

Analysis

Practising evidence-based medicine in an era of high placebo response: number needed to treat reconsidered

Steven P. Roose, Bret R. Rutherford, Melanie M. Wall and Michael E. Thase

Summary

The number needed to treat (NNT) statistic was developed to facilitate the practice of evidence-based medicine. Placebo was assumed to be therapeutically inert when the NNT was originally conceived, but more recent data for conditions such as major depressive disorder (MDD) suggest that the placebo control condition can have considerable therapeutic effects. Complications arise because the NNT calculated from randomised controlled trials (RCTs) reflects a comparison between medication plus clinical management and placebo plus clinical management, whereas, in the clinical setting, physicians choose between prescribing open medication, observing a patient over time with a supportive approach, and doing nothing. Thus, NNTs derived from clinical trials are not directly relevant to clinical decision-making, because they are based on control conditions that do not exist in standard practice. Additional difficulties may arise when using NNTs to compare alternative treatments for MDD, such as medication and psychotherapy, since these comparisons require the control conditions upon which the respective NNTs are based to be similar. Whereas pill placebo conditions include intensive clinical management and elicit expectations of improvement, attention control conditions for psychotherapy research are less well developed. Often the effects of

psychotherapy are gauged against a wait-list control condition, which has substantially fewer therapeutic components than a pill placebo control condition. To improve the clinical utility of NNTs for the treatment of MDD, we advocate effectiveness studies that include treatment conditions resembling actual clinical practice, rather than using placebo-controlled RCTs for this purpose. Until such studies are performed, the effect of bias in comparing NNTs across treatments can be controlled by ensuring that the RCT control conditions upon which the NNTs are based are comparable.

Declaration of interest

S.P.R. has served as a consultant to Lundbeck. B.R.R. has received honoraria from MedAvante and Pfizer and served as a consultant to RCT Logic. M.E.T. has acted as an advisor/consultant to AstraZeneca, Bristol-Myers Squibb Company, Cerecor, Inc., Eli Lilly & Company, Forest Laboratories, Johnson & Johnson (includes Ortho-McNeil Pharmaceuticals), Merck and Co. Inc., Novartis, Otsuka, Pfizer, and Shire US Inc.

Copyright and usage

© The Royal College of Psychiatrists 2016.

The practice of evidence-based medicine requires clinicians to integrate their personal experience and expertise with the results of systematic research studies. Randomised controlled trials (RCTs) are the primary source of evidence with respect to treatments, and statistical tools such as effect sizes and ‘number needed to treat’ (NNT) statistics were developed to facilitate clinicians’ comparisons of results across studies and the application of these results to their practice. This integration of systematic research into clinical decision-making has been of great benefit to patients over the past 25 years, but temporal trends in research practices and study results necessitate ongoing revision of how evidence-based medicine is integrated into clinical practice.

In this analysis, we describe one such example, in which rising placebo response rates in RCTs over the past 30 years might influence how practising clinicians and other interested stakeholders interpret and apply the results of these studies. We discuss this problem using the example of major depressive disorder (MDD), which has become perhaps the paradigmatic illustration of a disorder for which increasing placebo response is the dominant issue in RCTs. In short-term RCTs of antidepressants in adults with depression, for example, placebo response has risen steadily over the past 30 years, concurrent with a decline in drug *v.* placebo difference.^{1,2} Studies conducted in children and adolescents with MDD typically observe even higher placebo response rates (mean of 46% compared with a mean medication response of 59%), which have also been increasing over time.³ Although our discussion is focused on MDD, we believe the

arguments considered are germane to the practice of evidence-based medicine in any disorder for which clinical trials have consistently demonstrated a high placebo response rate.

