Interpreting Estimates of Treatment Effects

Implications for Managed Care

Stephen V. Faraone, PhD

INTRODUCTION

How should health care professionals choose among the many therapies claimed to be efficacious for treating specific disorders? The practice of evidence-based medicine provides an answer. Advocates of this approach urge health care professionals to base treatment choices on the best evidence from systematic research on both the efficacy and adverse effects of various therapeutic alternatives. Ideally, health care professionals would compare different treatments by referring to randomized, double-blind, head-to-head trials that compared the treatment options. Although individual medications are typically well researched when these placebo-controlled studies are performed, studies that directly compare treatments are rare. In the absence of direct head-to-head trials, other evidence comes from indirect comparisons of two or more therapies by examining individual studies involving each treatment.

This article provides an introductory review of methods of such indirect comparisons of therapies across studies, provides examples of how these methods can be used to make treatment decisions, and presents a general overview of relevant issues and statistics for readers interested in understanding these methods more thoroughly.

STATISTICAL SIGNIFICANCE VERSUS MAGNITUDE OF EFFECT

Before one considers the meaning of a treatment effect, it is necessary to document that the effect is "statistically significant" (i.e., the effect observed in a clinical trial is greater than what would be expected by chance). If a treatment effect is not larger than that expected by chance, the "magnitude of effect" computed from the trial is questionable if one is making comparative therapeutic choices. Sometimes a small trial suggests a large benefit, but the result might not be statistically significant because the study is underpowered. In that case, the apparent large benefit should be viewed cautiously and should be considered when one is designing future studies aimed at replicating the finding.

When the results of clinical trials are statistically significant, treatment choices should not be made based on comparisons of statistical significance, because the magnitude of statistical significance is heavily influenced by the number of patients studied. Therefore, a small trial of a highly effective

therapy could have a statistically significant result that is smaller than a result from a large trial of a modestly effective treatment.

Although the results of statistical analyses provide crucial information, the magnitude of statistical significance does not necessarily indicate the magnitude of the treatment effect. As such, it is impossible to determine from the degree of statistical significance how, for example, a novel therapy evaluated in one study compares with the efficacy of other established or emerging treatments for the same condition.

This problem of interpretation of statistical significance can be addressed if we use the concept of magnitude of effect, which was developed to allow clinically meaningful comparisons of efficacy between clinical trials. The magnitude of an effect can help clinicians and P&T committee members decide whether the often modest increases in efficacy of newer therapies are important enough to warrant clinical or formulary changes. One way to make these determinations is to examine acceptable effects of widely recognized therapies for specific disorders. If this approach is not taken, comparing two clinical trials can be difficult. As the name suggests, an effect magnitude estimate places an interpretable value on the direction and magnitude of an effect of a treatment. This measure of effect can then be used to compare the efficacy of the therapy in question with similarly computed measures of effect of a treatment's efficacy in other studies that use seemingly noncomparable measures.

When indirect comparisons are conducted, measures of effect magnitude are essential in order to make sensible evaluations. For example, if one study measured the efficacy of a therapy for back pain using a five-point rating scale for pain intensity and another study used a 10-point rating scale, we could not compare the results, because a one-point decrease has a different meaning for each scale. Even if two studies use the same measure, we cannot simply compare changed scores between treatment and placebo, because these studies may differ in their standards for precision of measurement. These problems of differing scales of measurement and differences in precision of measurement make it difficult to compare studies. Fortunately, these problems can be overcome if we use estimates of effect magnitude, which provide the difference in improvement between therapy and placebo, adjusted for the problems that make the statistical significance level a poor indicator of treatment efficacy.

Although using measures of effect magnitude to indirectly

Dr. Faraone is Director of the Medical Genetics Research Center and of Child and Adolescent Psychiatric Research at SUNY (State University of New York) Upstate Medical University in Syracuse, New York. He is also a Professor of Psychiatry in the Department of Psychiatry and Behavioral Sciences.

compare therapies is helpful, this method has some limitations. Bucher et al.[1] presented an example of comparing sulfamethoxazole–trimethoprim (Bactrim, Women First/Roche) with dapsone/pyrimethamine for preventing *Pneumocystis carinii* in patients with human immunodeficiency virus (HIV) infection. The indirect comparison using measures of effect magnitude suggested that the former treatment was much better. In contrast, direct comparisons from randomized trials found a much smaller, nonsignificant difference.

