

## Perspective

# The Flawed Basis for FDA Post-Marketing Safety Decisions: The Example of Anti-Depressants and Children

Donald F Klein<sup>\*,1</sup>

<sup>1</sup>New York State Psychiatric Institute, Columbia University, New York, NY, USA

The FDA (February 3, 2005) issued a black box warning that all antidepressants increase the risk of suicidal thinking and behavior in children and adolescents that must be cited in all advertising as well as included in the package insert. Following this, there was a sharp decrease in antidepressant prescriptions for children with uncertain public health impact. The current black box does not claim that these medications increase the risk of completed suicides, although this is the clear implication of the term 'suicidality'. The interpretation by the press is unequivocal that lethal outcomes prompted this action. This review concludes the following: (1) Since no suicide occurred in clinical trials of approximately 4400 children, the analyses relied upon 'suicidality' as a surrogate. (2) The classification of adverse events by the Columbia group necessarily relied on inferences, because the available evidence was not prospectively collected for this purpose. (3) The data analysis relied on a composite variable labeled 'suicidality', an unvalidated, inappropriate surrogate. Specific criticisms of analytic procedures and inferences are presented. The failure of the FDA's post-marketing surveillance system is reviewed. The data necessary for objective evaluations of possible post-marketing harm cannot be gathered by the current process. Proper prospective post-marketing surveillance by linked computerized medical records is a crucial issue that deserves major public and political attention and prompt action. *Neuropsychopharmacology* (2006) **31**, 689–699. doi:10.1038/sj.npp.1300996; published online 14 December 2005

**Keywords:** FDA; antidepressants; drug safety; drug efficacy; child/adolescent suicide; postmarketing surveillance

## EXECUTIVE SUMMARY

The FDA (February 3, 2005 (US Food & Drug Administration, 2005)) issued a black box warning that all antidepressants increase the risk of suicidal thinking and behavior in children and adolescents. The black box warning template is appended (Appendix). It must be cited in all advertising as well as included in the package insert. In addition, a new medication guide must be distributed by pharmacists to all who obtain such medications. Close observation by therapists and families is considered necessary. Following this warning, there was a sharp and rapid decrease in antidepressant prescriptions for children (Vedantam, 2005b), with uncertain public health impact. The American Medical Association and American Psychiatric Association both warned against decreasing access to these drugs for patients who may benefit from them.

The basis for this FDA action is the following summary statement:

\*Correspondence: Dr DF Klein, New York State Psychiatric Institute, Columbia University, 1051 Riverside Drive, New York, NY 10032, USA, Tel: +1 212 543 6685; Fax: +1 212 543 6945; E-mail: donaldk737@aol.com

Received 11 August 2005; revised 24 October 2005; accepted 25 October 2005

Online publication: 31 October 2005 at <http://www.acnp.org/citations/Npp103105050501/default.pdf>

'Pooled analyses of short-term (4 to 16 weeks) placebo-controlled trials of 9 antidepressant drugs (SSRIs and others) in children and adolescents with major depressive disorder (MDD), obsessive compulsive disorder (OCD), or other psychiatric disorders (a total of 24 trials involving over 4400 patients) have revealed a greater risk of adverse events representing suicidal thinking or behavior (suicidality) during the first few months of treatment in those receiving antidepressants. The average risk of such events in patients receiving antidepressants was 4%, twice the placebo risk of 2%. No suicides occurred in these trials.' (FDA, February 3, 2005 (US Food & Drug Administration, 2005))

The current black box does not claim that these medications increase the risk of completed suicides, although that is the clear implication of the term 'suicidality'. The interpretation placed upon this statement by the press is unequivocal that lethal outcomes prompted this action. Further, many newspaper headlines explicitly refer to medication-induced suicide rather than 'suicidality'. For instance, the Washington Post headline: 'Drugs Raise Risk of Suicide; Analysis of Data Adds to Concerns on Antidepressants'. (Vedantam, 2005a)

Below, this critique concludes:

### (1) *Surrogate Markers: Role and Definition*

The central concern of the FDA and its advisory boards was whether antidepressants could be lethal, by causing

suicide in children and adolescents. Since no suicide occurred in these clinical trials that studied approximately 4400 children, the analyses relied upon 'suicidality' as a surrogate.

- (2) The classification of adverse events by the Columbia group necessarily relied on inferences, because the available evidence was not prospectively collected for this purpose. Thus, it does not fulfill the requirements of definitions used by standardized scales of 'suicidality', such as inquiry concerning intent to die and definite plan. In particular, the inclusion of 'ideation' does not attend to either intent or plan.
- (3) *Strategies used in data analysis*  
The data analysis relied on a composite marker of 'suicidality', including 'ideation'. This was an inappropriate, misleading surrogate for completed suicide that grossly overestimates the potential risk for a rare event.
- (4) The failure of the FDA's post-marketing surveillance system is reviewed.  
Alternative methods for systematically collecting safety data are cited. It is emphasized that the data necessary for objective evaluations of possible post-marketing harm cannot be gathered by the current FDA process. Proper, prospective, post-marketing surveillance by linked computerized medical records is a crucial issue. Major public and political attention and appropriate action are required.

## SURROGATE MARKERS: ROLE AND DEFINITIONS

The transcript of the Joint Meeting of the CDER Psychopharmacologic Drugs Advisory Committee and the FDA Pediatric Advisory Committee, held during September 13–14, 2004, vividly presents the concerns of FDA officials, committee members, and parents of children who killed themselves or others while taking prescribed SSRI medication (HHS, 2004). The families were outraged that they had not been informed of the increased likelihood of a suicidal fatality from antidepressant treatment. They argued that such notification would have led them to refuse such dangerous treatment, thus preventing their tragic, unnecessary, losses.

Representative quotes follow:

MS VAN SYCKEL: Good afternoon. My name is Lisa Van Syckel. The FDA and the pharmaceutical industries have repeatedly stated that it is the disease, not the drug, that causes our children to become violent and suicidal. It wasn't the disease that caused my daughter to viciously mutilate herself; it was the drug. It wasn't the disease that caused my daughter to become violent and suicidal and out of control. It wasn't the disease that caused her to scream the words 'I want to die.' And, it sure as hell was not the disease that caused Christopher Pittman to kill the two people he loved the most, his grandparents. He had been on Zoloft just three weeks and he was 12 years old. Christopher is now facing life in prison as an adult. Pfizer refers to me and others as a detractor of SSRIs and that I am misinforming legislators with oversight responsibilities. As an adult, I am considered fair game for verbal attacks but, ladies and gentlemen, Pfizer crossed the line the day they attacked a dead child.

