The Nortriptyline Therapeutic Window: A systematic review and meta-analysis

Steven C. Bagley, MD, MS*

Abstract

Background: Some studies have found a curvilinear relationship between the nortriptyline plasma concentration and the clinical rating of depression in those treated. The concentration interval of optimum response has been called the "nortriptyline therapeutic window."

Objectives: To analyze all original data relevant to the therapeutic window.

Data sources: English language articles found in PubMed using search terms "nortriptyline," and "plasma concentration," "serum concentration," or "blood concentration". The search was extended through the use of Google Scholar citation links, and reference mining of secondary articles and books.

Study selection: Clinical trials or observational studies of adults that reported patient-level or aggregate results of depression response by concentration or concentration range were retained.

Data extraction: Plasma concentration and depression ratings were extracted from text, tables and scanned figures.

Data synthesis: Studies reporting only qualitative results were reported narratively. Quantitative patient-level and aggregate data were meta-analyzed to compare treatment inside the plasma concentration window with treatment outside the window. A meta-regression was performed to address variation in study designs and other study-level variables.

Results: In thirty-three clinical trials or observational studies, some evidence was found for a plateau or decline effect at higher plasma concentrations, but the consistency of the effect varied with the depression scale. There was marked variation in response within and outside the window, and no sharp cutoff in response at either the lower or upper window limits. A random-effects meta-analysis found a statistically significant risk difference for treatment within the window compared to treatment above the window of 0.17 (0.04–0.30, 95% CI), corresponding to a number needed to treat of almost 6. The observational studies showed stronger response than the clinical trials; this depended on a single, anomalous outlier. The meta-regression showed a non-statistically significant trend for a stronger response in observational studies, and small and non-significant effects for studies with endogenous depression and those using fixed-dose regimens. Many studies were small and did not attempt to disconfirm the window's existence by including patients with concentrations above the upper window limit.

Conclusions: Weak, inconsistent evidence of highly varying quality for the nortriptyline therapeutic window was found. Its clinical utility over prudent prescribing practice is limited.

Keywords: nortriptyline, therapeutic window, dose-response, systematic review, meta-analysis

Introduction

The rise of evidence-based medicine to guide clinical practice, along with basic research providing novel

*Department of Psychiatry and Behavioral Sciences, Stanford University, 401 Quarry Road MC 5723, Palo Alto, CA 94304-1419. Email: steven.bagley@stanford.edu.

insights into neuropsychiatric mechanisms of action, and the curation and deployment of databases of drug-drug, and drug-food interactions, have all contributed to a rigorous evidence base that supports the safe and effective use of newly developed pharmaceuticals. However, older drugs, which may still be mainstays of treatment, are not always retrospectively subjected to the same scientific scrutiny as their newer counterparts. Some researchers have revisited the evidence supporting older drugs, such as trazodone [39], but this has not become routine practice, leaving critical gaps in our understanding.

The tricyclic antidepressants, named for their common chemical structure, have been effectively used for decades in the treatment of depression, and remain valuable second-line pharmacologic agents. By the early 1970's, a problem of pharmacokinetic variation of the tricyclics had been identified. Patients treated with identical doses of a tricyclic had differing plasma concentrations of the drug and its metabolites; sampling across patients revealed a wide range of drug and metabolite concentrations of up to forty-fold but the response of each patient to a given dose was generally consistent and repeatable [1]. These differences were presumed to be due to genetic determinants of drug metabolism. Thus, it seemed plausible that patients might exhibit more consistent responses to a specific drug plasma concentration than to a specific drug dose.

This idea was examined by Åsberg and colleagues [3] who reported an unusual pharmacologic feature of one of the tricyclics, nortriptyline. They found an unexpected curvilinear relationship between its plasma level and a patient's therapeutic improvement as measured by changes in scores on a standardized depression rating scale. With increasing nortriptyline plasma levels, clinical response first improved, but at higher levels the response plateaued and ultimately declined. When response was plotted against plasma level, their data described an inverted-U shape; the middle region, containing the region of most effective response, was named the "nortriptyline therapeutic window". Follow-up studies showed mixed results, some supporting a curvilinear relationship and others failing to replicate those findings, but belief in the window continues; unqualified statements about it appear in standard psychopharmacology texts [51, 55].

The purpose of this review is to evaluate all known published evidence relevant to the existence of the nortrityline therapeutic window using a systematic review, meta-analysis and meta-regression.

Method

Search strategy

The PubMed database was searched for English language articles in the interval 1966 through August 2011, using search terms "nortriptyline" AND ("plasma concentration" OR "serum concentration" OR "blood concentration"). Titles and abstracts were reviewed, and potentially relevant articles were retrieved. Included were articles, book chapters, or case series reporting original data on adults under acute treatment for depression that related nortriptyline concentration to patient depression response with corresponding depression rating scores in at least one arm of treatment. Excluded were individual case reports, as well as articles that involved treatment with amitriptyline, of which nortriptyline is a metabolite, in order to avoid confounding effects. Using Google Scholar, articles citing results in the first search group were examined. This process of expansion through citation links was iterated until no new articles were found, or until articles outside the scope of this review were encountered. Bibliographies of retrieved articles, reviews, and psychopharmacology texts and monographs were also searched by hand.

All studies meeting entry criteria were coded without regard to completeness or methodological quality following current recommended practice [34]; some study characteristics that might influence quality were explicitly modeled in the meta-regression.

Data extraction and coding

Articles were coded in a standard format [34] in a two-dimensional table, including information on populations, interventions, comparators, outcomes, and study designs. The population was adults, including geriatric patients; age ranges or means were recorded.

The inclusion diagnosis was coded as *endogenous depression* versus *other*. This choice was motivated by the observation that the window phenomenon might be more easily detectable in endogenously depressed patients [18]. Further distinctions were not useful, because of the wide variety of diagnostic systems (including the Feighner criteria [15], Research Diagnostic Criteria [58], DSM-III, DSM-IIIR, DSM-IV), and diagnoses (including major depression, both psychotic and non-psychotic, bipolar disorder, and poorly characterized states such as "primary depressive illness") employed.

The intervention was treatment with nortriptyline. The treatment was coded for dose (or dose range), and the dose regimen: *fixed* (patient assigned to a fixed-dose or dose titration schedule throughout the trial), *variable* (dose varied in unspecific way), *level* (patient's dose adjusted to produce a predetermined target plasma level), or *clinician-determined* (dose adjusted to achieve clinical response).

