Theoretical Review

EVIDENCE THAT THE SSRI DOSE RESPONSE IN TREATING MAJOR DEPRESSION SHOULD BE REASSESSED: A META-ANALYSIS

C. Bruce Baker, M.D., J.D.,[1,*] Richard Tweedie, Ph.D., D.Sc.,[2] Sue Duval, B.Sc., Ph.D.,[3] Scott W. Woods, M.D.[1]

The limitations in design and analysis of currently available dose-response studies of SSRI treatment of major depression have led to the conclusion that dose response is flat. We applied concepts from our companion article to determine if currently available data is consistent with a "potential" and an "expressed" dose response. Using these concepts, we performed a meta-analysis on all identifiable published fixed-dose and dose-escalation studies that reported the effect of different SSRI oral doses on efficacy. "Potential" dose response in fixed-dose studies with categorical response outcomes equaled a significant meta-analyzed slope of 3.1%/100 SSRI mg equivalents (SMEs) (SE=1.2%) or 7.8% across the dose range. Similar analysis in dose-escalation studies that reported categorical response data yielded a non-significant meta-analyzed slope of 3.7%/100 SMEs (SE=2.3%) or 9.3% across the dose range. Analyses of the "expressed" dose response demonstrated in the studies indicated a slope statistically equal to zero. The current analysis suggests a "potential" dose response can be demonstrated for SSRIs in treating major depression. The analysis suggests an "expressed" dose response could exist in best clinical practice. Study designs better tailored to address the relevant clinical question would test these hypotheses more appropriately than previous studies. Depression and Anxiety 17:1–9, 2003. *© 2003 Wiley-Liss, Inc.*

Key words: *depression; dose response; SSRI, meta-analysis*

INTRODUCTION

Whether it is reasonable to increase SSRI doses in treating major depression after an initial inadequate response is controversial and an issue of wide practical importance. SSRIs are the initial antidepressant for the majority of patients treated for major depression. Initial treatment will fail or yield an inadequate response in approximately 50% of cases [AHCPR, 1993]. Both in terms of patient suffering and cost-effectiveness, it is important to know what strategies are reasonable after an inadequate initial response. One popular option, and one endorsed as reasonable by leading reviews of treating refractory depression, is an increase in the SSRI dose [Post, 1995; Thase and Rush, 1995]. Despite the strategy's popularity, it has been widely stated that the available research data has not supported it [Amin et al., 1989; Benkert et al., 1997; Dornseif et al., 1989; Dunner and Dunbar, 1992; Fabre and Putman, 1987; Fabre et al., 1995; Fava et al., 1994; Preskorn and Lane, 1995; Schweizer et al., 1990; Thase and Rush 1995; Wernicke et al., 1987, 1988].

Using the arguments in our companion article [Baker and Woods, 2003] about improved study design and analysis, we reevaluate "potential" and "expressed" antidepressant dose responses of SSRIs with a meta-analysis of the available fixed-dose and dose-escalation studies.

[1] Yale University School of Medicine, Department of Psychiatry, Connecticut Mental Health Center, New Haven, Connecticut
[2] Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota
[3] Division of Epidemology, School of Public Health, University of Minnesota, Minneapolis, Minnesota

*Correspondence to: Dr. C. Bruce Baker, Yale University School of Medicine, Department of Psychiatry, Connecticut Mental Health Center, 34 Park Street, Room 38, New Haven, CT 06519. E-mail: bruce.baker@yale.edu

Received for publication 15 March 2002; Accepted 16 June 2002

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/da.10079

MATERIALS AND METHODS

CONCEPTS

We applied the following concepts regarding improved analysis of currently available data [see companion article, Baker and Woods, 2003]. First, both fixed-dose and dose-escalation study types are relevant to the dose-response question. Second, in theory one can use intent to treat (ITT) or evaluable patient (EP) study samples to evaluate the "expressed" dose response of patients treated according to "best clinical practice" who fail an initial SSRI dose, if the studies meet specific design criteria, e.g., preferably dose-escalation studies with dose titration and sufficiently long phases one and two. To date, no published studies meet these design criteria. Therefore, with available fixed-dose and dose-escalation data, one may find a "dose-tolerant" (DT) sample is a better approximation of the appropriate sample to use in performing "expressed" dose-response analysis relevant to the clinical context of raising SSRI doses. However, use of a DT sample only partially compensates for the design limitations of the studies used. A DT sample cannot compensate for a short phase one, which results in limited purification of a sample tolerant of but unresponsive to an initial SSRI dose, or for a short phase two, which results in limited separation of responders/nonresponders to higher doses.

Third, to answer the "potential" dose-response question, by definition, analysis of either study type should use a DT sample as the sample of first choice, not merely an approximation. However, the short phase one and two of currently available studies impose limitations. With available data analysis, using a DT sample will underestimate the potential dose response.