Song et al.[2] examined 44 published meta-analyses that used a measure of effect magnitude to compare treatments indirectly. In most cases, results obtained by indirect comparisons did not differ from results obtained by direct comparisons. However, for three of the 44 comparisons, there were significant differences between the direct and the indirect estimates.

Chou et al.[3] compared initial highly active antiretroviral therapy (HAART) with a protease inhibitor (PI) against a nonnucleoside reverse transcriptase inhibitor (NNRTI). The investigators conducted a direct meta-analysis of 12 head-to-head comparisons and an indirect meta-analysis of six trials of NNRTI-based HAART and eight trials of PI-based HAART. In the direct meta-analysis, NNRTI-based regimens were better than PI-based regimens for virological suppression. By contrast, the indirect meta-analyses showed that NNRTI-based HAART was worse than PI-based HAART for virological suppression.

From these studies, although it seems reasonable to conclude that indirect comparisons usually agree with the results of head-to-head direct comparisons, nevertheless when direct comparisons are lacking, the results of indirect comparisons using measures of effect magnitude should be viewed cautiously. Many variables, including the quality of the study, the nature of the population studied, the setting for the intervention, and the nature of the outcome measure, can affect the apparent efficacy of treatments. If these factors differ between studies, indirect comparisons may be misleading.

The magnitude of effect can be computed in a couple of ways:

1. *Relative* measures express the magnitude of effect in a manner that clearly indicates the relative standings of the two treatments being considered. This method results in comparative statements such as "the improvement rate for treatment X is five times the improvement rate for treatment Y."

2. *Absolute* measures express the magnitude of effect without making such comparative statements. Instead, they define a continuous scale of measurement and then place the observed difference on that scale. For example, a simple absolute measure is the difference in improvement rates between two groups.

Ideally, both relative and absolute measures would be reported by treatment studies.

EFFECT MAGNITUDE: ABSOLUTE MEASURES

Although absolute measures of effect seem to be associated with a straightforward interpretation, these measures can be misleading if the baseline rates of outcomes are not taken into account.[4] For example, let's suppose we know that a therapy

doubles the probability of a successful outcome. The absolute effect of the treatment depends on the baseline (or control) probability of a successful outcome. If it is low, say 1%, the therapy increases successful outcomes by only one percentage point to 2%, a fairly small increase in absolute terms. In contrast, if the baseline rate of success is 30%, the treatment success rate is 60%, a much large increase in absolute terms.

For continuous variables, one simple approach to computing an absolute measure is the *weighted mean difference*, which is created by pooling results of trials that have used the same outcome measure in a manner that weights the results of each trial by the size of the trial. The weighted mean difference is readily interpretable, because it is on the same scale of measurement as the clinical outcome measure. The problem with using this method is that different trials typically use different outcome measures even when they are focused on the same concept.

Standardized Mean Difference and Cohen's *d*: Effect Size Measurement

The standardized mean difference (SMD) measure of effect is used when studies report efficacy in terms of a continuous measurement, such as a score on a pain-intensity rating scale. The SMD is also known as Cohen's *d*.[5]

The SMD is sometimes used interchangeably with the term "effect size." Generally, the comparator is a placebo, but a similar calculation can be used if the comparator is an alternative active treatment. The SMD is easy to compute with this formula:

$$SMD = \frac{\textit{new treatment improvement} - \textit{comparator (placebo) improvement}}{\textit{pooled standard deviation}}$$

(1)

The pooled standard deviation (SD) in Equation 1 adjusts the treatment-versus-placebo differences for both the scale and precision of measurement and the size of the population sample used.