They viciously attacked a dead child and you all know it. And you, ladies and gentlemen, as adults, need to tell Pfizer that they need to stop.

MS MILLING-DOWNING: On January 10th, 2004 our beautiful little girl, Candice, died by hanging four days after ingesting 100 mg of Zoloft. She was 12 years old. The autopsy report indicated that Zoloft was present in her system. We had no warning that this would happen. This was not a child who had ever been depressed or had suicidal ideation. She was a happy little girl and a friend to everyone. ...Candice's death was entirely avoidable, had we been given appropriate warnings and implications of the possible effects of Zoloft. It should have been our choice to make and not yours.

We are not comforted by the insensitive comments of a corrupt and uncaring FDA or pharmaceutical benefactors such as Pfizer who sit in their ivory towers, passing judgments on the lives and deaths of so many innocent children. The blood of these children is on your hands. To continue to blame the victim rather than the drug is wrong. To make such blatant statements that depressed children run the risk of becoming suicidal does not fit the profile of our little girl (HHS, 2004).

The aggrieved families assumed that taking the medication directly caused the tragic outcome. The suggestion that a child's suicide may have been the result of their illness, rather than the medication, was dismissed as an outrageous aspersion. Some believed that the FDA failed by allowing such dangerous medications to be on the market, or by failing to recall them. Others believed that, since these drugs are apparently useful under some circumstances, a stern warning of pharmacological risk would suffice.

These presentations deeply concerned the Committee. However, its members recognized that self-selected clinical anecdotes cannot establish causality because of the *post hoc ergo propter hoc* fallacy. Their central dilemma was that, in the clinical trials that might allow a causal inference, there had been no completed suicides. The advisory committees' transcripts show their concern for first arriving at objective estimates of causal risks, before recommending FDA policy actions. Did the clinical trial data actually support the families' outrage about deaths due to suicide-provoking drugs (which might also induce homicide)? Therefore, if these randomized, blinded clinical trial data were to cast light on lethal outcomes, a reliable surrogate marker for completed suicide had to be defined. At this crucial point, the FDA process became questionable.

There is a large literature on the clinical, logical, and statistical necessities for defining a surrogate. Most discussion of surrogate utility occurs in the context of allowing FDA Accelerated Approval of investigational drugs. The general logical framework is that the surrogate is a more immediate and accessible variable that is on the direct causal pathway to an undesirable event. The assumption is that, if the surrogate can be diminished, then the clinical end point of interest also will be benefited.

There are examples of both good and bad decisions utilizing surrogates. A good decision was that agents lowering cholesterol levels would also lessen the likelihood of atherosclerotic disease. It is far easier to show that a medication can lower cholesterol levels than that it prevents myocardial infarction. However, bad predictions also occur.

Since arrhythmias were known precursors of cardiac death, the approval of certain anti-arrhythmic drugs was based on the presumption that this lowers the risk of cardiac death. However, the drugs were not beneficial. They actually increased the rate of cardiac death.

When the Advisory Committees and FDA needed to develop a useful surrogate for completed suicide, the cautions learned in this area should have been the framework for discussion (Fleming, 2005; Fleming and DeMets, 1996; Prentice, 1989; Freedman *et al*, 1992; DeGruttola *et al*, 1997; Lin *et al*, 1997; Buyse *et al*, 2000; Baker and Kromer, 2003). Instead, an ill-defined term, loosely labeled 'suicidality', was treated as an established causal surrogate. Numerous comments by Advisory Committee members illustrate confusion about this issue.

Following are transcripts from the FDA Advisory Committee's meeting (HHS, 2004):

DR IRWIN (Committee Member): Is there a word suicidality?

DR GOODMAN (Committee Chairman): Every time I write it in Word, it gets red underlined.

DR IRWIN: It seems to me, I mean to me, I am not certain anyone really knows what it is that we are saying and what you are voting on, or, to me, I would like to know what suicidality is.

DR GOODMAN: I don't think it is in an Oxford Dictionary either.

MS GRIFFITH (SGE Patient Representative): It is not in Webster's.

DR IRWIN: In a sense, it confounds things by, you know, the front page of the paper today, I think may lead to kind of a misrepresentation.

DR POLLOCK (Committee Member): Can't we just use the explicit language?

DR GOODMAN: That is, in part, what I would favor, is that if we use it, I think we need to at least parenthetically define what we mean when we are answering the question.

DR TEMPLE (FDA Associate Director for Medical Policy): Yes, that is what we do. I think that is what we actually did in labeling. Whether we should coin a new word is debatable, obviously, but it means suicidal behavior plus suicidal ideation. That is what we use it to mean as those items. (sic)

DR GOODMAN: Would it be fair for us to slightly modify the question, or do we have to take it as it is, because what I would say, if we could use the definition that corresponds to Outcome 3, I would feel most comfortable, because that corresponds to the reclassification and the way you approach the dataset. So, suicidality, suicide attempt, preparatory action/or suicidal ideation.

DR KATZ (FDA Supv. Medical Officer and Director of Division of Neuropharmacological Drug Products): Yes, you can certainly amend the question. We called it suicidal behavior and ideation, but it is clearly what is embodied in Codes 1, 2, and 6. (pp 213–215)

DR GOODMAN: Also, in talking about suicidality, there are some definitional questions that have come up all along, and I think we need to be, among ourselves, as clear as possible what we mean when we say suicidality. Maybe there will be some benefit from the work that the Columbia group has done to help us make sure that we are using the same language.

Also, I think it behooves us to try to translate what suicidality means to the general public. In looking at some samples of the morning papers, front page New York Times, front page USA Today, there are headlines about how it has been concluded already, based on yesterday's discussion, that the antidepressants increase suicidality in children.

I am trying to imagine. I would be interested in how a parent, in reading that, what they would think, what do they mean by suicidality. My guess is that they are going to think that it is suicide. As we discussed yesterday, this includes suicide, it includes suicide attempts, but our definitions also includes preparatory actions and ideation.

So, I think we need to be very clear that we are using the same terminology, and maybe Dr Posner (Columbia Task Force Chair) will be able to help us along the way in that. (pp 164–165)

First, I had mentioned earlier that there is some lack of clarity about the definition of suicidality. In fact, as we can see on the other screen, although there is quite clarity there, you can set the brackets either narrowly or broadly in terms of what we mean by suicidality.

For the most part, in the analysis that was presented yesterday, the definition of suicidality corresponded to Outcome 3, which included evidence of suicide attempt, preparatory actions or suicidal ideation.