SELECTION OF STUDIES

We sought to identify all published studies that reported the effect of dose on efficacy of the four SSRIs with a Food and Drug Administration indication for major depression (citalopram, fluoxetine, paroxetine, and sertraline). We identified studies using the Cochrane Library, Medline, and Health Star databases supplemented by manual searches based on the references associated with studies located through the databases. A larger number of studies deal with issues such as adverse advents [Altamura et al., 1988; Beasley et al., 1992], blood levels [Beasley et al., 1990], or treatment of depression subtypes [Tignol et al., 1992]; however, we identified ten studies that present original data and address SSRI dose vs. response for the general diagnosis of major depression. We dropped one study because it did not report numerical outcome data [Montgomery et al., 1992]. We excluded one study because its design was markedly dissimilar from all the remaining studies [Hebenstreit et al., 1989]. Of the remaining eight studies, four were fixed-dose studies [Dunner and Dunbar, 1992; Fabre et al., 1995; Wernicke et al., 1987, 1988] and four were dose-escalation studies [Benkert et al., 1997; Dornseif et al., 1989; Fava et al., 1994; Schweizer et al., 1990].

DESCRIPTION OF STUDY DESIGNS

Designs of the fixed-dose studies were all relatively similar (Tables 1 and 2). All studies were double-blinded, placebo-controlled, parallel-group, randomized trials. Patients were started immediately on higher doses without gradual escalation. The *n*s varied from 30 to 460. Length varied from 3 to 7 weeks.

Designs of the dose-escalation studies were somewhat more variable (Tables 3 and 4). All studies divided treatment into two phases: in phase one, patients received an initial SSRI dose, and in phase two, phase one non-responders were randomized either to continue the original dose (to control for additional time on the original dose) or receive a higher dose. Treatment assignment in phase two was blinded in all the studies. In the four studies phase one varied from 3–8 weeks and phase two from 3–5 weeks. N's in phase two ranged from 41–371. Only one study [Fava et al.,

TABLE 1. Potential dose response of current fixed dose SSRI studies and approximation of expressed dose response in best clinical practice; dose tolerant samples—unselected depressed patients: response rates[†]

Study[a]	Medication	Lower dose (mg)	Med dose (mg)	Higher dose (mg)	Lower dose, % (ratio)	Med dose % (ratio)	Higher dose, % (ratio)	Slopes, % (*SE*)
Fabre [1995]	Sertraline	50	100	200	56.5 (48/85)	61.0 (47/77)	56.9 (33/58)	3.4 (3.8)
Wernicke [1988]	Fluoxetine	5	20	40	48.9 (43/88)	48.9 (45/92)	53.7 (44/82)	3.6 (1.3)
Wernicke [1987]	Fluoxetine	NA	20	40–60	NA	41.3 (38/92)	46.7 (43/92) for 40 mg 39.2 (31/79) for 60 mg	1.3 (4.3)
Meta-analytic synthesis								3.1 (1.2)[b]

[†]Response rated based on decrease of HAMD > 50% among patients in the reported sample who did not drop out due to adverse events.
[a]Reported samples based on various evaluable patient definitions (details available on request).
[b]Significant positive slope by weighted linear regression. P=0.0075 (two-sided Z=2.68); slope=3.1%/100 SMEs (SE=1.2%) or 7.8% over dose range reported in studies.

TABLE 2. Expressed dose response of current fixed dose SSRI studies (incorporating study limitations); evaluable or ITT patient samples—unselected depressed patients: group mean HAMD changes

					Mean (SD)/change HAMDN			
Study*	Medication	Lower dose (mg)	Med dose (mg)	Higher dose (mg)	Lower dose (mg)	Med dose	Higher dose	Slope (*SE*)
Fabre [1995]	Sertraline	50	100	200	10.6 (8.6) 90	9.8 (8.9) 89	9.2 (8.2) 82	−0.9 (0.2)
Dunner [1992]	Proxoxetine	20	30	40	12.5 (NA) 95	11.5 (NA) 85	11.5 (NA) 87	−0.6 (0.5)
Wernicke [1988]	Fluoxetine	5	20	40	11.2 (8.3) 94	10.0 (8.4) 91	11.2 (8.2) 92	−0.2 (0.9)
Wernicke [1987]	Fluoxetine	NA	20	40–60	NA	9.8 (8.5) 97	9.6 (8.2) 97 for 40 mg 7.2 (9.1) 103 for 60 mg	−1.3 (0.7)
Totals[a]					11.5	10.3	9.8	−0.8 (0.2)[b]

*Reported samples based on various evaluable patient definitions (details available on request).
[a]Means weighted by *ns*.
[b]Significant negative slope by weighted linear regression. P=0.00002 (two-sided Z=−4.10); slope=−0.8 HAMD/100SMEs (*SE*=0.2) or −0.2 HAMD units over dose range of reported studies.