An SMD of zero means that the new treatment and the placebo have equivalent effects. If improvement is associated with higher scores on the outcome measure, SMDs greater than zero indicate the degree to which treatment is more efficacious than placebo, and SMDs less than zero indicate the degree to which treatment is less efficacious than placebo. If improvement is associated with lower scores on the outcome measure, SMDs lower than zero indicate the degree to which treatment is more efficacious than placebo and SMDs greater than zero indicate the degree to which treatment is less efficacious than placebo.

Examination of the numerator of Equation 1 shows that the SMD increases as the difference between treatment and placebo increases, which makes sense. Less intuitive is the meaning of the denominator. Because the SD assesses the precision of measurement, the denominator of Equation 1 indicates that the SMD increases as the precision of measurement increases.

The SMD is a point estimate of the effect of a treatment. Calculation of 95% confidence intervals (CIs) for the SMD can facilitate comparison of the effects of different treatments.

Interpreting Estimates of Treatment Effects

When SMDs of similar studies have CIs that do not overlap, this suggests that the SMDs probably represent true differences between the studies. SMDs that have overlapping CIs suggest that the difference in magnitude of the SMDs might *not* be statistically significant.

Cohen[5] offered the following guidelines for interpreting the magnitude of the SMD in the social sciences: small, SMD = 0.2; medium, SMD = 0.5; and large, SMD = 0.8.

Probability of Benefit

The probability-of-benefit (POB) statistic was originally proposed by McGraw and Wong.[6] They called it the *common language effect size statistic*, which they defined as the probability that a randomly selected score from one population would be greater than a randomly selected score from another. For treatment research, this becomes the probability that a randomly selected treated patient would show a level of improvement exceeding that of a randomly selected placebo patient.

To calculate the POB for continuous data, we first compute:

$$Z = \frac{SMD}{\sqrt{2}}$$

(2)

This *Z* statistic is normally expressed as a standard normal distribution, and the POB is computed as the probability that a randomly selected standard normal variable is less than *Z*. The POB is equivalent to the area under the receiver operating characteristic curve and is also proportional to the Mann–Whitney statistic comparing drug and placebo outcome scores.[7,8] Further details and prior applications of the method are available elsewhere.[9–15]

For binary variables, the POB statistic can be computed from the absolute difference (AD) in treatment response as follows: POB = 0.5(AD+1).

AD is computed by subtracting the percentage of patients who improved with placebo from the percentage who improved with treatment.
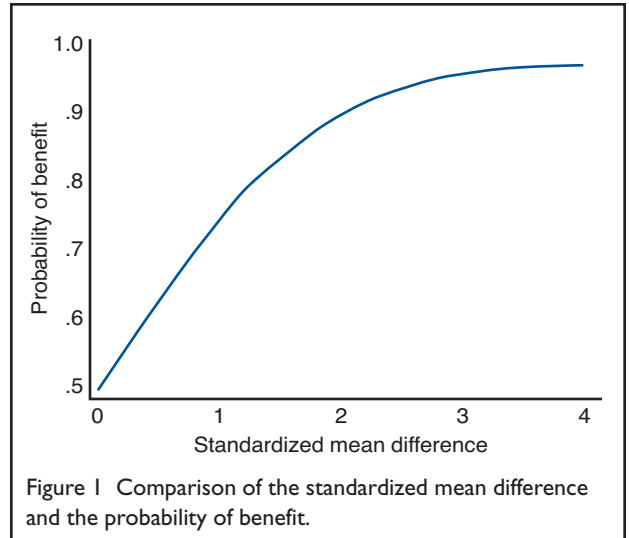
Using the formula in Equation 2, we can show how the POB corresponds to varying values of the SMD. Stata Corporation's statistical software was used to make the computations.[16] Figure 1 presents the results.

When the SMD equals zero, the probability that treatment outperforms placebo is 0.5, which is no better than the flip of a coin. When the SMD equals 1, the probability that treatment outperforms placebo is 0.76. The incremental change in the probability of benefit is small for SMDs greater than two.