So, I think before we take a vote on that question, there should be some discussion and maybe some guidance from the FDA, as well, is which definition of suicidality are we adopting for the purpose of that vote. (pp 208–209)

My feeling is—again, I pose this to the FDA—we cannot ignore the other information we heard from the public testimony about cases of completed suicide, and obviously, those are not from the trial, yet we can in some ways extrapolate from the ideation and behaviors in the trials to the risk of completed suicide that perhaps would exist in the absence of a carefully controlled environment, such as is the case in a clinical trial.

So, maybe I could start by posing the two questions to the FDA. One has to do with which definition of suicidality should we be entertaining, and, secondly, should we limit this answer to what we know from the clinical trials.

DR LAUGHREN (FDA Medical Officer in Psychiatry; Team Leader, Psychiatric Drug Products, Division of Neuropharmacological Drug Products): Our intent was that you focus on Outcome 3. (Adverse events coded as Suicide Attempt, Preparatory Action, or Suicidal Ideation. See below.) That was our primary endpoint in the trials, so that is what we intended by suicidality. I think for the purposes of this question, we would like you to focus on the clinical trials. (pp 209–210)

Note Dr Laughren puts aside the issue of surrogate validation. The term 'suicidality' is stipulated to refer only to a composite of Columbia-defined inferences re suicidal intent, derived from unstructured adverse event narratives. Whether 'suicidality', so defined, is relevant to completed suicides—or even suicide attempts—has not been established.

DR GOODMAN: I think we have a clarification on that and hopefully, the public will understand what we mean, too, and that, I think we will leave it to the press to do their job in trying to best define what we mean and don't mean by that term, specifically, that we are not talking about

actual completed suicide if we are restricting our deliberations to the clinical trials, because there weren't any instances. (pp 215–216)

DR FOST (Committee Member): Suppose there were no SSRIs, suppose they were contraindicated, that is, prohibited, approximately, let me just ask the question about suicides, about completed suicides, and I understand there is no suicides in the FDA data, but based on everything that we know, approximately, would there be more suicides, fewer suicides, or the same amount if there were no SSRIs in children? (p 31)

Note Dr Fost asks, even if it is assumed that SSRIs cause suicide, would banning them be a net loss or gain? Dr Temple's reply indicates that such data are unavailable and probably unresearchable. He also indicates the completely assumptive basis for considering 'suicidality' as a valid surrogate for completed suicide or attempt.

DR TEMPLE: There is not going to be any way to answer that, in part because you can't do rigorous studies of the kind that would answer that. No one is going to let you not treat, not institutionalize, et cetera, someone who is getting worse and worse, and it would require long-term studies presumably against no treatment, and it is not easy to figure out how anybody is going to do those. (pp 31–32)

There were no completed suicides in the pediatric data, so that doesn't give you a clue. You can form your own judgment about whether increased suicidal behavior or thinking is going to lead to suicides in a certain fraction of cases. It is hard to imagine that it couldn't, but you don't know what that ratio is.

The success rate of suicidal attempts is relatively low. I gather it is higher in males than females, but I don't think there is going to be ways to put numbers on that. (p 32)

DR TEMPLE: The difficulty in dealing with the question of completed suicides is that while, unquestionably, some of the cases reported sound pretty interesting and persuasive on the point, you have no idea how persuasive the decrease in suicide that other people alleged, how large that is.

So, how to say whether there is a net benefit or harm on completed suicides certainly is unclear to me. Those data are very hard to analyze quantitatively. That is not the same as saying that some people don't seem to get worse when they are on these drugs, but some people seem to get better also. So, how to put that in numbers that addresses that question, increasing, say, the risk of suicide, that seems very hard to do. (p 211)

Note concerns for lethal outcomes and well-defined suicide attempts were replaced by focussing on the composite variable, termed 'suicidality', used in FDA data analyses.

This definition incorporates not only actions evaluated as suicide attempts (often in the absence of articulated intent, plan, or injury) but also 'ideation', which cannot yield firm inferences about lethal intent. Although explicitly distinguished from completed suicide, 'suicidality' still played this central role in the voting about FDA black box warnings.

DR ROBINSON (Committee Member): I would vote yes in the sense that if we are really saying that there is a potentially fatal side effect that might occur in 2, 3 percent of children taking these drugs, I think we have to in some way make sure that that information gets out. I am not

really as concerned in some ways of black box bolding. I just think that we need to make sure that a potentially fatal side effect with 2 or 3 percent of the population needs to get out. (p 387)

## PREDICTING RARE EVENTS

### Validating Surrogate Markers of Rare Events

Surrogate validation becomes even more daunting when the clinical end point of interest is a rare event. The classic paper analyzing these difficulties, by Meehl and Rosen, delineated two types of error: false negatives and false positives (Meehl and Rosen, 1955). Correctly discerning the condition in question, in those who actually have the condition, is called 'sensitivity'. If sensitivity is low, many 'false negatives' will be missed. Declaring the absence of a condition, in those who actually do not have it, is termed 'specificity'. Low specificity incurs many false positives. In rare conditions, false positives far outnumber true positives—even with high sensitivity and specificity.

Subjects labeled with 'suicidality' will be a mixture of a few true, and many false, positives. Case-control studies compare those with a specific condition to those who do not have that condition, on a variety of antecedent variables. Those who complete suicide frequently made prior suicide attempts, whereas those who do not commit suicide rarely do so. However, 93% of 302 near fatal 'attempters' did not subsequently kill themselves, during a 5-year followup (Beautrais, 2004). Thus, statistical demonstration of a risk factor is far from predicting marked, frequent danger.

The findings for suicide confirm previous conclusions that, even among high-risk samples, the occurrence of suicide is too low to identify those individuals who are likely to die by suicide from those who are not. This conclusion has important clinical implications since it suggests the need for high-quality followup, treatment, and surveillance of all patients making serious suicide attempts rather than approaches that focus on providing care to those clinically deemed to be at risk of further suicidal behavior (Beautrais, 2004).

Those with the risk factor in question, who actually reach the end point of interest, compared (as a ratio) to those without the risk factor who nonetheless reach it, is expressed as 'relative risk' (RR): (the rate of completed suicide in those with 'suicidality') divided by (the rate of completed suicide in those without 'suicidality'). An RR of 1 indicates no evidence that the risk factor predicts the end point of interest. An RR greater than 1 indicates that such an association may exist, but may be a chance fluctuation requiring statistical analysis. Clearly, a tiny denominator can lead to a very high risk ratio. In rare disorders, those labeled by the risk factor will largely be false positives (low positive predictive value) with regard to clinical outcome, as illustrated by Gould *et al* (2003):

Each year 1 in 5 teenagers in the United States seriously considers suicide. ...5% to 8% of adolescents attempt suicide, representing approximately 1 million teenagers. ...and approximately 1600 teenagers die by suicide. (Gould *et al*, 2003) Therefore, a conservative estimate is 12.5 million teenagers, of whom 2.5 million consider suicide annually.