TABLE 3. Potential dose response of current dose escalation SSRI studies and approximation of expressed dose response in best clinical practice; dose tolerant samples—depressed patients: selected for initial inadequate response: response rates†

					Response rate, % (ratio)		
Study*	Length (wk)	Medication	Initial dose (mg)	Raised dose (mg)	Initial dose	Raised dose	Slope, % (*SE*)
Benkert [1997]	3/3	Paroxetine	20	40	61.1 (11/18)	75.0 (24/32)	9.3 (9.1)
Fava [1994]	8/4	Fluoxetine	20	40–60	26.9 (7/26)	53.3 (8/15)	17.6 (10.3)
Schweizer [1990]	3/5	Fluoxetine	20	60	53.8 (21/39)	60.0 (18/30)	3.1 (6.1)
Dornsief [1989]	3/<5	Fluoxetine	20	60	34.4 (62/180)	39.4 (63/160)	2.5 (2.6)
Meta-analytic Synthesis							3.7 (2.3)[a]

†Response rates based on decrease of HAND > 50% among patients in the reported sample who did not drop out due to adverse events.
*Reported samples based on various intent to treat or evaluable patient definitions (details available on request).
[a]Non significant positive slope by weighted linear regression. P=0.102 (two-sided Z=1.64); slope=3.7%/100 SMEs (*SE*=2.3%) or 9.3% across the dose range of the studies.

1994] gradually escalated patients to the higher dose. This same study differed in its phase two design [Fava et al., 1994]. In this study, patients were randomized to one of three arms: 1) increase the SSRI, 2) maintain the lower dose and add lithium, or 3) maintain the lower dose and add desipramine. We included this last study in our analysis because despite the additional comparisons inherent in its design, it still can provide a comparison between maintaining the original SSRI dose vs. increasing the dose. Most importantly for the present analysis, if lithium or desipramine were effective, they would tend to lessen the difference between groups whose SSRI dose was escalated, and groups whose initial SSRI dose was continued, they would not increase the difference. Thus, if inclusion of this study biases the current analysis, the bias is against finding a dose-response effect. Additionally, it is unclear whether the augmentation strategies employed in this study would have had a chance of altering the response rates because the lithium dose in this study was low and evidence on effectiveness of adding desipramine is unclear.

DESCRIPTION OF STUDY OUTCOME TYPES AND SAMPLE SIZES

Categorical response rates are reported in three of four fixed-dose (Table 1) and all four dose-escalation studies (Table 3). In all studies, response was defined as a >50% decrease in HAM-D. The three fixed-dose studies had a total n of 745; the four dose-escalation studies had a total n of 500 (DT samples).

Changes in group mean HAM-Ds were reported in all four fixed-dose studies (Table 2) and all four dose-escalation studies (Table 4). The four fixed-dose studies had an n of 1,102; the four dose-escalation studies had an n of 573 (EP samples). (In the present article, EP data is for patients with one evaluation post baseline, or patients remaining in the study at least 1 week.)

STATISTICAL METHODS

Adjustments to address design limitations. In our primary analysis of potential dose response using categorical data from current fixed-dose and dose-escalation studies, we employed the DT sample,

TABLE 4. Expressed dose response of current dose escalation SSRI studies (incorporating study limitations); evaluable or ITT patient samples—depressed patients selected for initial inadequate response: group mean HAMD changes

Study*	Length (wk)	Medication	Initial dose (mg)	Raised dose (mg)	Initial dose mean (*SD*) / change HAMDN	Raised dose mean (*SD*) / change HAMDN	Slope (*SE*)
Benkert [1997]	3/3	Paroxetine	20	40	13.6 (6.6) 36	13.9 (6.5) 50	0.2 (1.0)
Fava [1994]	8/4	Fluoxetine	20	40–60	6.8 (5.1) 26	10.4 (4.9) 15	2.4 (1.2)
Schweizer [1990]	3/5	Fluoxetine	20	60	10.1 (NA) 41	11.6 (NA) 36	0.8 (0.7)
Dornsief [1989]	3/<5	Fluoxetine	20	60	6.8 (7.2) 189	8.2 (7.3) 180	0.7 (0.4)
Total[a]					8.1	9.8	0.8 (0.3)[b]

*Reported samples based on various intent to treat or evaluable patient definitions (details available on request).
[a]Means weighted by *n*s.
[b]Significant positive slope by weighted linear regression. P=0.0096 (two-sided Z=2.59); slope=0.8 HAMD units/100SMEs (SE=0.3) or 1.93 HAMD units over dose range reported studies.

subtracting the *n* dropping due to adverse events from the original *n*.