The conceptual basis of NNT

In a highly influential article, Laupacis *et al* reviewed several methods by which practising clinicians could ‘measure and compare the benefits and risks of various preventive, diagnostic, therapeutic, rehabilitative approaches.’⁴ These ‘yardsticks’ were intended to help clinicians transform the otherwise often difficult-to-interpret results of RCTs, considered the most valid data on which to base measurements of benefit and harm for treatments, into information directly applicable to the clinical situation. The ideal yardstick, according to Laupacis and colleagues, would have four attributes: it would (a) compare the consequences of doing nothing with the potential benefits of doing something; (b) summarise the potential harm of a treatment; (c) identify patients who are at high risk of an adverse event and who are likely to be responsive to therapy; and (d) permit comparisons between different treatment approaches.⁴

The metric most closely approximating this combination of properties was deemed to be the NNT. The NNT is defined mathematically as the reciprocal of the absolute risk reduction (ARR). For a therapeutic intervention, the ARR is calculated by subtracting a post-treatment measurement of disease severity in the treatment being studied from measured disease severity in the comparison group.

$$\text{ARR} = \text{disease severity in the control or comparison group} \\ - \text{disease severity in treated group} \\ \text{NNT} = 1/\text{ARR}$$

For example, in an antidepressant clinical trial in which the response rate in the medication group was 50% and the response rate in the placebo group was 35%:

$$\text{NNT} = 1/(0.5 - 0.35) = 6.67$$

or 7, since it is customary to round the NNT to a whole number.

As this example suggests, the most easily interpretable NNTs are calculated from objectively defined outcome data that are categorical, such as death *v.* survival or stroke *v.* no stroke.⁴ In many psychiatric conditions (e.g. MDD), categorical outcomes such as response *v.* non-response do exist, but difficulties may arise in other conditions (e.g. schizophrenia) for which there is no clear consensus on the definition of response. It is sometimes necessary to derive NNTs from continuous data (e.g. a change score on an outcome measure), such as for older studies that do not report response and remission rates as well as for meta-analyses that report summary mean difference statistics between interventions and controls. Complex statistical methods have been developed for these cases to permit calculation of NNTs from continuous outcome data, and in many reports these derived NNTs seem to be valid across a range of treatment response definitions.⁵ However, these approaches may require patient-level data from clinical trials to which most clinicians do not have access, and their underlying assumptions may not be valid in all conditions. For the practicing clinician, these considerations tend to limit the applicability of the NNT to disorders that have clinically meaningful, categorical outcomes and studies that report such data.

The NNT tells the clinician the number of patients one would expect to treat with intervention A to have one more success (or one fewer failure) than if the same number were treated with intervention B (either a comparator treatment or an inactive control). As stated in the original article by Laupacis *et al.*, a 'number needed to be treated of 11 means that 10 of 11 patients either do not need therapy or will not respond to it'.⁴ An NNT of 2 indicates that a treatment is very effective and that this effect will be readily apparent to the clinician. By contrast, an NNT of 100 reflects a very small treatment effect that would contribute little to everyday clinical practice (unless, of course, that one patient could be identified prospectively). However, a particular NNT value does not in itself establish whether a treatment is clinically important, and the NNT alone should not be used to make comparative judgements across areas of therapeutics. Variables such as the prevalence of the illness being treated, the acuity and severity of the illness, and the time-frame of observation are critical. For example, despite having a very low NNT of 2, an anti-emetic medication relieving nausea for 2 h may be less clinically important than an anticoagulant medication used to prevent death from stroke, even if the latter had an NNT of 100.

Similarly, the number needed to harm (NNH) is a variation that determines the number of patients the clinician needs to treat with intervention A compared with B before they will see one additional case of a specified severe adverse event.⁶ In order to make an informed assessment of the risk-benefit ratio of an intervention, both the NNT and NNH must be considered. For example, a ratio is sometimes computed for these two numbers, with higher ratio values indicating greater relative benefit.

Application of NNT: is antidepressant treatment worthwhile for this patient?

The meaning of the NNT for any treatment is always expressed relative to a control condition. In the original formulation, the

control condition was pill placebo and, since placebo is inactive, the NNT was explicitly interpreted as informing clinicians about the consequences of prescribing a treatment as compared with doing nothing. Laupacis and colleagues argued that a benefit of the ARR (and thus, the NNT) as compared with relative risk reduction is that the ARR and NNT constructs are 'expression[s] of the consequences of giving no treatment'.⁴ The notion of comparing a treatment with doing nothing is very relevant for clinicians, since this is often the decision confronting them in their offices. Do I prescribe aspirin for this patient at risk of myocardial infarction or not?