Study design was recorded as clinical trial when an explicit experimenter-controlled intervention was deployed in a prospective fashion; others were coded as observational studies. When relevant data were a subset of the fully reported data, such as when one arm of a clinical trial was treated with nortriptyline, only the data from that subset were retained. Therefore, the sample sizes used for the meta-analysis in some cases vary from the total size reported in the original article.

Individual patient data were taken from the text, tables, and figures, extracting the data points from scanned images where necessary. Following current practice, outcomes were taken as final depression rating scores [27]; when available, initial or change scores were also recorded and analyzed.

Three depression rating scales were encountered: the Crönholm-Ottosson (CO) scale [4], the Hamilton Depression Rating Scale (HAMD) [21], and the Montgomery-Åsberg Depression Rating Scale (MADRS) [40]. To convert individual (patient-level) data into aggregate remission data, scores were thresholded; patients were assigned remission status if their scores were less than or equal to 3 for CO scores, 7 for HAMD scores, and 9 for MADRS scores [54, 29]. The consensus range of 50–150 ng/mL was used as the therapeutic window [35]. When aggregate results were reported, the stated depression cutoffs (or other criteria) for determining remission (or recovery) were recorded, however, in some cases, the reported criterion was not what is now accepted as the depression cutoff for remission. Patient-level data were plotted along with visual demarcations of the appropriate remission cutoff (shown in the plots as a horizontal rectangle), the therapeutic window (vertical rectangle), and a locally weighted regression smoother (loess curve) [12] as a non-parametric summary of the overall trend.

Data analysis

The effect size was defined to be the risk difference for recovery in the treatment group compared to recovery in the control group. The treatment group were those patients whose plasma concentrations were inside the therapeutic window. The control group were those patients with concentrations above the window; an alternative control group including patients either above or below the window was also examined.

Formally, with the following definitions: in is the number of subjects with plasma concentrations within the therapeutic window, in_{rec} is the number of such patients who have depression ratings in the recovered range, $above$ is the number of subjects with concentrations above the window's upper limit, and $above_{rec}$ is the number of such patients recovered, then the risk difference (RD) effect size is calculated as:

$$RD = \frac{in_{rec}}{in} - \frac{above_{rec}}{above}$$

Because of the diversity of populations and study designs considered, DerSimonian-Laird random-effects modeling [13] was employed. Publication bias

was assessed using a funnel plot and the trim and fill method of Duval and Tweedie [14]. Heterogeneity was computed using the Cochrane Q and I^2 statistics [7]. Various continuity corrections were applied to avoid zero in the denominators in the risk difference (and also in the odds ratio when used); typically 0.5 was added to each value. All of these choices of meta-analysis parameters were varied, as detailed below in the sensitivity analysis. The canonical interpretation of the significance of statistical tests with a threshold of 0.05 (5%) was employed. The quality of evidence was assessed according to the GRADE criteria [20].

Software

The analysis was conducted using the R language, [52]. The `metafor` package (version 1.6-0) was used to perform the meta-analysis [60], with additional programming by the author.

Results

A PubMed search was performed on September 5, 2011, returning 272 articles. Thirty-three original studies on the nortriptyline therapeutic window meeting the inclusion criteria were found, and are shown in Table 1. Three pairs of duplicate publications were identified: [29, 28], [56, 31], [63, 64]; data from the duplicate studies were merged, leaving 30 studies for this review. The 30 studies fall into three classes: 16 reported patient-level data; these data sets are shown collectively in a subsequent section. They were also converted to aggregate data using the thresholding scheme described in the methods section, and combined with the 7 studies reporting only aggregate data; these 23 studies entered into the meta-analysis. The 7 remaining studies reported results without any supporting patient-level or aggregate data, and are described qualitatively in the following section.

Some problems with the data were uncovered. Åsberg et al. [3] reported that one patient made a suicide attempt in the second week of their study, "which made formal rating impossible". However, they chose to keep that patient's data and arbitrarily assigned an amelioration (rating scale change) score of zero. As this imputation seems unjustified, that point was removed from the data set used here. Montgomery et al. [41] provided few details about the design of their study; details of recruitment and inclusion were not reported beyond noting that the patients were depressed and not taking antidepressants. Recovery was assessed "globally" at one hospital site and retrospectively by chart review at the other. Because of the uncertainty of its study design, it was assigned to the "observational" category.

Studies with only qualitative results

Burrows and colleagues published two trials in 1974. In the first [9], 80 patients with "primary depressive illness" were treated with nortriptyline with doses 75–250 mg per day (as determined by the clinician) for four weeks. They concluded that their results showed: "a lack of relationship between clinical response and plasma nortriptyline levels". In the second [10], 40 patients were treated in a sequential matched-pair design, where one patient of the pair was randomly assigned to a low plasma level (below 49 ng/mL) and the other to a high level (above 140 ng/mL), each for four weeks. Their trial, "showed no differences between these plasma levels and clinical response in twenty pairs of depressed subjects".

Fensbo [16] described a trial in which 23 patients with endogenous depression were treated with nortriptyline 50 mg tid for 4 weeks. Depression was rated on an idiosyncratic scale using scores 0–5. Concentration of nortriptyline in whole blood was measured (although the article title mentioned "serum"). No patient-level data were reported. Their averaged aggregate data do appear in a figure in that paper suggesting that there may be a decline in response above 200 ng/mL. However, Fensbo concluded: "[N]o statistically significant correlation between any blood concentration level and effect could be found. Moreover, neither a lower or an upper bound limit for a therapeutic response could be defined."

Lipsey and colleagues [36] randomly assigned depressed post-stroke patients to nortriptyline (11 com-

pleters) or placebo (15 completers) for 4–6 weeks of treatment with a nortriptyline dose titrated up to 100 mg at night. All patients who completed the trial were within the serum concentration range of 50–140 ng/mL by the end. The average HAMD score fell from 13.9 to 2.7; this was statistically significant when compared to their placebo group, but no patients were treated outside of the therapeutic window.

Ng Ying Kin et al. [46] provided results of a trial (with additional details reported in [45]). In this trial geriatric patients (29 completers) were treated with nortriptyline for up to 7 weeks. The dose was adjusted to maintain a serum level between 50 and 170 ng/mL. They used a depression remission threshold of a Hamilton depression rating score less than or equal to 10. Their conclusion about therapeutic window was limited to the following statements: "The rate of clinical remission in this study was relatively high compared to placebo. These results tend to lend support for a similar therapeutic window in the depressed elderly."