To differentiate the influence of using DT samples versus the general influence of simply increasing the n through meta-analysis, we also ran the categorical analyses using EP n's. For HAM-D change data in both fixed-dose and dose-escalation studies, only expressed dose-response analysis based on EP samples was possible. HAM-D data for dose-tolerant samples could not be identified through the published data.

General methodology and available data. Categorical response data for the fixed-dose and dose-escalation studies are presented as proportions (Tables 1 and 3). The n's were calculated both for an EP and a DT analysis. Data on changes in HAM-D scores were represented in the studies in terms of change in means of group HAM-D scores (Tables 2 and 4). The n's and *SD*s for the HAM-D scores were available and used in six of the eight studies. For the two studies with missing *SD*s, we used estimates described below.

To calculate an effect across studies using different SSRIs, we arranged dose levels of the different SSRIs along a common dose structure in terms SSRI mg equivalents (SMEs). We represented this common dose structure in a continuous form for use in a weighted linear regression of both the categorical and the continuous outcome data. After generating linear regressions for each of the individual studies within each type of study design, we carried out a meta-analysis within each study design of the slopes of these regressions using the method of Tweedie and Mengersen [1995]. This method in effect calculates a weighted average of the slope coefficients within each study type, and the slope generated from this meta-analysis gives an overall estimate of the rate of change in categorical or continuous outcome vs. dose within the fixed-dose or dose-escalation studies.

We used a random-effects model for the meta-analyses. Alpha=.05. All statistical tests were run with Minitab v. 12.1.

Equivalent dosage levels. For all the base case weighted linear regressions, we generated the continuous SMEs using the following anchor points: 50 SMEs=sertraline 50 mg=fluoxetine 5 mg or paroxetine 20 mg; 100 SMEs=sertraline 100 mg=fluoxetine 20 mg or paroxetine 30 mg; and 200 SMEs=sertraline 200 mg=fluoxetine 40 mg or paroxetine 40 mg. We chose this SME structure based on the lowest effective doses shown in any of the reported studies as the first anchor point.

We also carried out a sensitivity analysis of the SME structure because there is little evidence on appropriate dose equivalencies. We performed the weighted linear regression using five additional alternative sets of anchor points for the SMEs. Sensitivity analysis showed the originally chosen SMEs resulted in the greatest consistency among the slopes of the individual studies within each type of study as reflected in the lowest *SE* of the meta-analyzed slope for each study

type. The original and five alternative SME structures showed the same pattern of results. Consequently, we report specific results only for the base case SME structure. Exact slope values in the dose-response analyses depend on the SMEs chosen.

Estimated *SD*s for the weighted linear regression. For appropriate weighted regression, results must have weights based on *SD*s and *n*s. *N*s were available in all cases and binomial *SD*s were available for all categorical response data. For continuous HAM-D data, three *SD*s were missing or not fully reported and had to be estimated.

For Fava and Rosenbaum [1994] (Table 4), we averaged the DMI/Li *SD*s to give a *SD* of 5.1 based on n=26. For Schweizer [1990] (Table 4), we assumed the baseline and post-treatment *SD* for the HAM-D were the same and calculated the SD for the difference assuming independence. For Dunner [1992] (Table 2), we estimated a *SD* of 8.4 which is the mean of all the studies with reported post-baseline HAM-D *SD*s.

RESULTS

Analysis of the potential dose response (DT samples) for fixed-dose studies reporting categorical data are reported in Table 1. The meta-analyzed slope is 3.1%/100 SMEs (*SE*=1.2%) or 7.8% across the studies' dose range (significant dose response, *P*= 0.0075; two-sided *Z*=2.68).

Analysis of the expressed dose response using EP samples with the fixed-dose studies with group mean HAM-D changes are reported in Table 2. The meta-analyzed negative slope is −0.8 HAM-D units/100 SMEs (*SE*=0.2 HAM-D units) or −2.0 HAM-D units across the study's dose range (significant inverse dose response, *P*=0.00002; two-sided *Z*=−4.10). Apparent conflict between analysis of the categorical and continuous measures may be due to the unavailability of a DT sample for the continuous measures (see Discussion).

Analysis of the potential dose response (DT samples) for the dose-escalation studies with categorical response outcomes is reported in Table 3. The meta-analyzed slope is 3.7%/100 SMEs (*SE*=2.3%) or 9.3% across the studies' dose range (no significant dose response at .05 level, *P*=.10; two-sided *Z*=1.64).

Analysis of the expressed dose response that employed EP samples, with the dose-escalation studies with group mean HAM-D changes is reported in Table 4. The meta-analyzed slope is 0.8 HAM-D units/100 SMEs (*SE*=0.3 HAM-D units) or 1.9 HAM-D units across the dose range (significant dose response, *P*=.0096; two-sided *Z*=2.59).