However, it must be noted that this interpretation of the NNT for a placebo-controlled trial (i.e. as an expression of the consequences of prescribing a treatment compared with doing nothing) is based on the assumption that the placebo intervention is tantamount to doing nothing – that placebo response is roughly equivalent to natural history (the course an illness would take if undisturbed by treatment). In the original examples illustrating the utility of the NNT concept, placebo response probably was a reasonable approximation of doing nothing. For example, participants in clinical trials of aspirin to prevent myocardial infarction received aspirin or placebo. Placebo effects do not seem to exist when the outcome is mortality caused by myocardial infarction, and participants in these studies typically do not receive other potentially therapeutic elements that are standard parts of the intensive follow-up comprising the placebo condition in RCTs for treatment of MDD (e.g. weekly clinic visits, advice about diet and exercise, feedback about lipid levels and blood pressure measurements).

By contrast, a substantial proportion of the therapeutic benefit of participating in an RCT evaluating a novel antidepressant is conveyed by the so-called nonspecific elements of treatment. In fact, 60–80% of benefit can be explained by nonspecific factors.^{1,9} RCTs in psychiatric disorders may be more influenced by these nonspecific effects because symptom reports are more subjective (e.g. depressive symptom reports) and physiological pathways seem to exist by which placebo effects may influence the target disorder.¹⁰ For example, taking a pill believed to represent an effective treatment may induce expectancies in depressed patients that are capable of causing both side-effects and therapeutic effects.¹¹ Besides the expectancy effects generated in placebo-controlled RCTs for MDD, the study pills are administered as part of a multi-component clinical management program that typically includes weekly sessions with a clinician, self-reports, structured ratings, medical evaluation, and considerable contact with caring and helpful clinic staff.¹²

Increasing evidence suggests this therapeutic contact may also be a potent contributor to symptomatic improvement in patients with depression, particularly patients in the placebo arm of a medication *v.* placebo RCT.^{13,14} In adolescents with depression, a greater number of visits during the acute treatment period is associated with significantly increased placebo response but not medication response.¹⁵ Similarly, an analysis of antidepressant clinical trials enrolling adults aged 18–65 years found a cumulative and positive therapeutic effect of additional follow-up visits on placebo response, but the effect of this increased therapeutic contact was approximately 50% less in the medication groups.¹⁴ Analysis of response rates in clinical trials of antidepressants in older patients demonstrated that there is a significant interaction between study visits and treatment assignment (OR = 0.89, $t = -2.186$, d.f. = 36, $P = 0.035$), such that each additional visit over the grand mean for the sample increased average placebo response by 2.5%, but did not significantly affect medication response.¹⁶

Thus, increasing the number of study visits significantly increases placebo response while leaving medication response

generally unaffected. The effect of this arm-assignment/visits interaction is to dramatically decrease average medication *v.* placebo differences with increasing visit frequency. For example, in a 12-week, placebo-controlled study of antidepressants for late-life depression, providing 6 clinic visits resulted in an average medication response rate of 51.7% and placebo response rate of 39.5%.¹⁷ Providing 8 clinic visits for the same duration of study decreased the average medication response rate to 50.7% and increased placebo response to 44.4%, while providing 10 clinic visits resulted in a medication and placebo response of 49.7% and 49.3%, respectively. Thus, intensifying supportive care from 6 to 10 visits over 12 weeks resulted in a reduction of the average medication/placebo difference from 12.2% to 0.4%.

The weekly visits that are typical in RCTs are far more frequent than are commonly practised in the community treatment of depression. Claims-based analyses of the follow-up received by patients filling a new prescription for an antidepressant report that patients with depression have an average of 2.4 to 4.5 face-to-face visits with a physician in the 12 weeks after starting an antidepressant (i.e. they are seen approximately monthly).^{17–19} Most patients (73.6%) are treated exclusively by their general medical provider, as opposed to a psychiatrist,²⁰ fewer than 20% of patients have a mental healthcare visit in the first 4 weeks after starting an antidepressant, and fewer than 5% of adults beginning treatment with an antidepressant have as many as 7 physician visits in their first 12 weeks on the medication. These differences partially explain why results of RCTs are not always generalisable to standard clinical treatment:^{18,19} the placebo response in an RCT is enhanced by the frequent clinic visits. If RCTs included a frequency of patient visits similar to that of clinical practice, this would decrease the placebo response while not significantly affecting the medication response. Thus, far from representing the case of doing nothing in clinical practice, the placebo arm in an antidepressant RCT represents an intensive form of clinical management that has therapeutic effects.

