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 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 nortriptyline 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-III-R, 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 unspecified 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 Cröhn-Ottosson (CO) scale [4], the Hamilton Depression Rating Scale (HAMD) [21], and the Montgomery-Asberg 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. Åberg 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 completers) 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 ($\chi^2=3.53$, 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 an-
tidepressant 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 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 Åberg 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.

---

**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.
<table>
<thead>
<tr>
<th>Study</th>
<th>Risk Difference [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asberg 71</td>
<td>−0.56 [−1.18, 0.06]</td>
</tr>
<tr>
<td>Burrows 72</td>
<td>−0.04 [−0.37, 0.28]</td>
</tr>
<tr>
<td>Kragh–Sorensen 73</td>
<td>0.29 [−0.04, 0.61]</td>
</tr>
<tr>
<td>Lyle 74</td>
<td>0.03 [−0.27, 0.33]</td>
</tr>
<tr>
<td>Kragh–Sorensen 76</td>
<td>0.40 [0.03, 0.77]</td>
</tr>
<tr>
<td>Burrows 77</td>
<td>0.07 [−0.25, 0.39]</td>
</tr>
<tr>
<td>Montgomery 77</td>
<td>0.66 [0.41, 0.91]</td>
</tr>
<tr>
<td>Ziegler 77</td>
<td>0.41 [0.04, 0.78]</td>
</tr>
<tr>
<td>Montgomery 78</td>
<td>0.47 [−0.04, 0.98]</td>
</tr>
<tr>
<td>Sorensen 78</td>
<td>0.25 [−0.22, 0.72]</td>
</tr>
<tr>
<td>Hollister 79</td>
<td>0.13 [−0.88, 1.14]</td>
</tr>
<tr>
<td>Hollister 80a</td>
<td>−0.06 [−0.51, 0.39]</td>
</tr>
<tr>
<td>Hollister 80b</td>
<td>−0.34 [−1.34, 0.65]</td>
</tr>
<tr>
<td>Pedersen 80</td>
<td>0.28 [−0.20, 0.76]</td>
</tr>
<tr>
<td>Smith 80</td>
<td>0.49 [−0.02, 0.99]</td>
</tr>
<tr>
<td>Hollister 82</td>
<td>0.36 [−0.12, 0.83]</td>
</tr>
<tr>
<td>Lehmann 82</td>
<td>−0.25 [−0.99, 0.49]</td>
</tr>
<tr>
<td>Gold 83</td>
<td>0.00 [−1.02, 1.02]</td>
</tr>
<tr>
<td>Murphy 85</td>
<td>0.04 [−0.96, 1.05]</td>
</tr>
<tr>
<td>Perry 85</td>
<td>0.05 [−0.95, 1.06]</td>
</tr>
<tr>
<td>Malmgren 87</td>
<td>−0.11 [−0.62, 0.40]</td>
</tr>
<tr>
<td>Young 88</td>
<td>−0.11 [−0.43, 0.20]</td>
</tr>
<tr>
<td>Weintraub 01</td>
<td>0.18 [−0.84, 1.20]</td>
</tr>
<tr>
<td>RE Model</td>
<td>0.17 [0.04, 0.30]</td>
</tr>
</tbody>
</table>

**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.

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 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 precisions. 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.

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 win-
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.

dow 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 non-responders 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 nortriptyline 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.

Acknowledgements

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,
References


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

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=Cröhn-Ottosson depression scale, HAMD=Hamilton depression rating scale.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>SE</th>
<th>P-value</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.1842</td>
<td>0.1240</td>
<td>0.1375</td>
<td>-0.0589</td>
</tr>
<tr>
<td>Diagnosis: non-endogenous</td>
<td>-0.0970</td>
<td>0.1509</td>
<td>0.5203</td>
<td>-0.3927</td>
</tr>
<tr>
<td>Design: observational</td>
<td>0.3078</td>
<td>0.2141</td>
<td>0.1559</td>
<td>-0.1158</td>
</tr>
<tr>
<td>Dose: not fixed</td>
<td>-0.0522</td>
<td>0.1641</td>
<td>0.7504</td>
<td>-0.3738</td>
</tr>
</tbody>
</table>

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.