The alternative analysis of the expressed dose response in each group of studies performed to check the effect of sample size versus the effect of analysis based on DT samples of both the fixed-dose and dose-escalation categorical data using EP rather than DT samples showed a non-significant negative slope (−3.8%/100 SMEs *SE*=3.7% or −9.5% across the reported dose range, *P*=.303; two-sided *Z*=−1.0323) for the fixed-dose studies and a non-significant positive slope (2.4%/100 SMEs *SE*=2.5% or 6% across the dose range, *P*=.3472; two-sided *Z*=.937) for the dose-escalation studies.

DISCUSSION

We suggest that the individual results and the overall pattern of results that we report are consistent with existence of a potential SSRI dose-response relation. The same data suggest an expressed dose response could exist in studies that more closely reflect best clinical practice as well as in best clinical practice itself. These conclusions are inconsistent with the conclusions drawn from each prior individual study used in our analysis and with prior general conclusions drawn from these studies as a group. We believe the difference in the results of our analysis and all previous analyses is explainable in terms of two principles described in our companion article. We suggest the role played by these principles is illustrated in the patterns of significance, lack of significance, and slope directions observable when one contrasts the different groups of data in our analysis. We consider the data in detail with the potential dose-response relation and then consider the magnitude of the factors that would influence existence of an expressed dose-response relation.

EXISTENCE OF A POTENTIAL SSRI DOSE-RESPONSE RELATION

The weighted linear regression meta-analysis of the fixed-dose studies with categorical data (Table 1) and the dose-escalation studies with change in HAM-D data (Table 4) directly support existence of an SSRI potential dose response.

We believe the categorical data for the dose-escalation studies are also consistent with these conclusions. The weighted linear regression meta-analysis of the dose-escalation studies with categorical data actually showed a larger slope for these studies than the group of fixed-dose studies. We suggest that the dose-escalation studies did not reach the significance criterion due to the larger standard errors associated with the dose-escalation studies given their smaller *n*, 500 vs.745.

Beyond the fact that the slopes were positive for the meta-analyzed data of the fixed-dose studies and dose-escalation studies with categorical response data and the dose-escalation studies with change in HAM-D data, other characteristics of the data are also consistent with a potential dose response. Slopes of the individual studies are similar within each study type, slope directions of all the individual studies are consistent within each study type, and the meta-analyzed slope

values for the fixed-dose studies and dose-escalation studies with categorical response data are in the same direction.

PRINCIPLE 1: CORRELATION OF ADVERSE EVENTS WITH DOSE AFFECTS

Appropriate dose-response analysis: this issue is magnified if the dose is not escalated gradually. Why are dose-response slopes positive for meta-analyzed data of fixed-dose and dose escalation studies' categorical data using DT samples vs. inconsistent for the same studies using EP samples? This pattern suggests two points. First, this pattern is consistent with the effect exerted by the relation of dose with adverse events as reflected in DT versus EP samples [see companion article, Baker and Woods, 2003]. Analysis based on DT samples demonstrates several consistent patterns that were described immediately above that support a potential dose response. However, analyses based on EP samples show inconsistent results. Within both study designs, when EP samples are used the slopes of the individual studies are mixed between positive and negative slopes; the magnitude of the slopes is very inconsistent as reflected in the *SE*s; and whereas neither of the following is statistically significant, the fixed-dose data as a group shows a negative slope, while the dose-escalation data shows a positive slope.

Second, the fact that EP meta-analyzed data is not significant whereas DT meta-analyzed data is significant demonstrates that the significance for DT meta-analyzed data versus the uniform lack of significance with each of the individual studies EP samples is not simply due to increased *n* from meta-analyzing the studies.

Why are results for fixed-dose categorical response data consistent with a positive dose-response relation, as opposed to results for fixed-dose HAM-D data that are consistent with a negative dose-response relation? This result is particularly remarkable because the number of studies is larger and the sample size is almost 70% larger for fixed-dose HAM-D data. However, this result is completely consistent with our first principal regarding association of dose and adverse events. Dropout rates due to adverse events increase across the dose range, and adverse events could be compensated for in categorical response data by constructing the DT sample but not in the continuous HAM-D data. Higher numbers of early dropouts due to adverse events with less time to improve would bias results from higher dose groups downward in an EP last observation analysis.

Why does dose-escalation categorical data show a larger change in outcome compared to dose-escalation HAM-D data? Again, patients that do not have a chance to improve due to adverse events could be removed from categorical data but not from HAM-D data.

Why does dose-escalation HAM-D data reveal a positive dose-response, whereas fixed-dose HAM-D data reveals a negative dose response? This result is consistent with our first principal and the fact that dose-escalation studies tended to have fewer dropouts due to adverse events as opposed to fixed-dose studies.