Number Needed to Treat

Another statistic that has a straightforward interpretation is the number needed to treat (NNT), or the number of patients who must be treated to prevent one adverse outcome. The poor outcome would be not showing improvement in symptoms. The NNT is computed as follows:

$$NNT = \frac{100}{percent\ improved\ with\ treatment - percent\ improved\ with\ placebo}$$

(3)

Figure 1 Comparison of the standardized mean difference and the probability of benefit.

In Equation 3, the denominator of the NNT is the absolute difference, which is also used to compute the POB for binary variables. As Kraemer and Kupfer showed, NNT can be defined for continuous variables if we assume that a treated patient has a successful outcome if the outcome is better than a randomly selected nontreated patient.[10] We can then compute:

$$NNT = \frac{1}{2\ (POB - 1)}$$

(4)

Using these formulas in Equations 2 and 4, we can show how the POB corresponds to varying values of the NNT. Stata's software was used to make the computations.[16] Figure 2 presents the results.

As the SMD approaches zero, the NNT becomes very large. When the SMD = 1, the NNT = 2. As with the POB, the incremental change in the NNT is small for SMDs greater than two.

Ideally, the NNT would be very small. For example, when NNT = 2, for every two patients who are treated, one will have a good outcome. The NNT exceeds one for effective therapies

Figure 2 Comparison of the standardized mean difference and the number needed to treat.
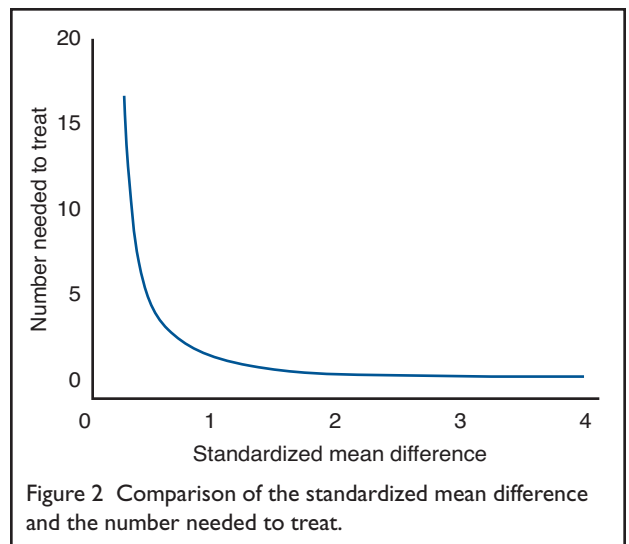
and is negative if a treatment is harmful. The main advantage of the NNT is its straightforward clinical interpretation. It also can be easily used to weigh the costs and benefits of treatment. In this framework, the cost of achieving one successful treatment is not only the amount incurred for the patient who improved; the cost also includes the amount incurred in treating patients who do not improve. Therefore, if the cost of a single therapy is C, the cost of achieving one successful treatment is $C*NNT$. If the costs of two therapies are C_1 and C_2 and their NNTs are NNT_1 and NNT_2, the relative cost of the first treatment, compared with the second, is:

$$\frac{C_1 * NNT_1}{C_2 * NNT_2}$$

(5)

As Equation 5 indicates, doubling the NNT has the same effect as doubling the cost of the treatment.

EFFECT MAGNITUDE: RELATIVE MEASURES
Relative Risk

A simple measure of effect magnitude would be to use the percentage of patients who improve in the treated group as a comparative index of efficacy in different studies; however, this would be wrong. Although statistics involving percentages of improvement are easy to understand, they cannot be interpreted meaningfully without referring to the percentage of improvement observed in a placebo group, especially when improvement is defined in an arbitrary manner such as a 30% reduction in a symptom outcome score.