Another difference between the exemplar RCTs originally used to illustrate the utility of the NNT (e.g. aspirin to prevent myocardial infarction, warfarin for the prevention of stroke) and RCTs for disorders such as MDD is that placebo responses are extremely variable.¹ Most of the variability in drug–placebo differences in antidepressant RCTs can be attributed to fluctuation in placebo response rather than medication response.³ For example, in a recent meta-analysis of antidepressant RCTs, placebo response ranged from 13% to 56% (mean 34.7%), and NNTs for antidepressants calculated from individual RCTs within this data-set ranged from 2 to 25.¹³ As a result, even when medication response seems to be relatively constant across studies, the clinical estimate of the importance of medication treatment changes dramatically on the basis of placebo response rates.

Insofar as the placebo condition significantly diverges from doing nothing, clinicians' application of the NNT from a given RCT to their practice becomes problematic. For example, the clinical question for a practicing psychopharmacologist might be construed as 'would it be worthwhile to prescribe an antidepressant for this individual with mild MDD?' The options available are (a) prescribe an antidepressant or (b) not prescribe any medication and follow up with the patient at a later date. To decide between these options, the psychopharmacologist might consult the literature, where they would find an average medication response rate from placebo-controlled RCTs for mild depression of approximately 45–50% and a placebo response rate of 30–35%, translating into NNTs of between 10 and 20.²¹ Controversially, it has been suggested that these NNTs are high considering the nature of the benefit (i.e. moderate symptom improvement), the potential for side-effects, and the cost

associated with the treatment.²² Thus, the psychopharmacologist may decide not to prescribe treatment, judging the 10% response rate difference between medication and placebo to be too small in light of the cost and possibility of side-effects.

In our view, this treatment decision is confounded by a misinterpretation of the NNT construct. The NNT from an antidepressant RCT compares prescribing a medication plus intensive clinical management to prescribing a placebo plus intensive clinical management. The options available in the clinical situation, including to do nothing, do not exist in these RCTs, and for the reasons outlined above, there is reason to think that placebo plus intensive clinical management is substantially more therapeutic for patients with depression than doing nothing. Whereas response rates reported for an acute course of placebo plus clinical management are 30–40%,^{1,2} the expected response rate for an acute depressive episode without treatment over 8–12 weeks' follow-up is approximately 10%.²³ The true difference between the alternative treatment options available in the clinical situation, therefore, is more likely to be:

$$45\% \text{ medication response} - 10\% \text{ spontaneous remission} \\ = 35\% \text{ (NNT} = 3)$$

rather than:

$$45\% \text{ medication response} - 35\% \text{ placebo response} \\ = 10\% \text{ (NNT} = 10).$$

In fact, the NNT that best represents the clinical situation would be calculated using medication response rates from comparator trials, which are significantly higher than the response rates in placebo controlled trials.¹³ Higher response rates may occur in comparator trials because these patients know they are receiving active medication as opposed to possibly receiving placebo, and this situation is more generalisable to clinical practice. A medication response rate of 60% from a comparator trial yields an NNT of 2 (where the control is 10% spontaneous remission).

It is therefore misleading to apply an NNT derived from an RCT, in which an expectation of improvement is induced and frequent study visits are administered for both medication- and placebo-treated subjects, to a clinical setting, in which placebo plus clinical management is not an available treatment option. To inform the clinician about the relative effectiveness of treatments in a real-life clinical setting, the NNT for an antidepressant or antipsychotic should be calculated from data from a study that compares open medication with a no-treatment control group. Unfortunately, no such studies have been done in the past, and to propose such a study design now would raise significant feasibility and, more importantly, ethical concerns. Thus, while placebo-controlled RCTs are the gold standard for demonstrating the efficacy of an intervention, they have serious, possibly fatal limitations when used to guide clinical decision-making in disorders with high placebo response rates.

Another application of NNT: which of the available treatments should be prescribed?

A second clinical application of the NNT is to help clinicians decide between alternative interventions. Among the treatments available for a given condition, which would be the most useful to prescribe for this patient? A clinician might choose by simply comparing the NNTs calculated from clinical trials of each intervention and selecting the intervention with the lowest NNT. This would be cogent when choosing between antidepressants because in an RCT comparing medications all patients receive the same management, with the only variable being medication assignment.