PRINCIPLE 2: IT IS IMPORTANT TO HAVE A LARGE ENOUGH, PURIFIED SAMPLE OF LOWER DOSE NON-RESPONDERS TO DETECT A DOSE RESPONSE

Why has no individual dose-escalation study detected a dose response with categorical or HAM-D data, and no individual fixed-dose study detected a dose response with categorical data? The answer appears to be sample size, power, and type of sample. In all four individual dose-escalation studies, response rates in higher dose groups were higher than in initial dose groups. Three of four dose-escalation studies had sample sizes of fewer than 40 patients per cell. The fourth study had a harmonic mean sample of 170 patients per cell. This largest study would have a power of only 0.45 to detect the roughly 10% difference in response rates at α=.05 detected in the meta-analysis. Power to detect the change in HAM-D would be much lower.

The exact power calculation is more complicated for the fixed-dose studies. The following approximation makes it clear that lack of power is equally a problem in the fixed-dose studies. Assuming the difference in response rates is roughly 9%, even a study with 120 patients per cell, the harmonic mean of the largest study, would have a power of only 0.35 at α=.05.

Why is the slope of dose-escalation HAM-D data significantly positive versus the significantly negative slope of fixed-dose HAM-D data? This pattern holds true despite the much larger *n* for the fixed-dose data. Dropouts due to adverse events could not be subtracted from either of the data sets; however, fewer adverse events tended to affect the dose-escalation data to interfere with an EP analysis. Also, the dose-escalation data has a more purified sample of lower dose non-responders, which helps to strengthen the available signal.

IS THERE AN EXPRESSED SSRI DOSE RESPONSE IN BEST CLINICAL PRACTICE? IF SO, IS IT CLINICALLY MEANINGFUL?

Does the potential dose-response analysis of the current studies provide a good approximation for the expressed dose response in best clinical practice, and in studies which reflect that practice? The potential dose-response analysis we report, based on the DT samples from the available studies, directly represents the potential dose response of these study

designs and not that of studies that more closely reflect best clinical practice or best clinical practice itself. There are two generally applicable countervailing factors pushing in opposite directions, which can cause particular potential dose responses to diverge from the expressed dose response one would expect in best clinical practice or studies that reflect such practice. First, the less a study purifies non-responders prior to study entry, the more the study will push the potential dose response generated from these studies toward understating the potential dose response from best clinical practice or studies designed to reflect this practice. Second, to the extent a particular potential dose-response estimate excludes irreducible dose-related side-effects, the estimate tends to overstate the expressed dose response in best clinical practice and studies that reflect such practice.

Accordingly, the likely existence and magnitude of an expressed dose response in best clinical practice can be analyzed in two sub-questions: 1) is the increased response to raising the dose likely to be greater in best clinical practice and better designed studies than the amount demonstrated in the studies used in the current meta-analysis, and 2) how large is the higher dose adverse events inherent dropout rate in best clinical practice and is it large enough to offset the increased response rate?

Is the difference in response rates for maintaining the initial dose vs. increasing the dose larger in appropriate clinical practice than the meta-analyzed rate of 7–10% across the dose range? There is reason to believe this is so. As noted previously, the most important design limitation of the available dose-escalation studies for which DT samples cannot compensate is that the available studies incorporate limited purification of the non-responder sample. In three [Benkert et al., 1997; Dornseif et al., 1989; Schweizer et al., 1990] of the four dose-escalation studies, the dose was increased after only 3 weeks. This would result in a limited separation of patients who would or would not respond to the initial dose within recommended trial periods of 6 weeks, thus diminishing the signal. Consistent with this view, the dose-escalation study [Fava et al., 1994] with the longest duration on the initial dose (8 weeks) had the lowest response to continuing the original dose in phase two and the largest difference in response between continuing the original dose vs. increasing the dose. The response rate of 53% in the increased dose group is notable, as the sample is composed purely of lower dose non-responders.

Clearly, limited weight should be placed on one small study; however, it most closely approaches the ideal study design we described in our companion article for testing expressed dose response. Consequently, these results may be closer to what would be observed in clinical practice that followed treatment guidelines, which in general recommend evaluation for 6 weeks with mild to moderate depression, 4 weeks in severe depression, before declaring failure and changing therapy [AHCPR, 1993].

Additionally, across all dose-escalation studies, phase two varied from 3 to 5 weeks. Relatively short periods on the increased doses limits numbers of dose-escalation responders that have time to emerge.

Finally, the increased response rate at higher doses may not need to be markedly higher than the 7–10% difference observed in the currently available studies for the difference to be clinically significant. The 7–10% difference in response approximates the difference between placebo and active medication frequently reported in studies of SSRIs [e.g., Khan et al., 2000; Thase et al., 2001].