Bondareff et al. [6] described a trial involving outpatients (70 completers) with DSM-III-R major depressive disorder. They were treated with nortriptyline for 12 weeks, with the dose adjusted to achieve clinical response. They concluded: "The rate of response did not differ in patients with nortriptyline levels below 50 ng/mL or those with plasma levels 50–150 ng/mL, but patients with plasma levels above 150 ng/mL showed suggestive but nonsignificant lower rates of improvement (χ^2=3.53, df=1, p=0.06)."

Streim et al. [59] reported on the treatment of 69 frail, elderly, possibly demented, nursing home residents who were randomized to treatment with a fixed dose after titration of either low dose or normal dose of nortriptyline and treated for 10 weeks. For patients with relatively intact cognitive function (Mini-Mental State Exam score > 18) [17], they found a quadratic (inverted-U) relationship between plasma concentration and improvement measured as fractional improvement (i.e., percent change) in the HAMD score. Using steady-state values at 28 days, they found a therapeutic window of 40–107 ng/mL. Using data for trial completers at 10 weeks, they found a window of 42–111 ng/mL. No patients were treated above the upper limit of the consensus therapeutic window, 150 ng/mL.

Thus, of the studies reporting only qualitative results, or aggregates not amenable to more detailed analysis, three had negative results (that is, no evidence for the therapeutic window), three treated all patients within the therapeutic window (and thus were unable to disconfirm the window by testing higher concentrations), and one found a nonsignificant trend towards lower response at higher plasma levels. The GRADE quality of evidence for this set of studies with widely disparate patient populations, trial designs, and treatment regimens is very low.

Studies with quantitative patient-level data

Sixteen studies reported individual patient-level data in tables or figures. All but one of these studies used either the Crönholm-Ottosson depression rating scale or the Hamilton depression rating scale. For each of those two scales, composites of all data using that scale are shown in Figure 1. One remaining study used the Montgomery-Åsberg depression rating scale, and is not shown.

At low plasma concentrations (less than 50 ng/mL), the antidepressant response increases as the plasma concentration increases, as would be expected.

When the plasma concentrations falls within the therapeutic window (50 to 150 ng/mL), the average response (shown as the loess smoother) shows a mildly inverted-U shape for the CO scale and a very slight increase with concentration for the HAMD scale. The MADRS data set shows no clear trend (data not shown). However, focusing on the average response ignores the tremendous variation in antidepressant response. In general, under conditions of high variance, the average is not a very enlightening summary statistic. The distribution of the final depression scores is extremely broad, making it hard to use the window to guide treatment. Most importantly, many patients who were treated with concentrations in the window remained depressed. When

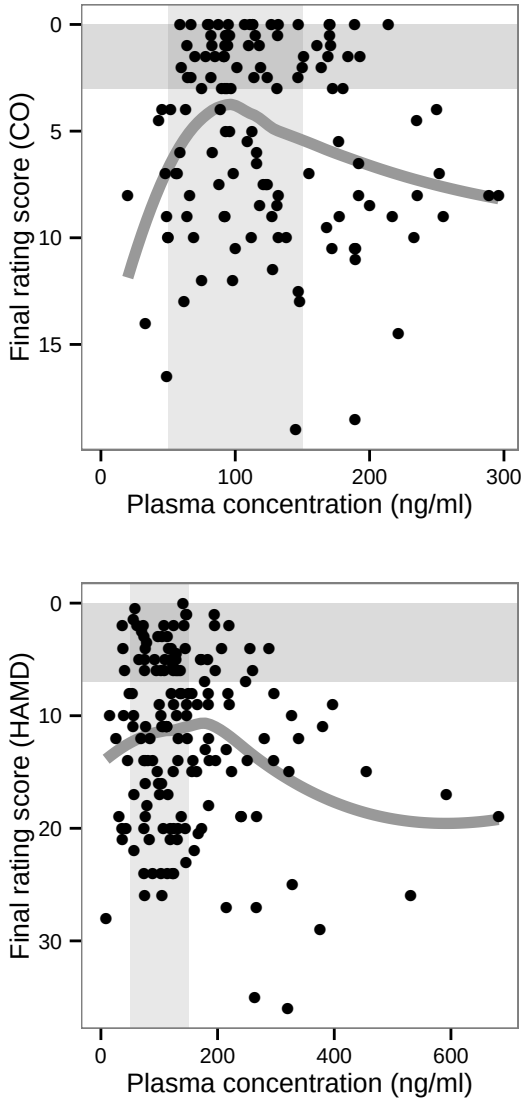

FIGURE 1: Composite of all data sets using Crönholm-Ottosson (upper) and Hamilton (lower) depression rating scales. Remission window (horizontal bar) and therapeutic window (vertical bar) are highlighted in gray. Loess smoother is also plotted.

the patient level data are combined with the aggregate data, the positive predictive value is 52%, so knowing that the treatment is within the therapeutic window predicts a clinical response about half of the time. The corresponding positive predictive value for concentrations greater than the upper limit of the window is 33%.

At plasma concentrations above the window (greater than 150 ng/mL) the response seems to plateau, and slightly declines. The possibility of a confounding effect arising from dose escalation for nonresponsive patients will be explored below.

The GRADE quality of evidence for the patient-level data was assessed to be low because of the inconsistency and high variance of the results.

Meta-Analysis

The meta-analysis combined all studies for which patient-level or aggregate data were reported. The results are shown in Figure 2 for the default choices of the meta-analysis parameters: risk difference as the effect size measure, using the DerSimonian-Laird random effects method, a continuity correction for all values, and comparing the window treatment to a control group, defined to contain those whose concentration is greater than the window upper limit.

The risk difference was calculated to be 0.17 (0.04–0.30, 95% CI), which is statistically significant at a 0.0105 level, rejecting the null hypothesis of no effect. This risk difference corresponds to a number-needed-to-treat of 5.9 (3.4–25, 95% CI). Incidental note is made that the Åsberg et al. study, the original publication promoting the idea of the nortriptyline therapeutic window, has the lowest risk difference (smallest effect size) of all the studies examined.