A history of a prior suicide attempt is one of the strongest predictors of completed suicide, confirming a particularly higher risk for boys (30-fold increase) and a less elevated risk for girls (3-fold increase). (Ibid.)

Assuming 12.5 million adolescents, the suicide rate is 0.0128%. In boys, approximately 0.5 million attempted suicide and 5.75 million did not. To allow maximum predictability, we assume that all 1600 completed suicides occurred in boys.

Calculation indicates that, among those who attempt suicide, completing suicide occurs in 0.232%. Among those who did not previously attempt suicide, the suicide rate is 0.008%. Therefore, the rate of completed suicide is 30 times greater in attempters than in nonattempters. Also, 72.5% of completed suicides had an antecedent attempt. Although attempts are clearly a risk factor for completed suicides, 99.77% of attempters will not commit suicide (false positives). This seemingly large RR = 30 is taken by some to show that suicide attempts are directly on the causal path to completed suicide. However, since less than 0.3% of attempts proceed to completed suicide, ‘attempts’ must be extraordinarily heterogeneous.

The predictive ability of suicide attempts toward completed suicide can also be estimated by the correlation; Phi coefficient = 0.038. For the even more common suicidal ideation, predictability is, of course, worse. Therefore, only very few—or perhaps none—of the small number of events labeled ‘suicidal attempts’ in the FDA-reviewed studies are actually on a suicide path. The attempt to demonstrate a causal attribution to medication, specifically for suicide attempts as inferred from adverse events, did not approach statistical significance (see below).

Notably, the interview scale items regarding suicide actually point in the opposite direction. Consistent terminology would refer to this as an ‘anti-signal’. Intermediate algebra demonstrates that the term ‘suicidality’ carries an altogether erroneous impression of a predictable public health menace.

An opposing argument might be that these estimates are based on annual rates, which may be substantially increased in depressed adolescents. There is limited prospective data incorporating both attempts and suicide rates. The Maudsley 20-year followup of 245 depressed adolescents (with 36% comorbid conduct disorders) found a 2.45% risk of completed suicide (N=6), but a 44.3% lifetime risk of attempted suicides.

Both the maximum possible Phi coefficient and true positive rate are equal to 0.056 (Fombonne *et al*, 2001). These estimates of predictability are quite similar to the previous example.

**CLASSIFICATION OF EVENTS BY THE COLUMBIA GROUP**

Unstructured adverse event narratives from 24 pediatric placebo controlled trials—conducted over a nearly 20-year period, with durations ranging from 4 to 16 weeks—were provided to the Columbia group. There was no common format, nor were data collected by a standard procedure across these trials. Most trials addressed MDD, but OCD, generalized anxiety disorder, social phobia, and attention

deficit hyperactivity disorder were also studied. All adverse events identified by industrial sponsors, as well as all serious adverse events, all accidental injuries, and all accidental overdoses, were independently, blindly, categorized by the experts assembled under the leadership of Dr Kelly Posner (Columbia University).

Dr Posner described their approach to categorization to the FDA committee. Owing to limited narrative information, and since stated suicidal intent was often absent, inferences were necessary. Considered in the inference of lethal intent were: method used, clinical circumstances, ‘past history of suicide attempt, past history of self-injurious behavior/self-mutilation, and family history of suicide/suicide attempts’.

This departed from criteria required in standardized suicide assessment scales for children, such as the K-SADS, and adults, for example, Beck Suicide Intent Scale. In such scales, explicit lethal intent and objective plan are used to assess suicidal behavior (‘suicidality’). Other considerations include lethality of method, precautions against interference, and failure to notify others of the plan. In contrast, the utility of historical factors—that is, ‘past history of suicide attempt, past history of self-injurious behavior/self mutilation, and family history of suicide/suicide attempts’—used by the Columbia group to determine if a current act shows ‘suicidal intent’, is dubious.

The pertinent Columbia codes for the clinical trial events used by the FDA are in Table 1.

Dr Posner stated,

Suicide attempt, of course, which is defined as a self-injurious behavior associated with some intent to die. Intent can be stated or inferred by the rater. It is important to know that no injury is needed. (HHS, 2004, p 121)

**Table 1** Categorization of Suicidality

FDA used adverse event columbia codes	N	FDA Columbia code composites plus item analyses	N
(1) Suicide attempt	27	(1) Definitive suicidal behavior (1+2)	33
(2) Preparatory actions toward imminent suicidal behavior	6	(2) Suicidal ideation without behavior (6)	45
(3) Self-injurious behavior, intent unknown	24	(3) Primary outcome: definitive suicidal behavior/ ideation (1+2+6)	78
(4) Self-injurious behavior, no intent—primarily to affect circumstances	2	(4) Possible suicidal behavior/ ideation (1+2+3+6+10)	109
(5) Self-injurious behavior, no intent—primarily to affect internal state	5	(5) Self-injurious Behavior—nonsuicidal (4+5+11)	11
Analyses		Scale item	
(6) Suicidal ideation	45	(6) Worsening of suicidality—item scores	434
(10) Not enough information	7	(7) Emergence of suicidality—item scores	349
(11) Self-injurious behavior, no suicidal intent	4	No events	4436

Following are abbreviated narrative examples of ‘Suicidal Attempt’, cited by Dr Posner (HHS, 2004). These indicate how difficult inferring ‘intent to die’ is when no specific inquiry has been made, or clinical description is taken to overrule stated intent.

- The subject attempted suicide by immolation. Her siblings doused the flames immediately. She was left with minor burns on her abdomen and one on her left shoulder that were treated. The subject admitted that she was angry with her parents for going away and leaving her alone at home, because she was fearful. The subject admitted that she had acted impulsively and had not intended to kill herself.
- ...a 16-year-old who claimed to have ingested 100 tablets of study med, after a fight with her mother. The patient informed her mother. The mother brought the patient to the ER. The patient reported feeling shaky. Emergency room physician said she ‘was slightly tachycardic’, with a pulse of 100. The tox screen was negative, but the patient did have some illness and she stayed in the ER until she was asymptomatic, and then later admitted to a psych unit.
- After a conflict with her father, the patient, age 17, took an overdose of 20 (several) tablets. In her father’s opinion, the overdose was five tablets. The patient did not have any symptoms of an overdose, ‘not even nausea’.
- The patient, age 15, impulsively slit her wrists following an altercation with her mother. The wounds were superficial and were not stitched.