Is the higher dose adverse event dropout rate likely to be large enough to make the strategy of raising the dose inefficient? The ultimate answer awaits a properly designed trial. It is likely that dropouts due to adverse events will be greater on higher doses. However, among the medications in our meta-analysis several pieces of currently available evidence suggest that higher dose adverse event dropouts probably would be markedly less than almost all the studies in our meta-analysis. The one study in our meta-analysis that did escalate patients to higher doses reported no adverse event dropouts [Fava et al., 1994]. In this study, after tolerating and failing fluoxetine at 20 mg for 8 weeks, patients in the "increase dose" arm were raised to 40 mg for 1 week and then in the absence of response the patients were raised to 60 mg if they could tolerate side effects. One other study, which was too dissimilar in design to include in our meta-analysis, also escalated to higher doses [Hebenstreit et al., 1989]. It reported that no adult patients withdrew due to adverse events. The adverse event rate was approximately the same for lower and higher doses and fell over time as doses increased. Additionally, if the potential dose response were sufficiently promising in properly designed studies employing titration, then it could be worth investigating further ways to minimize adverse event dropouts so the expressed dose response could be increased.

It is worth remembering that, from a clinical point of view, one may specifically be interested in the response of patients who do tolerate higher doses as long as a large proportion of patients can tolerate higher doses. All the reported studies in our analysis, even with little or no attention to lessening side effects, show that a large majority of patients do tolerate higher doses. The dose response of dose-tolerant patients in studies that more closely reflect best clinical practice, as well as in best clinical practice itself, could only be higher than the potential dose response based on current studies.

LIMITATIONS OF OUR ANALYSIS

First, any meta-analysis is at best an approximation of potential data from an actual large-scale trial.

Studies comparing meta-analyses' results with the results from actual large scale RCTs that address the same questions have shown limited agreement [LeLorier et al., 1997].

More specifically, the present meta-analysis has several limitations. First, it is based on the small number of studies available. Second, missing or partial information in three studies required us to estimate the *SD*s in these three cases [Dunner and Dunbar, 1992; Fava et al., 1994; Schweizer et al., 1990]. Two of these studies [Fava et al., 1994; Schweizer et al., 1990] are relatively small. It is unlikely any reasonable assumption about their *SD*s would have altered the overall results.

Our meta-analysis only applies to the type of patients included in the original studies. The overall analyses of these studies do not address special sub-populations of depressed patients. Secondary analyses in some of the already discussed studies and other studies have reported data that perhaps suggests patterns of greater response to increased doses in various sub-populations of depressed patients such as recurrent depression [Montgomery et al., 1994], melancholic depression [Tignol et al., 1992], more severe depression [Benkert et al., 1997; Montgomery et al., 1994], or partial responders vs. complete nonresponders [Fava et al., 1994]. Most of the analyses either have had very small *n*s and the patterns have been statistically nonsignificant or the reports give limited accounts of their results. However, such sub-populations may have a significant effect on response rates and any future prospective studies should describe these factors in their populations.

The study designs within each group were quite similar but not identical. The study that was most different from others in its group was Fava [Fava et al., 1994]. However, for reasons expressed earlier, we believe if any bias exists in this design, the bias is against finding a difference between doses. We did have to combine the outcome data and *SD*s for the two SSRI lower dose groups in this study.

The SSRI dose-equivalence structure that we used could be debated. The structure can alter the numerical value of the slope, but our sensitivity analysis suggested that the overall conclusions would not be altered.

Our inference from our statistical analysis that there is a relation between dose and response is clearly subject to an alternative explanation. Use of a DT sample for analysis results in a systematic relation between dose groups on one hand and tolerance of higher doses on the other. Therefore, if one believes there is a relation between tolerance and response, one could argue that our results are explained by the fact that selecting DT samples also selects for a subgroup of subjects who are more likely to respond at a higher rate across dose levels. Although we know of no evidence to support the theory of a relation between tolerance and response, we cannot exclude this possibility given the design of the currently available studies. Finally, to address the expressed dose-response question for best clinical practice and studies that reflect such practice, we initially use the results from the current studies' DT sample as an approximation (as justified in our companion article) and then qualitatively address the likely effect of current studies' limited purification of the non-responder sample and of unavoidable dose-related side-effect dropouts. Clear estimation of each of these effects will require appropriately designed new trials.

CONCLUSIONS

First, we suggest that we have demonstrated the conceptual and practical usefulness of distinguishing clearly between "potential" and "expressed" dose response as outlined in the companion article. Second, we believe our analysis, acknowledging its limits, suggests it is likely there is a potential dose response for SSRIs in treating major depression in the currently available studies, best clinical practice, and potentially in studies designed to reflect such practice. Third, we believe better evidence is needed, but there is a clear possibility that an expressed SSRI dose-response relation could exist in best practice and studies reflecting this practice, if patients are selected appropriately and SSRIs are escalated appropriately.