Heterogeneity is that part of the observed variation between studies due to true differences between them; the remainder of the variation is attributed to random sampling. The I^2 statistic reports the percentage of total variation not due to sampling. In this meta-analysis, I^2 is 41.7, a level of heterogeneity typically considered "moderate" [22]. For the set of studies considered in this meta-analysis, one important factor to consider is that because many of the studies have small sample sizes, and therefore small

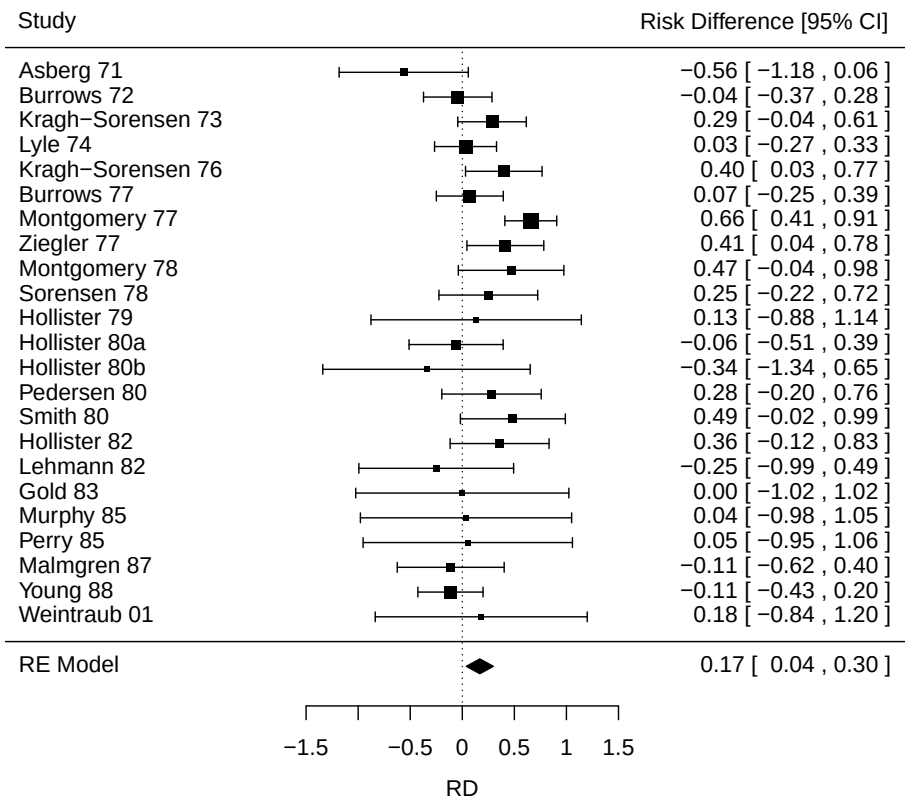

FIGURE 2: Forest plot for meta-analysis of treatment within nortriptyline therapeutic window versus treatment at concentrations above the window. Effect size is computed as RD, the risk difference.

precision weights, they contribute relatively little to the summary effect. This leads to wide confidence intervals, with most of the observed variation ascribed to random variation, not underlying heterogeneity. Given that these studies vary in many aspects of their design and implementation, it seems appropriate that the heterogeneity is at least moderate.

Sensitivity analysis

The meta-analysis was re-run using different parameters. The choices include different methods (fixed-effects or restricted maximum-likelihood estimation (REML) instead of DerSimonian-Laird) different effect size measures (odds ratio instead of risk difference), a different treatment group (inside the window) to those outside (either above or below) instead of only above, and various options for the continuity correction. Using no continuity correction left

only 11 studies, and was therefore omitted. There was no difference in using REML in the place of DerSimonian-Laird, and little difference in using a fixed-effects model (results not shown). Different values of the continuity correlation and the control group also had little effect. The results are numerically different when using an odds ratio effect size, but qualitatively the same, and similarly unchanged in corresponding sensitivity analyses.

Studies varied quite a bit in their effect size values and in their precision. The influence of each study was tested by leaving it out of the analysis in turn. Doing so reveals that the Montgomery study [41] had the most influence; it led to the lowest estimate of the effect size when omitted, and produced the lowest values for Q and I^2. Its influence was also suggested by its outlying position in the forest plot. It produced an I^2 of zero, meaning that the observed variation of all the studies except for the Montgomery study could be due to chance, with no difference in the true effect sizes. The Montgomery study is distinctive and not representative; some of the issues with this study were mentioned at the beginning of the results section. That the Montgomery study occupies an outlying position is not to suggest that the other studies are coherent evidence of a single, robust signal. It is more likely that many of the studies are not precise enough to matter much when assessing heterogeneity.

The GRADE quality of evidence for the meta-analysis data was assessed as low, because of the small study sizes, and the sensitivity analysis results just mentioned.

Meta-Regression

To assess the effect of several of the study variables on the effect size, a meta-regression was performed, using diagnosis (nonendogenous versus endogenous), design (observational versus clinical trial), and dose regimen (not fixed versus fixed) as covariates. These were chosen because they seemed a priori to be related to study quality; additional variables were not introduced because of the limited amount of data available for the regression analysis.

The meta-regression results appear in Table 2. The coefficient for observational studies is largest in magnitude—this is due to the presence of the anomalous Montgomery study [41]—but falls short of statistical significance. The results for design and diagnosis are small and not statistically significant.

Publication bias

Publication bias occurs when some results (usually those favoring the treatment) are preferentially published. One method for detecting publication bias is through visual inspection of a funnel plot, which graphs the standard error against the risk difference. Small studies will have large standard errors; as the study size grows towards infinity the sampling error approaches zero, so the funnel plot should appear as a symmetrical funnel pointed upwards. Publication bias is suspected when small studies with negative results (here, lower values of the risk difference) are missing. The funnel plot for the nortriptyline studies is shown in Figure 3.

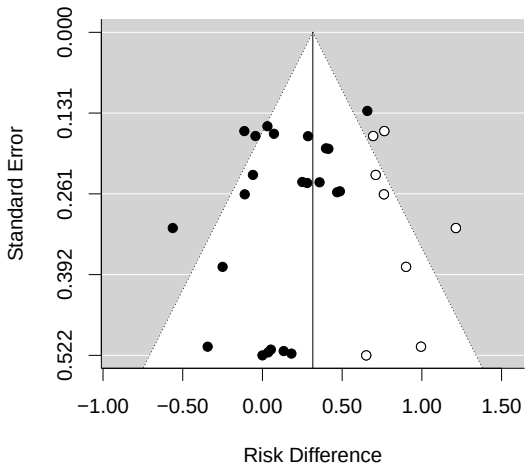

FIGURE 3: Funnel plot, using the trim and fill method to impute possibly missing studies (open circles); eight such studies with a positive risk difference may be missing.