However, the situation is different if the clinician wants to compare antidepressants with psychotherapeutic modalities. Taking the treatment of mild MDD as an example, the NNTs in RCTs for adults with mild MDD range from 10 to 25, with an average of approximately 16.²¹ By contrast, much lower NNTs have been reported for the psychotherapeutic treatment of MDD. A series of influential meta-analyses have reported NNTs for cognitive-behavioural therapy (CBT) and other psychotherapeutic treatments for MDD in the 2–5 range.²⁴ On the basis of this substantially lower NNT, a clinician might reasonably opt to prescribe an evidence-based psychotherapy over medication.

In fact, guidelines for the treatment of mild MDD, based substantially on NNT comparisons, state that psychotherapy should be the first-line treatment for this condition. For example, National Institute for Health and Care Excellence (NICE) guidelines advise that practitioners ‘not use antidepressants routinely to treat persistent sub-threshold depressive symptoms or mild depression because the risk–benefit ratio is poor’.²⁵ Others have made stronger statements, suggesting the continued high rates of antidepressant prescriptions for mild to moderate MDD in the face of these unfavorable NNT numbers reflects improper motives on behalf of physicians and/or pharmaceutical companies.²²

While evidence-based psychotherapies are inarguably a valuable component of the clinician’s armamentarium in the treatment of MDD, confusion in the interpretation of NNTs from medication and psychotherapy trials makes blanket statements in favor of psychotherapy difficult to justify. As discussed previously, the NNT is calculated relative to a control condition. It follows that, in order to compare NNTs across different treatment interventions (e.g. antidepressants and CBT), the control conditions upon which the NNTs are based must be comparable. As stated in a recent commentary by Garcia, ‘to directly compare NNTs one needs to ensure that . . . the control or comparison groups to which the treated group was compared were equivalent. The need for similar control groups is no small feat in psychotherapy research where there is not a readily available analog to the pill placebo conditions that are frequently used in psychopharmacology research’.²⁶

In the case of an antidepressant RCT, the placebo control condition replicates every element of the medication condition except for the actual pharmacological effect of medication (e.g. serotonin reuptake inhibition). By contrast, no comparable placebo control condition exists for psychotherapy, partly because the specific ‘active ingredients’ of psychotherapy are unknown. Instead, a psychotherapy typically is compared with the effects of being placed on a treatment waiting list (or other attention control condition), receiving a different active treatment, or a placebo pill.

A control condition such as a wait-list has substantially fewer therapeutic components than a pill placebo control condition. Subjects on a wait-list do not usually have frequent clinic visits and are not masked to the fact that they are not receiving an active treatment (and probably, therefore, have minimal expectancy of improvement), both of which stand in stark contrast to the treatment received by individuals in a pill placebo cell. Indeed, a recent meta-analysis reported that the relative benefit of psychotherapies *v.* their control conditions (and the NNT calculated) depends greatly upon the nature of the control condition.²⁷ Thus, there is good reason to believe that comparisons between NNTs for CBT *v.* wait-list control and for antidepressants *v.* pill placebo are seriously confounded. Such comparisons would be biased in favour of psychotherapy over medication, given the much less effective control conditions in the psychotherapy trials.

The problem of mismatched control conditions is not necessarily avoided even when the treatments are administered

within the same study. The Treatment for Adolescents with Depression Study (TADS) randomly allocated adolescents with MDD to CBT alone, fluoxetine alone, combined CBT and fluoxetine, or pill placebo.²⁸ The authors state that ‘participants and all study staff remained masked in the ‘pills only’ conditions (fluoxetine therapy and placebo) . . . [but] patients and treatment providers in the combination and CBT conditions were aware of treatment assignment’.²⁸ Thus, the study compared blinded treatments (fluoxetine and pill placebo) with un-blinded treatments (CBT and combined CBT/fluoxetine). It has been reported that comparisons of openly administered treatments with double-blind treatments are significantly biased against the blinded treatment.¹³ The most clinically informative and fair study design would compare psychotherapy, open medication, and no treatment in individuals with mild MDD. However, no such study has been performed.

The point of the above discussion is not to advocate antidepressants as the treatment of choice for mild MDD, but rather to illuminate the difficulties posed by comparing treatments that do not have comparable control conditions. While NNTs can be calculated for any treatment and control condition, the validity of between-treatment comparisons and therefore of treatment decisions based on NNTs depends on careful verification that the controls upon which each NNT is based are comparable.