One solution to this problem is to express drug–placebo differences in improvement as the relative risk for improvement, which is computed as the ratio of the percentage of patients who improved in the treatment group divided by the percentage of patients who improved in the placebo group:

$$relative\ risk\ for\ improvement = \frac{percentage\ improved\ with\ treatment}{percentage\ improved\ with\ placebo}$$

(6)

Although it seems odd to view improvement as a "risk," the term is widely used in the sense of probability rather than risk of a negative outcome. Relative risk is easily interpreted: it is the number of times more likely a patient is to improve with treatment compared with placebo. Therefore, a relative risk for improvement of 10 would mean that 10 treated patients improved for every untreated patient.

Odds Ratio

As its name indicates, the odds ratio is computed as the ratio of two odds: the odds of improvement with treatment and the odds of improvement with placebo. The formula is:

$$odds\ ratio = \frac{odds\ of\ improvement\ with\ treatment}{odds\ of\ improvement\ with\ placebo}$$

(7)

In Equation 7, the odds of improving with the drug are the ratio of the percentage of patients improved with treatment to the percentage not improved with treatment. The odds of improvement with placebo are computed in a similar manner. The odds ratio indicates the increase in the odds for improve-

ment that can be attributed to the treatment. An odds ratio can also have an associated CI to allow one to decide on the reliability of the comparison.

Because thinking about "percentage improvement" is more intuitive than thinking about "odds of improvement," most people find it easier to understand and interpret the relative risk for the improvement statistic compared with the odds ratio. However, relative risk has an interpretive disadvantage, which is best explained with an example.

Suppose a new treatment has a 40% improvement rate compared with a standard treatment (at 10%). The relative risk for improvement is 40/10, so the new treatment seems to be four times better than the standard treatment; however, the treatment failure rates would be 60% for the new treatment and 90% for the standard treatment. The relative risk for failure is 90/60, which indicates that the old treatment is 1.5 times worse than the new treatment.

As this example shows, estimating relative risk depends on whether one examines improvement or failure to improve. In contrast, this problem of interpretation is not present for the odds ratio.

DISCUSSION

Table 1 summarizes measures of magnitude of effect. Even though using a measure of effect magnitude to compare the efficacy of different treatments is clearly an advance beyond qualitative comparisons of different studies, it would be a mistake to compare magnitudes of effect between studies without acknowledging the main limitation of this method. The computation of effect measures makes sense only if we are certain that the studies being compared are reasonably similar in any study design features that might increase or decrease the effect size. The usefulness of comparing measures of effect between studies is questionable if the studies differ substantially on design features that might plausibly influence differences between drug and placebo.

For example, if a study of drug A used double-blind methodology and found a smaller magnitude of effect than a study of a drug B that was not blinded, we could not be sure whether the difference in magnitude of effect was a result of differences in drug efficacy or differences in methodology. If endpoint outcome measures differ dramatically among studies, that could also lead to spurious results.

As another example, if one study used a highly reliable and well-validated outcome measure and the other used a measure of questionable reliability and validity, comparing measures of effect would not be meaningful. Meta-analyses of effect measures need to either adjust for such differences or to exclude studies with clearly faulty methodology.

Effect magnitude can be useful in making treatment and formulary decisions, but clinicians and managed care organizations must consider whether superiority of effect for a treatment reported from one or more studies would apply to all types of studies. The randomized, controlled trials from which effect measures are often derived might not reflect real-world conditions of clinical practice. For example, patients enrolled in randomized, controlled trials typically are subject to inclusion and exclusion criteria that would affect the clinical composition of the sample.

Interpreting Estimates of Treatment Effects

Table 1 Measures of Magnitude of Effect

Effect Size	Meaning of Values	Interpretation
Relative measures		
Relative risk for improvement (RR)	1: no difference Less than 1: placebo better than drug Greater than 1: drug better than placebo	The number of times more likely a patient is to improve with a drug compared with placebo
Odds ratio (OR)	1: no difference Less than 1: placebo better than drug Greater than 1: drug better than placebo	The increase in the odds for improvement that can be attributed to the treatment
Absolute measures		
Standardized mean difference (SMD)	0: no difference Negative: placebo better than drug Positive: drug better than placebo	Gives the difference between drug and placebo outcomes adjusted for measurement scale and measurement imprecision; can be translated into the probability of treated patients improving more than the average placebo patient
Probability of benefit	0.5: no difference 0–0.49: placebo better than drug 0.51–1.0: drug better than placebo	The probability that a randomly selected member of the drug group will have a better result than a randomly selected member of the placebo group
Number needed to treat (NNT)	Negative: placebo better than drug Positive: drug better than placebo	The number of patients, on average, who need to be treated for one patient to benefit from treatment