The examples cited on the slides presented at the hearing are segments of fuller reports. However, validly distinguishing impulsive gestures, angry retaliation, manipulative statements, and true intent to die remain clearly problematic. Further, ‘ideation’, without any validating behavior, does not provide a reasonable basis for inferring lethal intent. Nonetheless, the reliabilities reported are impressive. Intraclass correlations were Suicide Attempt 0.81, Preparatory Actions 0.89, and Suicidal Ideation 0.97. Calculation details and instructional methods were not presented. One possible explanation for high reliability is that the threshold for judging intent to die was set at a minimum. If only plainly accidental adverse events were excluded from being considered suicidal, high reliability (but low validity) would result. It is easier to measure, than to know exactly what is being measured.

## STRATEGIES USED IN DATA ANALYSES

### Single Trials and End points

Dr Hammad, the FDA statistical analyst (Hammad, 2004), repeatedly noted that treating different trials as one large trial may fail to preserve randomization, thus introducing bias and confounding. Therefore, individual trials were the units for the primary analytical approach. However, this approach largely failed to identify drug–placebo differences. Almost all later conclusions required pooled trials for statistical significance ( $p=0.05$ ). The criteria for trial poolability were not specified.

Preliminary analyses explored possible important covariates, which were regularly found unimportant. For instance,

in the MDD trials, a history of suicide attempt did not suggest increased suicidality risk. This surprising negative finding may be due to low power within trial analyses, or indicate the lack of construct validity of these surrogate outcome criteria. The ‘suicidality’ rate was higher in some placebo groups than in some drug groups, indicating marked sample heterogeneity across trials.

### Aggregated Trials, Composite End points

The statistical analysis took the Columbia categorization as a framework, but grouped the categories into dubious composite variables. The Primary Outcome composite measure is said to be postulated *a priori*, but the reasoning is never specified.

If the pooled analyses were restricted to the composite’s components, no findings approach statistical significance.

As Freemantle and Calvert emphasized, analyses of all the individual components of a composite should be tabulated (Freemantle and Calvert, 2004). Montori *et al* further question, ‘Are the component end points of similar importance to patients?’ (Montori *et al*, 2005). These issues are not addressed. It seems likely that the composite end point was chosen for FDA analysis because adverse events were so infrequent and diverse, that only by agglomerating events could an analyzable variable be constructed. If ‘intent to die’ could be validly estimated, a drug-induced increment in suicidal attempts would justify a warning. However, amalgamating attempts (that often caused no injury) with verbalizations subverts clinical meaningfulness.

Montori *et al* state, ‘When large variations exist between components, the composite end points should be abandoned’ (Montori *et al*, 2005). Regarding SSRI effects on MDD, the RR for attempts plus preparatory actions is 1.76; but for ideation it is 1.0, which is not even a signal. An editorial concerning cardiac outcomes questions using composites where one component is nonsignificant (Freemantle and Calvert 2004).

What composite credibility exists when no component is significant? In short, the FDA primary outcome composite variable is an inadequate, misleading surrogate—not only for completed suicide but also for supposed attempts.

## FDA GROUPINGS (SEE TABLE FOR SUMMARY)

### OUTCOME 1: ‘Definitive Suicidal Behavior’

Grouped Columbia Codes 1 and 2 as a composite:

Columbia Code 1 ( $N=27$ ): ‘Suicide attempt’

Defined as ‘Self-injurious behavior associated with some intent to die. Intent can be stated or inferred by a rater. No injury needed’. Columbia Code 2 ( $N=6$ ): ‘Preparatory acts toward imminent suicidal behavior’ Defined as ‘Person takes steps to injure self but is stopped by self or other. The intent to die is either stated or inferred’.

OUTCOME 2: ‘Suicidal Ideation without Behavior’ ( $N=45$ )

Columbia Code 6.

OUTCOME 3: Primary Outcome—‘Definitive Suicidal Behavior/Ideation’ ( $N=78$ )

Outcome 3 groups Columbia Codes 1, 2, and 6. Therefore, ‘Ideation’ is 57.7% (45/78) of the ‘Primary Outcome’, central

to the statistical analyses. When 'suicide attempts' (Columbia Code 1) and 'preparatory acts' (Columbia Code 2) were aggregated across studies, no difference was found between drug and placebo groups. However, when 'Suicidal Ideation' (Columbia Code 3) was added to Codes 1 and 2, this composite was important to the FDA.

**OUTCOME 4: 'Possible Suicidal Behavior/ Ideation' (N = 109)**

Combined Columbia Codes 1, 2, 3, 6, and 10, with 3 being 'Self-injurious behavior, unknown intent' and 10 being 'Not enough information, unable to classify'.

Outcome 4 is clearly heterogeneous. That it is statistically significant when pooled across trials, and numerically superior to the primary outcome, questions both surrogate adequacy and construct validity. There may be (pooled) evidence for a drug effect of some sort, but the label 'suicidality' is inappropriately specific and misleading.

**OUTCOME 5: 'Self-Injurious Behavior, Non-Suicidal' (N = 11)**

This grouped Columbia Codes 4, 5, and 11, considered by the Columbia group as showing no evidence, whatsoever, of suicidal intent (11 events). The RR comparing all SSRIs to placebo, for all indications, is 1.61—numerically similar to the nonsignificant RRs considered to support 'suicidality'.

That 'Self-Injurious Behavior, without Possible Suicidal Intent' shows this parallel pattern questions construing medication as specifically engendering 'suicidality'.

Other outcomes were derived from interview rating scales (HAM-D, CDRS-R, MADRS, and K-SADS) rather than Adverse Event Reports. These scales incorporated suicide relevant items. The number of putative 'suicidality' events far exceeds those derived from adverse reaction reports.

**OUTCOME 6: 'Worsening of Suicidality Score' (N = 434)**

Consists of a worsening (variously defined as one or two points) of items referring to suicide, from baseline values, regardless of subsequent change.

**OUTCOME 7: 'Emergent Suicidality' (N = 349)**

Refers to the subset of patients who, at baseline, had no suicide-relevant items; but then met criteria for worsening.

Outcomes 6 and 7 were both nonsignificant, with RRs less than 1. Symmetrical language would consider this an antisignal. Further, these contrasts are clearly the most powerful of all contrasts performed.

Dr Hammad states,

Those suicidality items were collected regularly at study visits ...A caveat is that the information gathered by the suicidality items might not have been collected at the time the suicidal behavior or ideation was manifest which might explain, to some extent, the lack of signal strength based on these outcomes. (Hammad, 2004) However, similar caveats could be raised about Adverse Event Reports, which were often not contemporaneous with the described events. Further, the item endorsements require focused inquiry.

