2001WE, 239806, [2001] 3 All ER 289.
- 484 [24] M.M. Stahl, M. Lindquist, M. Pettersson, I.R. Edwards, J.H. Sanderson and N.F. Taylor, Withdrawal reactions with selective
485 serotonin re-uptake inhibitors as reported to the WHO system, *European Journal of Clinical Pharmacology* **53** (1997),
486 163–169.
- 487 [25] G. Venning, Validity of anecdotal reports of suspected adverse drug reactions: The problem of false alarms, *BMJ* **284**
488 (1982), 249–253.
- 489 [26] D.W. Ogden, Brief for the Pharmaceutical Research and Manufacturers of America and the biotechnology industry orga-
490 nization as Amici Curiae supporting petitioners. In *Matrixx Initiatives et al v James Siracusano and NECA-IBEW Pension*
491 *Fund*, Supreme Court of the United States, 09–1156, 2010.
- 492 [27] Supreme Court of the United States (2011). *MATRIXx INITIATIVES, INC., ET AL. v. SIRACUSANO ET AL.* March
493 22nd, Cite as 563 US.
- 494 [28] M.N.G. Dukes, The case of Mrs Grundberg: A retrospective view, *International J of Risk and Safety in Medicine* **22** (2010),
495 215–225.
- 496 [29] J.P. Vandenbroucke, Observational research, randomized trials, and two views of medical science, *PLoS Medicine* **5** (2008),
497 e67.