By understanding the different measures and how they are computed, health care professionals can make sense of the literature that uses these terms; perhaps more importantly, they can compute some measures that they find particularly useful for decision-making. Thinking in terms of magnitude of effect becomes particularly useful when effect sizes for other treatments are understood.

Table 2 presents examples of measures of effect magnitude from meta-analyses of the psychiatric literature. The measures clearly show that some disorders can be treated more effectively than others. For example, the stimulant therapy for attention-deficit/hyperactivity disorder (ADHD) has a much lower NNT than the antidepressant treatment of generalized anxiety disorder. The table also shows differences among medications for the same disorder. The nonstimulant treatment of ADHD has nearly twice the NNT as the stimulant therapy for this disorder.[17–22]

The magnitude of effect is important when we consider the costs of various treatment options. One approach is to consider the costs of failed treatments. Based on the NNT, we can compute the probability of a failed treatment as:

$$\frac{NNT-1}{NNT}$$

(8)

## Table 2 Magnitude-of-Effect Measures for Classes of Psychiatric Medications

Medication Class	Relative Measures		Absolute Measures		
	Relative Risk for Improvement	Odds Ratio	SMD	Probability of Benefit	Number Needed to Treat
Nonstimulants for ADHD	2.5	3.1	.62	.67	7
Immediate-release stimulants for ADHD	3.6	5	.91	.74	4
Long-acting stimulants for ADHD	3.7	5.3	.95	.75	4
SSRIs for obsessive-compulsive disorder or depression	2.2	2.5	.5	.64	9
Atypical antipsychotic drugs* for schizophrenia	1.5	1.6	.25	.57	20
Antidepressants for generalized anxiety disorder	1.9	2.1	.39	.61	12

ADHD = attention-deficit/hyperactivity disorder; SMD = standardized mean difference; SSRI = selective serotonin reuptake inhibitor.

* Olanzapine, quetiapine, risperidone, and sertindole.

Data from Gale C, Oakley-Browne M. *Clin Evid* 2002(7):883–895;[17] Faraone SV, et al. *Medscape General Medicine e Journal.* 2006;8(4):4;[18] Faraone SV. *Medscape Psychiatry Mental Health* 2003;8(2);[19] Geddes J, et al. *Clin Evid* 2002(7):867–882;[20] Soomro GM. *Clin Evid* 2002(7):896–905;[21] and Leucht S, et al. *Schizophr Res*1999;35(1):51–68.[22]

Interpreting Estimates of Treatment Effects

Figure 3 Number of wasted treatments (in thousands) and the number needed to treat.
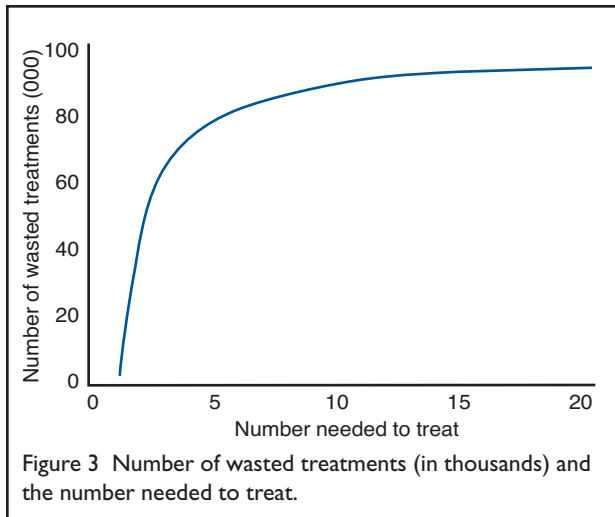
Now let's consider a health care system that treats 100,000 patients annually. Because we can compute the probability of a treatment failure, we can easily compute the number of wasted treatments as a function of the NNT. The results in Figure 3 indicate that treatment wastage increases dramatically as the NNT increases from one to five. For NNTs greater than five, treatment wastage is very high, but incremental increases in wastage are small with incremental increases in the NNT.