Publication bias can be tested quantitatively using the trim and fill method, which imputes missing studies and adds them to the funnel plot (shown in the funnel plot as eight open circles). Both visual inspection and the trim and fill method suggest that studies may be missing from the lower right area of the funnel, a region corresponding to studies with large standard errors (small sample size), with relatively large risk differences; that is, small studies showing response in the window. Such studies are unlikely to have been withheld from publication. That is, in order for there to have been publication bias, studies supporting the therapeutic window would have to have been suppressed, which seems improbable. Thus, no evidence of publication bias favoring the therapeutic window was found.

Discussion

This systematic review of all known published studies relating patient depression response to the nortriptyline therapeutic window examined 16 data sets reporting patient-level data, and found a slight decline in response in those using the Crönholm-Ottosson depression rating scale, starting within the window. In those using the Hamilton depression rating scale, response improved throughout the window, extended a bit beyond the window, then showed a modest decline. These are group-level effects; pronounced variation in response was seen both within and outside the therapeutic window. There is no evidence for a sharp cutoff at either the lower or upper window limits. Many of the studies were small and of low precision. Many provided data of limited evidential power as most or all patients were treated within the therapeutic window, not allowing for disconfirmation of a declining response above the window's upper limit.

Twenty-three studies had aggregate data that were used in a meta-analysis. Using commonly accepted parameters in the meta-analysis produced a statistically significant effect for response within the window compared to response above the window (risk difference of 0.17, p=0.01). This effect, as with that reported in the previous paragraph, was an aggregate one, based on disparate studies of generally large variance and weak design. In the meta-regression, observational studies showed a stronger, but not statistically significant, response over clinical trials, but those studies were quite disparate and the strong response hinged on a single outlier. The effect of endogenous depression and fixed-dose regimen was small and not statistically significant. A priori, one would expect that fixed-dose studies would be isolated from the confounding effect of response on level, as clinicians might escalate doses for nonresponders when allowed to do so; however, this effect was not detectable in the limited data available. The evidence reported here shows that even if a therapeutic window does exist, it was not reliably and robustly detected in studies that mimic typical clinical conditions, across types of depression, and independent of the rating scale used for measuring depression response.

The veridicality of the nortriptyline therapeutic window was an issue of some dispute in the 1970's [18, 33, 2]. One prominent explanation for the disparity between those studies finding a window and those that did not was that the window response was primarily seen in endogenously depressed patients. As noted, the data reported here show a trend towards this, but one not reaching statistical significance.

Overall, it does not appear that there is a clear role for clinical use of the therapeutic window. The predictive value of knowing that the patient's concentration is within the window is limited, and there is no strong cutoff at the consensus upper limit of 150 ng/mL that would make it a useful target guiding prescribing practice. Independent of the therapeutic window, prudent practitioners would initiate treatment at lower doses, increase the dose when indicated and as tolerated, assessing for signs of response and side effects. It is not clear that the therapeutic window concept contributes much beyond this procedure. As with other medications, there remains value in the qualified use of plasma concentration monitoring for nortriptyline to address issues of patient adherence to recommended treatment, to investigate in cases where abnormalities of metabolic function are suspected, and for toxicologic

purposes. Since nortripytline is metabolized into 10-hydroxy-nortriptyline, which appears to have antidepressant activity [5], a therapeutic window measuring nortriptyline alone would be compromised by the variability in the balance between the competing pharmacokinetic and pharmacodynamic processes for those two compounds, making the whole idea of a window for nortriptyline a somewhat dubious one.

Previous reviews in this area have reached less nuanced conclusions. Typically they did not survey all extant studies, or use quantitative methods of evidence synthesis. Perry et al. [49] modeled the results of 7 studies using quadratic regression; the quadratic term was found to be not statistically significant, although they argued for the existence of the window on other grounds. Perry et al. [50] used a receiver operating characteristic (ROC) analysis, finding optimal cutpoints for the therapeutic window of 58–148 ng/mL. However, the use of ROC curves in this manner requires that costs be assigned to correct and incorrect classifications, which, in their analysis, was implicit and amounted to finding the location of crossing points of responder and nonresponder density functions, an arbitrary and poorly motivated choice. Ribeiro et al. [53] conducted a more comprehensive meta-analysis; however, they missed two studies [47, 19], and produced a best estimate of the window using an ROC analysis, again with the same failure to make explicit the costs of classification errors. They did not study the effect of study-level covariates using a meta-regression. They reported a best-fit therapeutic window of 46–236 ng/mL, but this provides little clinical guidance as it would encompass the vast majority of those in treatment.

The standard cautions about systematic reviews and meta-analyses apply to this review. The results depend on the completeness of the article search, although there was no evidence of publication bias that would have withheld negative results. The strengths of the surveyed primary studies are in the main quite limited. Many of the studies date from the 1970's, a time when clinical studies were less rigorously designed, and standards for reporting were lax. Many of the studies were small, one reporting relevant data on only 5 patients. Authors often assumed the window hypothesis to be correct, and did not design their studies to collect evidence that could disconfirm it by placing patients outside of the window limits. In spite of the very limited evidence, texts and reviews routinely cite the nortriptyline therapeutic window as if it were an established fact of psychopharmacology.

Thus, weak, inconsistent evidence of highly varying quality for a nortriptyline therapeutic window was found. The inconsistency of the evidence warrants caution in generalization to patients outside of the original treatment groups. Its clinical utility for prospective prediction of depression response over routine, prudent prescribing practice appears to be limited.

Acknowledgments

These efforts have spanned more than a decade. The original Åsberg paper on the nortriptyline therapeutic window was brought to my attention by Ian Cook, MD, in a reading group during my psychiatry residency. UCLA Residency Training Director James Spar, MD, provided both funds to support my encyclopedic photocopying efforts, and a very careful reading of the logic of the paper. Robert V. Ashley, MD, Joel Braslow, MD, PhD, and Stephen L. Read, MD, were enthusiastic supporters at various stages of this endeavor.