Suggestions for improving the clinical utility of NNT

After reviewing the difficulties faced by clinicians who are using NNTs calculated from RCTs to practice evidence-based medicine, the question naturally arises: what can be used instead? The answer is far from clear. The most obvious solution is to conduct studies that more closely resemble the treatment options available to doctors and patients in clinical practice. For example, a study comparing openly administered antidepressants and monthly clinic visits (the average follow-up visit frequency in community treatment²²) with periodic follow-up without pills (i.e. doing nothing) would more closely replicate the clinical situation and would permit the calculation of an NNT for antidepressants that would better guide clinical decision-making. Treatment development could shift to a two-stage model in which an intervention is first tested for efficacy in a placebo-controlled RCT, and then tested in an effectiveness study to inform clinicians how different therapeutic options would work in their clinical practice.

Unfortunately, effectiveness studies are difficult to execute and perhaps even more difficult to fund in today’s research environment, so alternatives must be explored. Although all of the available alternatives may have significant drawbacks, in many cases they offer more valid clinical guidance than NNTs calculated from standard, placebo-controlled RCTs. One possibility would be to use information from psychotherapy studies utilising a wait-list control group as a means of estimating the option of doing nothing for MDD. We analysed the acute symptom change in wait-list control groups and found that individuals with MDD experience an average improvement of 4 Hamilton Rating Scale for Depression points (Cohen’s *d* effect size=0.5) over a mean follow-up of 10 weeks.²³ This figure could be construed as a type of ‘historical control’ and used in place of a placebo control condition to estimate a modified NNT for an active treatment condition in a different RCT. The obvious downside to this course is that there may be multiple confounds, making comparisons across different patient samples invalid.

Some steps can be taken to attenuate the problems faced by clinicians using the NNT to decide between different treatment

options. Clinicians can reduce the effect of bias in their NNT comparison by ensuring the comparability of the control conditions. These conditions should be similar to one another in terms of blinding (i.e. do not compare open *v.* blinded), credibility to participants, and the amount of therapeutic contact (e.g. visits) provided. The control condition should also be appropriate to the treatment being assessed. Attention to these considerations can enhance the validity of using the NNT to compare treatments.

Conclusions

In many medical fields, but particularly in psychiatry, long-standing concerns have been expressed about how applicable the results of RCTs are to clinical practice. Although the NNT has proved itself a useful tool in the practice of evidence-based medicine in some circumstances, it is perhaps less useful with RCTs, which have complex control and comparator conditions and high, variable placebo responses. The purpose of this analysis has been to raise the field's awareness of certain difficulties in the proper interpretation of NNTs derived from RCTs having high placebo response rates and illustrate how NNTs may be misleading in certain circumstances.

Steven P. Roose, MD, **Bret R. Rutherford**, MD, **Melanie M. Wall**, PhD, Columbia University College of Physicians and Surgeons, New York State Psychiatric Institute, New York, USA; **Michael E. Thase**, Departments of Psychiatry, Perelman School of Medicine of the University of Pennsylvania, Philadelphia Veterans Affairs Medical Center and University of Pittsburgh Medical Center, USA.

Correspondence: Dr Steven P. Roose, Columbia University College of Physicians and Surgeons, New York State Psychiatric Institute, 1051 Riverside Drive, Box 98, New York, NY 10032, USA. Email: spr2@cumc.columbia.edu

First received 7 Jan 2015, final revision 27 May 2015, accepted 3 Jun 2015

Funding

This work was supported by National Institute of Mental Health grants K23 MH085236 (BRR) and T32 MH015144 (SPR).