Ackowledgments. We gratefully acknowledge the assistance of Jeanne Dubin in preparation of the manuscript, and of Roy Money M.S. and Eliazbeth Neuse M.A. in preliminary analysis of the data. This work was supported by the Public Health Service (KO8 MH-01718 to C.B.B. and R24 MH54446 to S.W.W.) and NARSAD (R03051 to C.B.B.).

REFERENCES

AHCPR. 1993. Depression in primary care, vol 2: treatment of major depression, Washington, D.C.: U.S. Department of Health and Human Services.

Altamura AC, Montgomery SA, et al. 1988. The evidence for 20mg a day of fluoxetine as the optimal dose in the treatment of depression. Br J Psychiatry (Suppl 3):109–112.

Amin M, Lehmann H, et al. 1989. A double-blind, placebo-controlled dose-finding study with sertraline. Psychopharmacol Bull 25:164–167.

Baker BC, Woods SW. 2003. Is there a ssri dose response in treating major depression? The case for re-analysis of current data and for enhancing future study design. Depress Anxiety 17:10–18.

Beasley CM, Jr, Bosomworth JC, et al. 1990. Fluoxetine: relationships among dose, response, adverse events, and plasma concentrations in the treatment of depression. Psychopharmacol Bull 26:18–24.

Beasley CM, Jr, Sayler ME, et al. 1992. Fluoxetine: activating and sedating effects at multiple fixed doses. J Clin Psychopharmacol 12:328–333.

Benkert O, Szegedi A, et al. 1997. Dose escalation vs. continued doses of paroxetine and maprotiline: A prospective study in depressed out-patients with inadequate treatment response. Acta Psychiatr Scand 95:288–296.

Dornseif BE, Dunlop SR, et al. 1989. Effect of dose escalation after low-dose fluoxetine therapy. Psychopharmacol Bull 25:71–79.

Dunner DL, Dunbar GC. 1992. Optimal dose regimen for paroxetine. J Clin Psychiatry 53:21–26.

Fabre LF, Abuzzahab FS, et al. 1995. Sertraline safety and efficacy in major depression: A double-blind fixed-dose comparison with placebo. Biol Psychiatry 38:592–602.

Fabre LF, Putman HP. 1987. A fixed-dose clinical trial of fluoxetine in outpatients with major depression. J Clin Psychiatry 48:406–408.

Fava M, Rosenbaum JF, et al. 1994. Lithium and tricyclic augmentation of fluoxetine treatment for resistant major depression: A double-blind, controlled study. Am J Psychiatry 151: 1372–1374.

Hebenstreit GF, Fellerer K, et al. 1989. A pharmokinetic dose titration study in adult and elderly depressed patients. Acta Psychiatr Scand 80:81–84.

Khan A, Warner HA, et al. 2000. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials. Arch Gen Psychiatry 57:311–317.

LeLorier J, Gregoire G, et al. 1997. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 337:536–542.

Montgomery SA, Pedersen V, et al. 1994. The optimal dosing regimen for citalopram: A meta-analysis of nine placebo-controlled studies. Int Clin Psychopharmacol 9:35–40.

Montgomery SA, Rasmussen JGC, et al. 1992. Dose response relationship of citalopram 20 mg, citalopram 40 mg and placebo in the treatment of moderate and severe depression. Int Clin Psychopharmacol 6:65–70.

Post RM. 1995. Mood disorders: somatic treatment. In: Kaplan HI, Sadock BJ. Comprehensive textbook of psychiatry/IV. Baltimore, Williams & Wilkins. 1:1153–1178.

Preskorn SH, Lane RM. 1995. Sertraline 50 mg daily: the optimal dose in the treatment of depression. Int Clin Psychopharmacol 10:129–141.

Schweizer E, Rickels K, et al. 1990. What constitutes an adequate antidepressant trial for fluoxetine. J Clin Psychiatry 51:8–11.

Thase ME, Entsuah AR, et al. 2001. Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. Br J Psychiatr 78:234–241.

Thase ME, Rush AJ 1995. Treatment-resistant depression. In: Bloom FE, Kupfer DJ. Psychopharmacology: The fourth generation of progress. New York: Raven Press, Ltd. p 1081–1097.

Tignol J, MJ Stoker, et al. 1992. Paroxetine in the treatment of melancholia and severe depression. Int Clin Psychopharmacol 7:91–94.

Tweedie RL, Mengersen KL. 1995. Meta-analytic approaches to dose-response relationships, with application in studies of lung cancer and exposure to environmental tobacco smoke. Statistics Med 14:545–569.

Wernicke JF, Dunlop SR, et al. 1988. Low-dose fluoxetine therapy for depression. Psychopharmacol Bull 24:183–188.

Wernicke JF, Dunlop SR, et al. 1987. Fixed-dose fluoxetine therapy for depression. Psychopharmacol Bull 23:164–168.