## RISK BY DRUG

Table 10 of Dr Hammad's analysis summarizes overall risk estimates for each drug, for primary Outcome 3 (Behavior/Ideation), across all diagnoses, as well as across MDD trials

(Hammad, 2004). Except for fluoxetine, all analyses had RR estimates exceeding 1, considered by the FDA as a 'signal'. But, of these 12 analyses, only the two for venlafaxine were statistically significant, with lower confidence limits exceeding 1.

These two trials did not exclude patients with treatment resistance, history of suicide attempt, or homicide risk.

Dr Hammad clearly states the limitations of this underpowered *post hoc* analysis that is complicated by lack of statistical significance for many subanalyses. He advises caution, since pooling data across drugs assume a class effect. Further, the unanalyzable relationships to dosing, discontinuation, compliance, and hostility immediately prior to events labeled 'suicidal', may confound specific drug effects.

Dr Hammad's conclusions, relevant to this review, are:

- (1) No completed suicides.
- (2) No individual trial showed a statistically significant signal for 'suicidality', although risk ratios of 2 or more were found.
- (3) Table 13 of the Hammad report summarizes the overall risk estimates for treatment emergent agitation or hostility, by drug and MDD trials. All SSRI risk ratios are above 1. For paroxetine, this is significant.

Dr Hammad states,

Unfortunately, examining the likelihood of having an event of the primary outcome (Outcome 3) among patients with the symptoms of hostility or agitation was not evaluable because information of the timing of the latter events was not available in the data. Therefore, determining the time sequence was not possible. (Hammad, 2004) Nevertheless, the adverse event examples highlight this possibility. The Columbia group could have evaluated this sequence, to attempt to distinguish angry expressions, impulsive acts, and manipulations from the necessary 'intent to die'.

## FURTHER ANALYSES

Absolute risk differences between placebo and drug groups are not simply presented. Since Outcome 1 ('Actual or Near Attempts') seems closest to 'intent to die' issues, relevant sections of Table 16.2 (Hammad report) are summarized.

If antidepressants incite suicide attempts, this should be clearest in the nine studies of nondepressives, since suicide attempts are not characteristic of these illnesses. In six of these nine trials, there were zero Outcome 1's in both drug and placebo groups.

For SSRI trials, in nondepressives the Outcome 1 rate is 0.3% for SSRIs and 0.2% for placebo. Six of these eight trials had zero Outcome 1's for both drug and placebo. This finding does not support an SSRI toxicity that endangers all children. For the 10 MDD trials of SSRIs, the Outcome 1 percentage for drug is 1.9%, for placebo 1.3%—an absolute risk increment of 0.7%, equivalent to an RR of 1.5.

However, in four of the 10 SSRI trials, the absolute risk increments were zero or negative. The FDA analysis showed no significant effects for Outcome 1, but emphasized the RR 'signal', shown here to be inconstant.

## POSSIBLE ALTERNATIVE COMMITTEE CONCLUSIONS

What do the data provided to the Advisory Committee suggest? I conclude there is quite weak, possibly severely confounded, evidence from pooled trials that medication may cause behaviors that can arouse clinical concern. That these behaviors manifest an intent to die is dubious. That they are predictive of completed suicide is quite unlikely, and, at best, false positives would swamp any predictive efforts.

It was only by the amalgamation of diverse trials and questionable variables, of unlikely clinical or predictive significance, that statistical sanctification was approached.

There is no convincing evidence that antidepressants specifically increase suicide attempts. The committee's decision was probably influenced by the paucity of data (except for fluoxetine) demonstrating a specific benefit. In terms of risk/benefit ratio, the lack of specific benefit became misleadingly translated into a high specific risk.

However, the data were completely inadequate to determine if antidepressants even increase the risk that suicide attempts might occur, since the composite 'Primary Outcome' pooled findings did not warrant such a judgment.

A firm Committee conclusion could have been that concerns about lethality could not be adequately addressed by the data at hand. The simple fact is that firm conclusions about rare events cannot come from feasible randomized, double-blind, placebo-controlled trials—even if the relevant data were properly collected.

Heated public controversies about post-marketing drug toxicity hinge upon controversial rare or late events, for example, antidepressants in children, estrogen in postmenopausal women, antipsychotics for the elderly, long-acting stimulants for children, blindness with sildenafil, etc. Accusations of FDA corruption, bureaucracy, and pharmaceutical industry domination deflect the public's attention from this key issue.

Attaining adequate power in clinical trials to validly evaluate rare events is simply not realistic. Longer and larger pre-marketing trials decrease the period of profitable marketing exclusivity, and incur major practical problems. Patients drop out, making studies progressively more difficult to evaluate. Larger studies require multi-site protocols, with attendant severe administrative and clinical problems. Longer studies delay the public's access to useful treatments.

The public health value of spontaneously reported toxicity is speculative, since it is not known whether the suspected event is really attributable to the medication, or might occur in any case in such sick people. The FDA's system of monitoring and analyzing the infrequent, voluntarily reported, adverse events cannot succeed.

Typically, neither pre-marketing trials nor voluntary post-marketing reports demonstrate increased serious risks. However, large, independent, post-marketing trials, designed to study other end points, have unexpectedly revealed problematic toxicities. (Such blows to industry substantially lessen the likelihood of their voluntarily initiating post-marketing trials. Federal grant support for controlled trials of marketed medications has always been infinitesimal.)

Recently, medications are often released to the market on condition that the pharmaceutical firm conducts post-marketing surveys for potential side effects.

However, for the most part, these have either not been done or been ineffective. Current law (PDUFA) allows companies to speed premarketing drug evaluation by a petitioner's tax. However, Congress has forbidden using this money for post-marketing surveillance.

Horrifying, public outcries (at times well orchestrated) cannot be effectively met by protesting the lack of meaningful data. That leads to hasty worst-case restrictions that only resemble effective protections. Blindly shooting from the hip may play politically but is unlikely to improve matters. Whether black boxes, increased monitoring (largely of false positives), or drug withdrawals improve the public health remains unknown.

What should have dominated these proceedings is the overriding need for a system of post-marketing surveillance that really informs, as well as monitors, public policy decisions regarding safety and efficacy. The public is entitled to something much better than a buffed up *status quo*.

## STRATEGIES FOR COMPILATION OF VALID DATA REGARDING POST-MARKETING EVENTS

While multiple replications of randomized, controlled, clinical trials are the most secure basis for asserting specific benefits, their lack of feasibility for detecting rare events requires consideration of other methods. Attempts to derive causal risk estimates from naturalistic post-marketing epidemiological data incite heated debates. In my view, the creation of effective post-marketing surveillance must be brought to the forefront of public discussion, to deal with the confusing barrage of horror stories that spark regulations, warnings, and changes in medical practice, with unclear net effects.