References

- Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002; **287**: 1840–7.
- Dunlop BW, Thase ME, Wun CC, Fayyad R, Guice-Pabia CJ, Musgnung J, et al. A meta-analysis of factors impacting detection of antidepressant efficacy in clinical trials: the importance of academic sites. *Neuropsychopharmacology* 2012; **37**: 2830–6.
- Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA. Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry* 2009; **166**: 42–9.
- Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; **318**: 1728–36.
- Furukawa TA, Akechi T, Wagenpfeil S, Leucht S. Relative indices of treatment effect may be constant across different definitions of response in schizophrenia trials. *Schizophrenia Res* 2011; **126**: 212–9.
- Sedgwick P. What is number needed to harm (NNH)? *BMJ* 2013; **347**: f4869.
- Kraemer HC, Morgan GA, Leech NL, Gliner JA, Vake JJ, Harmon RJ. Measures of clinical significance. *J Am Acad Child Adolesc Psychiatry* 2003; **42**: 1524–9.
- Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry* 2006; **59**: 990–6.
- Rutherford BR, Pott E, Tandler JM, Wall MM, Roose SP, Lieberman JA. Placebo response in antipsychotic clinical trials: a meta-analysis. *JAMA Psychiatry* 2014; **71**: 1409–21.
- Mayberg HS, Silva JA, Brannan SK, Tekell JL, Mahurin RK, McGinnis S, et al. The functional neuroanatomy of the placebo effect. *Am J Psychiatry* 2002; **159**: 728–37.
- Rutherford BR, Marcus SM, Wang P, Sneed JR, Pelton G, Devanand D, et al. A randomized, prospective pilot study of patient expectancy and antidepressant outcome. *Psychol Med* 2013; **43**: 975–82.
- Miller MD, Frank E, Reynolds CF. The art of clinical management in pharmacologic trials with depressed elderly patients: lessons from the Pittsburgh Study of Maintenance Therapies in Late-Life Depression. *Am J Geriatr Psychiatry* 1999; **7**: 228–34.
- Rutherford BR, Sneed JR, Roose SP. Does study design affect outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychother Psychosom* 2009; **78**: 172–81.
- Posternak MA, Zimmerman M. Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials. *Br J Psychiatry* 2007; **190**: 287–92.
- Rutherford BR, Sneed JR, Tandler J, Peterson BS, Roose SP. Deconstructing pediatric depression trials: an analysis of the effects of expectancy and therapeutic contact. *J Am Acad Child Adolesc Psychiatry* 2011; **50**: 782–95.
- Rutherford BR, Tandler J, Brown P, Sneed JR, Roose SP. Clinic visits in late-life depression trials: effects on signal detection and therapeutic outcome. *Am J Geriatr Psychiatry* 2014; **22**: 1452–61.
- Valenstein M, Eisenberg D, McCarthy JF, Austin KL, Ganoczy D, Kim HM. Service implications of providing intensive monitoring during high risk periods for suicide among VA patients with depression. *Psych Services* 2009; **60**: 439–44.
- Morrato EH, Libby AM, Orton HD, deGruy FV, Brent DA, Allen R, et al. Frequency of provider contact after FDA advisory on risk of pediatric suicidality with SSRIs. *Am J Psychiatry* 2008; **165**: 42–50.
- Stettin GD, Yao J, Verbrugge RR, Aubert RE. Frequency of follow-up care for adult and pediatric patients during initiation of antidepressant therapy. *Am J Manag Care* 2006; **12**: 453–61.
- Mojtabai R, Olfson M. National patterns in antidepressant treatment by psychiatrists and general medical providers: results from the National Comorbidity Survey Replication. *J Clin Psychiatry* 2008; **69**: 1064–74.
- Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010; **303**: 47–53.
- Kirsch I, Sapirstein G. Listening to prozac but hearing placebo: a meta-analysis of antidepressant medication. *Prev Treat* 1998; **1**: Article 2a.
- Rutherford BR, Mori S, Sneed JR, Pimontel MA, Roose SP. Contribution of spontaneous improvement to placebo response in depression: a meta-analytic review. *J Psychiatr Res* 2012; **46**: 697–702.
- Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J Consult Clin Psychol* 2008; **76**: 909–22.
- National Institute for Health and Care Excellence. *Depression: The Treatment and Management of Depression in Adults (Update) (CG 90)*. NICE, 2009.
- Garcia AM. What does 'work' mean? Reopening the debate about clinical significance. *Clin Psychol Sci Pract* 2010; **17**: 48–51.
- Huang AX, Delucchi K, Dunn LB, Nelson JC. A systematic review and meta-analysis of psychotherapy for late-life depression. *Am J Geriatr Psychiatry* 2015; **23**: 261–73.
- March J, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, et al. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents with Depression Study (TADS) randomized controlled trial. *JAMA* 2004; **292**: 807–20.