Preliminary versions of a portion of this material were presented in the following two talks: "What Evidence Supports the Nortriptyline Therapeutic Window," Department of Veterans Affairs, Palo Alto Healthcare System MIRECC, Aug 26, 2009 and "Why We Think There is a Nortriptyline Therapeutic Window: Clinical Epidemiology Meets the Sociology of Science," Department of Veterans Affairs, Greater Los Angeles Healthcare System, West LA VA Department of Psychiatry Grand Rounds, Oct 22, 2009; participants at both sessions provided useful insights.

Conflicts of interest: None.

Version: Finished 2013, minor edits 2022. This document produced May 13, 2022.

References

[1] B. Alexanderson, D. Price Evans, and F. Sjöqvist. Steady-state plasma levels of nortriptyline in twins: influence of genetic factors and drug therapy. *British Medical Journal*, 4(5686):764–768, 1969.

[2] A. Amdisen. Blood concentration of cyclic antidepressants as a daily routine? A critical review from the doorstep of the clinical laboratory. *Acta Psychiatrica Scandinavica. Supplementum*, 61 Suppl 280:261–280, 1980.

[3] M. Åsberg, B. Crönholm, F. Sjöqvist, and D. Tuck. Relationship between plasma level and therapeutic effect of nortriptyline. *British Medical Journal*, 3(5770):331–334, 1971.

[4] P. Bech. The Cronholm-Ottosson depression scale: the first depression scale designed to rate changes during treatment. *Acta Psychiatrica Scandinavica*, 84(5):439–445, 1991.

[5] L. Bertilsson, B. Mellström, and F. Sjöqvist. Pronounced inhibition of noradrenaline uptake by 10-hydroxymetabolites of nortriptyline. *Life Sciences*, 25(15):1285–1292, 1979.

[6] W. Bondareff, M. Alpert, A. Friedhoff, E. Richter, C. Clary, and E. Batzar. Comparison of sertraline and nortriptyline in the treatment of major depressive disorder in late life. *American Journal of Psychiatry*, 157(5): 729–736, 2000.

[7] M. Borenstein, L. Hedges, J. Higgins, and H. Rothstein. *Introduction to Meta-analysis*. Wiley, 2009.

[8] G. Burrows, B. Davies, and B. Scoggins. Plasma concentration of nortriptyline and clinical response in depressive illness. *The Lancet*, 300 (7778):619–623, 1972.

[9] G. Burrows, B. Scoggins, L. Turecek, and B. Davies. Plasma nortriptyline and clinical response. *Clinical Pharmacology and Therapeutics*, 16(4):639–644, 1974.

[10] G. Burrows, L. Turecek, B. Davies, R. Mowbray, and B. Scoggins. A sequential trial comparing two plasma levels of nortriptyline. *Australian and New Zealand Journal of Psychiatry*, 8(1):21–23, 1974.

[11] G. Burrows, K. Maguire, B. Scoggins, J. Stevenson, and B. Davies. Plasma nortriptyline and clinical response—a study using changing plasma levels. *Psychological Medicine*, 7(01): 87–91, 1977.

[12] W. Cleveland and S. Devlin. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, pages 596–610, 1988.

[13] R. DerSimonian and N. Laird. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3): 177–188, 1986.

[14] S. Duval and R. Tweedie. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2):455–463, 2000.

[15] J. P. Feighner, E. Robins, S. B. Guze, R. A. Woodruff Jr, G. Winokur, and R. Munoz. Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26(1):57, 1972.

[16] C. Fensbo. A clinical trial of nortriptyline and maprotiline with cardiovascular monitoring and serum level estimations. In A. Jukes, editor, *The Biochemical and Physiological Role of Ludiomil*, pages 181–90. CIBA Laboratories: Horsham, England, 1977.

[17] M. Folstein, S. Folstein, and P. McHugh. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiat Res*, 12(3):189–198, 1975.

[18] A. Glassman, J. Schildkraut, P. Orsulak, D. Kupfer, R. Shader, J. Davis, C. B., J. Perel, G. Klerman, and D. Greenblatt. Tricyclic antidepressants—blood level measurements and clinical outcome: an APA task force report. *Am J Psychiatry*, 142:155–62, 1985.

[19] M. Gold, A. Pottash, A. Stoll, D. Martin, L. Finn, and I. Extein. Nortriptyline plasma levels and clinical response in patients with familial pure unipolar depression and blunted trh tests. *The International Journal of Psychiatry in Medicine*, 13 (3):215–220, 1983.

[20] G. Guyatt, A. Oxman, G. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello, and H. Schünemann. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*, 336(7650):924–926, 2008.

[21] M. Hamilton. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23(1):56–62, 1960.

[22] J. Higgins, S. Thompson, J. Deeks, and D. Altman. Measuring inconsistency in meta-analyses. *Bmj*, 327(7414):557–560, 2003.

[23] L. Hollister. Monitoring tricyclic antidepressant plasma concentrations. *JAMA: The Journal of the American Medical Association*, 241(23): 2530–2533, 1979.

[24] L. Hollister. Plasma concentrations of tricyclic antidepressants in clinical practice. *Journal of Clinical Psychiatry*, pages 66–69, 1982.

[25] L. Hollister, K. Davis, and P. Berger. Subtypes of depression based on excretion of MHPG and response to nortriptyline. *Archives of General Psychiatry*, 37:1107–1110, 1980.

[26] L. Hollister, A. Pfefferbaum, and K. Davis. Monitoring nortriptyline plasma concentrations. *American Journal of Psychiatry*, 137(4):485–486, 1980.

[27] M. Keller. Remission versus response: the new gold standard of antidepressant care. *J Clin Psychiatry*, 65 suppl 4:53–59, 2004.

[28] P. Kragh-Sørensen, M. Åsberg, and C. Eggert-Hansen. Plasma-nortriptyline levels in endogenous depression. *The Lancet*, 301(7795):113–115, 1973.

[29] P. Kragh-Sørensen, C. Hansen, and M. Åsberg. Plasma levels of nortriptyline in the treatment of endogenous depression. *Acta Psychiatrica Scandinavica*, 49(4):444–456, 1973.

[30] P. Kragh-Sørensen, C. Eggert Hansen, P. Baastrup, and E. Hvidberg. Self-inhibiting action of nortriptyline's antidepressive effect at high plasma levels. *Psychopharmacology*, 45(3): 305–312, 1976.

[31] V. Kumar, R. Smith, K. Reed, and D. Leelavathi. Plasma levels and effects of nortriptyline in geriatric depressed patients. *Acta Psychiatrica Scandinavica*, 75(1):20–28, 1987.