Is systematic post-marketing surveillance possible? The current FDA system (MedWatch) is inadequate, as FDA officials often state. However, poor funding or staffing is not to blame, since even excellent staffing with ample funding could not do the job. Given FDA's structure, the data necessary for informed conclusions cannot be collected. Rational attempts to develop systematic post-marketing surveillance do exist. These may help guide such a program for the United States. Three post-marketing surveillance approaches for collecting data, permitting rare event estimations, are briefly described.

(1) *Computerized Prescribing*: Can data from clinical practice help improve clinical choices and make practice self-correcting? Schiff and Rucker (1998), in a visionary article on Computerized Prescribing, suggest that orders to stop medications should indicate which of the following occurred:

- (a) adverse reaction;
- (b) symptom/disease resolved;
- (c) failure to achieve desired therapeutic response;
- (d) more desirable alternative.

By combining computer-entered drug indication data—that is, symptoms and diagnoses (automatically recorded



when the prescription is written)—with adverse effect data, usage patterns, and reasons for discontinuing therapy, a post-marketing surveillance system would be created. Expanding the number of trained physician assistants, capable of medically informed data collection, would facilitate this process. Linking this information with functional status, mortality, hospital admissions, clinical laboratories, and new diagnostic data opens significant outcomes research opportunities, as well as creates the infrastructure for longitudinal patient records and a knowledge-generating database.

(2) *UK General Practice Research Database (GPRD)*: A GPRD was established in the UK, requiring GPs to enter demographics, age, sex, medical diagnosis, comments, prescriptions, events leading to withdrawal of a drug or treatment, referrals to hospitals, treatment outcomes (including hospital discharge reports), and patient referrals into a computer format (Wood and Coulson, 2001; GPRD hyperlink: <http://www.gprd.com>).

Miscellaneous patient care information, for example, smoking status, height, weight, immunizations, and laboratory results, were included.

However, GPs primarily used these computers to manage patients, so data recording was incomplete. Regular audit and feedback accompanied by rewards and sanctions might have improved this system, but that expense was foregone. Also, after taking the initial responsibility, UK health ministers decided that the database should be self-financing.

(3) *PHARMO*: A seemingly more robust model is the PHARMO Record Linkage System, established in the early 1990s in the Netherlands (hyperlink: <http://www.pharmo.nl>).

Computerized medical histories are linked to the following: the use and cost of prescription drugs; diagnostic/therapeutic data from hospitals; clinical laboratory and pathological findings; family practitioner records; drug histories; all hospital admissions, with detailed information concerning primary and secondary discharge diagnoses, procedures, consultations, dates of admission and discharge; histological, cytological and autopsy examinations. To preserve privacy, records are anonymized.

Current data are collected from about two million representative Netherlands residents (12% of the population). A body of local academic experts regularly reviews this massive data bank for novel findings.

For rational policy decisions, cross-linked data on safety, efficacy, and cost are needed. Such a system could extend its scope from safety to effectiveness, as indicated by a PHARMO report (Herings *et al*, 2002) regarding drugs for hypertension, inhalation steroids for asthma, antidepressants, and cholesterol lowering. They find that the medical as well as the economical consequences of premature discontinuation of chronic, drug treatment are enormous. About 50–70% of the patients use drugs for too short a period to be effective and, therefore, use drugs ineffectively. On average, 50–70% of the patients being treated with one of these drugs discontinued using them within a year from starting treatment... investment loss of pharmaceuticals, runs to several hundreds of millions of zeros yearly, excluding the treatment costs of basically avoidable morbidity.

Meijer *et al* (2004) demonstrated the specific utility of the PHARMO cross-linked computerized files, by finding ‘a significant association between degree of serotonin reuptake inhibition by antidepressants in risk of hospital admission or abnormal bleeding as the primary diagnosis’. These conclusions required studying the 1992–2000 records of a cohort of 64 000 new antidepressant users in the Netherlands.

Such a provocative study is impossible for the current FDA structure. Even if an FDA scientist noticed case reports of bleeding—and became suspicious of drug toxicity—they would still have to design, activate, and find funds for a complex study limited to only this concern. The emphasis on increasing FDA staff, monitoring their activities, creating a new Department of Drug Safety, etc, obscures the need for an entirely new prospective data collection and analysis system.

That improved systematic post-marketing surveillance is necessary is hardly a new conclusion. A recent JAMA editorial by Fontanarosa *et al* (2004) documents this in hair-raising detail. They argue that restoring public trust requires developing a prospective, comprehensive, and systematic approach for monitoring, collecting, analyzing, and reporting, data on adverse events. Above all, the agency must be completely independent of influence from the pharmaceutical industry, biotechnology firms, and medical device manufacturers.

They propose that, under pain of severe legal penalty, industry should conduct extensive post-marketing studies on all new products, with rapid communication of detected adverse events. This falls short of what is both desirable and practical, since the wheel has to be reinvented for each new product. Models such as PHARMO, for effective prospective programs, already exist. The FDA currently sponsors a study by the National Academy of Science’s Institute of Medicine of the drug safety system, emphasizing the post-marketing phase. However, one answer is already in, the current post-marketing system is useless for rare and late toxicities that incite rancorous uncertainty and public apprehension.

Such a fully independent, federally supported agency must be insulated from likely political and economic interference. This may require an unusual hybrid of government and industry support, with a dedicated nonprofit public foundation. In any case, this issue should be at the forefront of public discussion rather than languishing in obscurity.

## ACTION IMPLICATIONS

Would it be too expensive? In the context of the many billions lost by industry because of the sudden withdrawal of popular products, as well as the blight upon sales of related drugs, this investment is actually prudent. Industry would be wise to foster this development, since it is likely that such losses will catastrophically increase.

Such a system would require a radical revision of the private practice, disconnected, paper-based, medical information systems of the United States. However, such a program could be initiated in federally supported, supervised medical services. Extensive hospital, outpatient,

and family services are provided by the armed services (Army, Navy, and Air Force), Veterans Administration, and Public Health Service. Instituting such a well-monitored system should be mandatory. The VA has already taken useful steps toward medical record computerization.

Two other grave, mounting, problems—the rising cost of medical care and the need for easily accessed complex, longitudinal, medical records for individual patients—also support this necessary development. A series of public meetings, including the range of stake-holders, should debate effective post-marketing surveillance—allowing public education, initiating relevant feasibility and cost/benefit studies, and gaining necessary legislative attention to this issue.