[32] L. Lehmann, C. Bowden, F. Redmond, and B. Stanton. Amitriptyline and nortriptyline response profiles in unipolar depressed patients. *Psychopharmacology*, 77(2):193–197, 1982.

[33] R. Levine. The role of plasma concentrations in the use of tricyclic antidepressant drugs. *Prog Neuro-psychopharmacol*, 3(1-3):211–222, 1979.

[34] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis, M. Clarke, P. Devereaux, J. Kleijnen, and D. Moher. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine*, 151(4): W–65–W–94, 2009.

[35] M. Linder and P. Keck Jr. Standards of laboratory practice: antidepressant drug monitoring. *Clinical Chemistry*, 44(5):1073–1084, 1998.

[36] J. Lipsey, G. Pearlson, R. Robinson, K. Rao, and T. Price. Nortriptyline treatment of post-stroke depression: a double-blind study. *The Lancet*, 323(8372):297–300, 1984.

[37] W. Lyle, P. Brooks, D. Early, W. Leggett, G. Silverman, R. Braithwaite, J. Cuthill, R. Goulding, I. Pearson, R. Snaith, and G. Strang. Plasma concentration of nortriptyline as a guide to treatment. *Postgraduate Medical Journal*, 50(583): 282–287, 1974.

[38] R. Malmgren, A. Åberg-Wistedt, and L. Bertilsson. Serotonin uptake inhibition during treatment of depression with nortriptyline caused by parent drug and not by 10-hydroxymetabolites. *Psychopharmacology*, 92(2): 169–172, 1987.

[39] W. B. Mendelson et al. A review of the evidence for the efficacy and safety of trazodone in insomnia. *Journal of Clinical Psychiatry*, 66(4): 469–476, 2005.

[40] S. Montgomery and M. Åsberg. A new depression scale designed to be sensitive to change. *Br J Psychiatry*, 134(4):382–389, 1979.

[41] S. Montgomery, R. Braithwaite, and J. Crammer. Routine nortriptyline levels in treatment of depression. *British Medical Journal*, 2(6080): 166–167, 1977.

[42] S. Montgomery, R. Braithwaite, S. Dawling, and R. McAuley. High plasma nortriptyline levels in the treatment of depression. I. *Clinical Pharmacology and Therapeutics*, 23(3):309–314, 1978.

[43] G. Murphy, A. Simons, R. Wetzel, and P. Lustman. Cognitive therapy and pharmacotherapy: Singly and together in the treatment of depression. *Archives of General Psychiatry*, 41(1):33–41, 1984.

[44] G. Murphy, A. Simons, and R. Wetzel. Plasma nortriptyline and clinical response in depression. *Journal of Affective Disorders*, 8(2):123–129, 1985.

[45] N. Nair, M. Amin, P. Holm, C. Katona, N. Klitgaard, N. Ng Ying Kin, P. Kragh-Sørensen, H. Kühn, C. Leek, and K. Stage. Moclobemide and nortriptyline in elderly depressed patients. A randomized, multicentre trial against placebo. *Journal of Affective Disorders*, 33(1):1–9, 1995.

[46] N. Ng Ying Kin, N. Klitgaard, N. Nair, M. Amin, P. Kragh-Sørensen, G. Schwartz, S. Ahmed, P. Holm, C. Katona, and K. Stage. Clinical relevance of serum nortriptyline and 10-hydroxynortriptyline measurements in the depressed elderly: A multicenter pharmacokinetic and pharmacodynamic study. *Neuropsychopharmacology*, 15(1):1–6, 1996.

[47] J. Pedersen and J. Sørensen. Therapeutic effect and side effects in patients with endogenous depression treated with oral nortriptyline once a day. *Neuropsychobiology*, 6(1):42–47, 1980.

[48] P. Perry, J. Browne, B. Alexander, B. Pfohl, F. Dunner, A. Sherman, and M. Tsuang. Relationship of free nortriptyline levels to therapeutic response. *Acta Psychiatr Scand*, 72(2):120–125, 1985.

[49] P. Perry, B. Pfohl, and S. Holstad. The relationship between antidepressant response and tricyclic antidepressant plasma concentrations. A retrospective analysis of the literature using logistic regression analysis. *Clinical Pharmacokinetics*, 13(6):381–392, 1987.

[50] P. Perry, C. Zeilmann, and S. Arndt. Tricyclic antidepressant concentrations in plasma: an estimate of their sensitivity and specificity as a predictor of response. *Journal of Clinical Psychopharmacology*, 14(4):230–240, 1994.

[51] P. Perry, B. Alexander, L. B.I., and C. DeVane. *Psychotropic Drug Handbook*. Lippincott Williams & Wilkins, eighth edition, 2006.

[52] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org`.

[53] M. Ribeiro, E. Pereira, R. Santos-Jesus, E. Sena, K. Petribu, and I. Oliveira. Nortriptyline blood levels and clinical outcome: meta-analysis of published studies. *Rev Bras Psiquiatr*, 22(2):51–6, 2000.

[54] A. Rush, H. Kraemer, H. Sackeim, M. Fava, M. Trivedi, E. Frank, P. Ninan, M. Thase, A. Gelenberg, D. Kupfer, et al. Report by the ACNP

task force on response and remission in major depressive disorder. *Neuropsychopharmacology*, 31(9):1841–1853, 2006.

[55] A. Schatzberg and C. Nemeroff, editors. *The American Psychiatric Publishing Textbook of Psychopharmacology*. Amer Psychiatric Pub Inc, fourth ed. edition, 2009.

[56] R. Smith, K. Reed, and D. Leelavathi. Pharmacokinetics and the effects of nortriptyline in geriatric depressed patients. *Psychopharmacology bulletin*, 16(3):54–57, 1980.

[57] B. Sorensen, P. Kragh-Sørensen, N. Larsen, and E. Hvidberg. The practical significance of nortriptyline plasma control. *Psychopharmacology*, 59:35–9, 1978.

[58] R. Spitzer, J. Endicott, and E. Robins. Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry*, 35(6):773–782, 1978.

[59] J. Streim, D. Oslin, I. Katz, B. Smith, S. DiFilippo, T. Cooper, and T. Ten Have. Drug treatment of depression in frail elderly nursing home residents. *Am J Geriatric Psychiatry*, 8(2): 150–159, 2000.

[60] W. Viechtbauer. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36:1–48, 2010.

[61] D. Weintraub. Nortriptyline in geriatric depression resistant to serotonin reuptake inhibitors: case series. *J Geriatr Psychiatry Neurol*, 14(1): 28–32, 2001.

[62] R. Young, G. Alexopoulos, R. Shindledecker, A. Dhar, and H. Kutt. Plasma 10-hydroxynortriptyline and therapeutic response in geriatric depression. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, 1 (3):213–215, 1988.

[63] V. Ziegler, P. Clayton, J. Taylor, B. Tee, and J. Biggs. Nortriptyline plasma levels and therapeutic response. *Clinical Pharmacology and Therapeutics*, 20(4):458–463, 1976.

[64] V. Ziegler, P. Clayton, and J. Biggs. A comparison study of amitriptyline and nortriptyline with plasma levels. *Archives of General Psychiatry*, 34(5):607–612, 1977.

Reference	*Diagnosis*	*N*	*Age*	*Design*	*Dosing*	*Dose, duration*	*Scale*
Asberg 1971[3]	endogenous depression	28	25–72	CT	Fixed	25–75 mg tid, 2 wks	CO
Burrows 1972[8]	primary depressive illness	32	16–63	CT	Fixed	50 mg tid, 4–6 wks	HAMD
Kragh-Sorensen 1973 [28, 29]	endogenous depression	30	18–71	CT	Fixed	50 mg tid, 4 wks	CO
Burrows 1974 [9]	primary depressive illness	80	40	CT	Variable	75–250 mg, 4 wks	HAMD
Burrows 1974 [10]	primary depressive illness	40	43	CT	Level	Adjusted for level, 4 wks	HAMD
Lyle 1974 [37]	referral to hospital for depression	45	20–77	Obs	Clin	55–200 mg, at least 2 wks	Clin
Kragh-Sorensen 1976 [30]	endogenous depression	24	24–69	CT	Level	Adjusted for level, 4 wks	CO
Burrows 1977 [11]	primary depressive illness	22	28–66	CT	Fixed	50 or 200–250 mg, 5 wks	HAMD
Montgomery 1977 [41]	depressed inpatients	36	21–74	CT	Clin	75–150 mg, 3–6 wks	Clin
Ziegler 1976 [63, 64]	Feighner criteria for primary or secondary affective disorder	19	20–40	CT	Fixed	75–150 mg, 6 wks	HAMD
Montgomery 1978 [42]	endogenous depression	18	23–71	CT	Fixed	100 mg, 4 wks	HAMD
Sorensen 1978 [57]	endogenous depression	27	34–86	CT	Level	Adjusted for level, 4 wks	CO
Hollister 1979 [23]	depressed patients	14	NA	Obs	Clin	One wk at given dose	Clin
Hollister 1980 [25]	primary affective disorder with depression	17	20–57	CT	Level	75–275 mg, 4 wks	HAMD
Hollister 1980 [26]	primary affective disorder, endogenous type	20	33–62	CT	Clin	Per clinician, 38 days	HAMD
Pedersen 1980 [47]	endogenous depression	21	22–63	CT	Fixed	150 mg, 4 wks	CO
Hollister 1982 [24]	various depressive syndromes	30	NA	Obs	Clin	>= 50 mg, at least 1 wk at dose	Clin
Lehmann 1982 [32]	primary major unipolar depression by RDC criteria	5	23–66	CT	Level	Adjusted for level, 4 wks	HAMD
Gold 1983 [19]	Unipolar, nonpsychotic depressed by RDC criteria	10	30–57	CT	Level	Adjusted for level, at least 21 days	HAMD
Lipsey 1984 [36]	Moderate or severe depression	11	62	CT	Fixed	Up to 100 mg, 4–6 wks	HAMD
Murphy 1984, 1985 [43, 44]	Primary unipolar affective disorder by Feighner criteria	16	21–54	CT	Level	Adjusted for level, 12 wks	HAMD
Perry 1985 [48]	Major depressive episode by DSM-III criteria	18	15–69	CT	Level	Adjusted for level, 21 days	HAMD
Smith 1980, 1987 [56, 31]	MDD, minor depressive disorder or SAD by RDC criteria	10	60–78	CT	Fixed	150 mg, 4 wks	HAMD
Malmgren 1987 [38]	MDD, RDC criteria	10	28–64	CT	Fixed	50 mg tid, 4–6 wks	MADRS
Young 1988 [62]	MDD, DSM-III criteria	37	76.5	CT	Fixed	Mean dose 1.1 mg/kg, 4 wks	HAMD
Kin 1996 [46, 45]	Major depression by DSM-III-R criteria	38	62–68	CT	Level	Adjusted for level, up to 7 wks	HAMD
Bondareff 2000 [6]	Major depression by DSM-III-R criteria	70	>= 60	CT	Clin	25–100 mg, 12 wks	HAMD
Streim 2000 [59]	MDD, dysthymia, or minor depression by DSM-IV	41	76–82	CT	Fixed	random assignment to low or high dose, up to 80 mg, 10 wks	HAMD
Weintraub 2001 [61]	MDD, DSM-IV criteria	10	65–93	Obs	Clin	up to 75 mg, up to 8 wks	CGI-I

TABLE 1: Studies reporting data on the nortriptyline therapeutic window. N is number of completers of treatment with nortriptyline at time closest to 4 weeks. Age is years, range or mean. NA=data not available. CT=clinical trial, Obs=observational study, Fixed=fixed dose of nortriptyline, Variable=variable dose, Clin=dose adjusted to clinical response, Level=dose chosen to achieve plasma level in range, CGI-I=Clinical Global Impression-Improvement, CO=Crönholm-Ottosson depression scale, HAMD=Hamilton depression rating scale.

Variable	*Estimate*	*SE*	*P-value*	*95%CI*	
intercept	0.1842	0.1240	0.1375	-0.0589	0.4272
Diagnosis: non-endogenous	-0.0970	0.1509	0.5203	-0.3927	0.1987
Design: observational	0.3038	0.2141	0.1559	-0.1158	0.7234
Dose: not fixed	-0.0522	0.1641	0.7504	-0.3738	0.2694

TABLE 2: Model statistics for mixed-effects meta-regression using the variables diagnosis, design, and dose on the estimate of risk difference for treatment within the nortriptyline therapeutic window. Estimate is the estimate of the risk difference in the regression, SE is standard error, Z-value is the standardized estimate, CI is the confidence interval