Beautrais (2004) emphasizes that a selective, intensive, clinical focus should be on those who make serious suicide attempts—rather than expending valuable clinical resources on the false positives generated by ‘suicidality’. Such a computerized system would facilitate many such preventive programs.

## CONCLUSIONS

- (1) Meeting public concerns about drug safety has failed.
- (2) It is completely impractical to attempt to answer questions about rare and late harms on the basis of clinical trials, even if the data were properly collected.
- (3) Computerized, cross-linked, population-based medical records should be mandated in federally supported medical facilities. Analyses should be carried out by an agency of independent experts, buffered from political and economic pressures.

## ACKNOWLEDGEMENTS

Judith Godwin Rabkin, PhD, provided a valuable critique of the manuscript. I have had many financially supported contacts with the pharmaceutical industry over 45 years, recently including Eli Lilly and Company, Forest Pharmaceuticals Inc., GlaxoSmithKline, Interferon Sciences, Inc., Pfizer Inc., Roche Pharmaceuticals, SmithKlineBeecham, The Upjohn Company, Vela Pharmaceuticals Inc., and VivoMetrics, Inc.

## REFERENCES

- Baker SG, Kromer BS (2003). A perfect correlate does not a surrogate make. *BMC Med Res Methodol* 3: 16.
- Beautrais AL (2004). Further suicidal behavior among medically serious suicide attempters. *Suicide Life-Threaten Behav* 34: 1–11.
- Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H (2000). The validation of surrogate endpoints in meta-analysis of randomized experiments. *Biostatistics* 1: 49–67.
- DeGruttola V, Fleming T, Lin DY, Coombs R (1997). Perspective: validating surrogate markers—are we being naïve? *J Infect Dis* 175: 237–246.
- Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (2004). *Joint Meeting of the CDER Psychopharmacologic Drugs*

- Advisory Committee and the FDA Pediatric Advisory Committee*, Bethesda, MD, September 14, 2004.
- Fleming TR (2005). Surrogate endpoints and FDA’s accelerated approval process—the challenges are greater than they seem. *Health Affairs* 24: 67–78.
- Fleming TR, Demets DL (1996). Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 125: 605–613.
- Fombonne E, Wostear G, Cooper V, Harrington R, Rutter M (2001). The Maudsley long-term follow-up of child and adolescent depression. *Br J Psychiatry* 179: ((1) Psychiatric outcomes in adulthood, pp 210–217; (2) Suicidality, criminality and social dysfunction in adulthood, pp 218–223).
- Fontanarosa PB, Rennie D, De Angelis KD (2004). Editorial: post-marketing surveillance—lack of vigilance, lack of trust. *J Am Med Assoc* 292: 2647–2650.
- Freedman LS, Graubard BL, Schatzkin A (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statist Med* 11: 167–178.
- Freemantle N, Calvert M (2004). Editorial. *Eur J Heart Failure* 6: 531–533.
- Gould MS, Greenberg T, Velting DM, Shaffer D (2003). Youth suicide risk and preventive interventions: a review of the past ten years. *J Am Acad Child Adolesc Psychiatry* 42: 386–405.
- Hammad TA (2004). Relationship between psychotropic drugs and pediatric suicidality. *Report to U.S. Food & Drug Administration*, August 16.
- Herings RMC, Leufkens HGM, Heerdink ER, Klungel OH, Breejkveldt-Postma NP (2002). Chronic pharmacotherapy forever—a study into the extent, determinants and possible economic consequences of non-persistence with chronic pharmacotherapy. *Report for PHARMO Institute for Drug Outcomes Research*, The Hague, Netherlands, February 2002.
- Lin DY, Fleming TR, DeGruttola V (1997). Estimating the proportion of treatment effect explained by a surrogate marker. *Statist Medicine* 16: 1515–1527.
- Meehl PE, Rosen A (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol Bull* 52: 194–216.
- Meijer WE, Heerdink ER, Nolen WA, Herings RM, Leufkens HG, Egberts AC (2004). Association of risk of abnormal bleeding with degree of serotonin reuptake inhibition by antidepressants. *Arch Intern Med* 164: 2367–2370.
- Montori VM, Permanyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D *et al* (2005). Validity of composite endpoints in clinical trials. *Br Med J* 330: 594–596.
- Prentice RL (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statist Med* 8: 431–440.
- Schiff G, Rucker D (1998). Computerized prescribing; building the electronic infrastructure for better medication usage. *J Am Med Assoc* 279: 1024–1029.
- US Food & Drug Administration (2005). *Templates for Anti-depressant Black Box Warning and Medication Guide* Posted on FDA website, February 3, 2005.
- Vedantam S (2005a). Fewer kids prescribed drugs for depression—sharp decrease seen after reports of risks. *The Washington Post* February 2, 2005, p A08.
- Vedantam S (2005b). Drugs raise risk of suicide; analysis of data adds to concerns on antidepressants. *The Washington Post* February 18, 2005, p A01.
- Wood L, Coulson R (2001). Original report: Revitalizing the general practice research database; plans, challenges, and opportunities. *Pharmacoepidemiol Drug Safe* 10: 379–383 (General Practice Research Database, Medicines Control Agency, London, UK).

## APPENDIX A. FDA TEMPLATE FOR BLACK BOX

### Suicidality in Children and Adolescents

Antidepressants increased the risk of suicidal thinking and behavior (suicidality) in short-term studies in children and adolescents with MDD and other psychiatric disorders. Anyone considering the use of (Insert established name) or any other antidepressant in a child or adolescent must balance this risk with the clinical need. Patients who are started on therapy should be observed closely for clinical worsening, suicidality, or unusual changes in behavior. Families and caregivers should be advised of the need for close observation and communication with the prescriber. (Insert established name) is not approved for use in pediatric patients. (See Warnings and Precautions: Pediatric Use.) (This sentence would be revised to reflect if a drug were approved for a pediatric indication(s). Such as, (Insert

established name) is not approved for use in pediatric patients except for patients with (Insert approved pediatric indication(s)). (See Warnings and Precautions: Pediatric Use.)

Pooled analyses of short-term (4 to 16 weeks) placebo-controlled trials of nine antidepressant drugs (SSRIs and others) in children and adolescents with MDD, OCD, or other psychiatric disorders (a total of 24 trials involving over 4400 patients) have revealed a greater risk of adverse events representing suicidal thinking or behavior (suicidality) during the first few months of treatment in those receiving antidepressants. The average risk of such events in patients receiving antidepressants was 4%, twice the placebo risk of 2%. No suicides occurred in these trials (FDA website, posted February 3, 2005 (US Food & Drug Administration, 2005)).