Because effect magnitude measures have implications for wasted treatments, they also have implications for costs. If the costs of two treatments are identical, the treatment that leads to more wasted treatments would be more costly. If the treatments differ in cost, we would need to translate treatment wastage into treatment costs by multiplying the amount of waste by the cost.

CONCLUSION

Taking an evidence-based approach to patient care requires that health care professionals consider adverse events along with efficacy when they select treatments for their patients. This discussion of efficacy does not diminish the importance of the many other questions health care providers must consider when choosing treatments, such as:

- Is the patient taking other treatments?
- Does the patient have a coexisting disorder that suggests the need for combined pharmacotherapy?
- Has the patient had a prior trial with any of the potential treatments? If so, what were the effects?

These and other questions remind us that even though quantitative methods, such as the computation of effect magnitude, play a crucial role in the practice of evidence-based medicine, they will never fully replace the intellectual expertise of compassionate and informed health care professionals.

REFERENCES

1. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50(6):683–691.
2. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: Empirical evidence from published meta-analyses. *BMJ* 2003;326(7387):472.
3. Chou R, Fu R, Huffman LH, Korthuis PT. Initial highly-active antiretroviral therapy with a protease inhibitor versus a non-nucleoside reverse transcriptase inhibitor: Discrepancies between direct and indirect meta-analyses. *Lancet* 2006;368(9546):1503–1515.
4. Cook RJ, Sackett DL. The number needed to treat: A clinically useful measure of treatment effect. *BMJ* 1995;310(6977):452–454.
5. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
6. McGraw KO, Wong SP. A common language effect size statistic. *Psychol Bull* 1992;111:361–365.
7. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143(1):29–36.
8. Colditz GA, Miller JN, Mosteller F. Measuring gain in the evaluation of medical technology: The probability of a better outcome. *Int J Technol Assess Health Care* 1988;4(4):637–642.
9. Faraone SV, Biederman J, Spencer TJ, Wilens TE. The drug–placebo response curve: A new method for assessing drug effects in clinical trials. *J Clin Psychopharmacol* 2000;20(6):673–679.
10. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry* 2006;59(11):990–996.
11. Faraone SV, Biederman J, Spencer T, Wilens TE. Efficacy of atomoxetine in adult attention-deficit/hyperactivity disorder: A drug–placebo response curve analysis. *Behav Brain Functions* 2005;1(1):16.
12. Faraone S, Short E, Biederman J, et al. Efficacy of Adderall and methylphenidate in attention deficit hyperactivity disorder: A drug–placebo and drug–drug response curve analysis of a naturalistic study. *Int J Neuropsychopharmacol* 2002;5(2):121–129.
13. Faraone SV, Pliszka SR, Olvera RL, et al. Efficacy of Adderall and methylphenidate in attention-deficit/hyperactivity disorder: A drug–placebo and drug–drug response curve analysis. *J Child Adolesc Psychopharmacol* 2001;11(2):171–180.
14. Short EJ, Manos MJ, Findling RL, Schubel EA. A prospective study of stimulant response in preschool children: Insights from ROC analyses. *J Am Acad Child Adolesc Psychiatry* 2004;43(3):251–259.
15. Tiihonen J, Hallikainen T, Ryynanen OP, et al. Lamotrigine in treatment-resistant schizophrenia: A randomized placebo-controlled crossover trial. *Biol Psychiatry* 2003;54(11):1241–1248.
16. *Stata User's Guide: Release 9*. College Station, TX: Stata Corporation, LP; 2005.
17. Gale C, Oakley-Browne M. Generalised anxiety disorder. *Clin Evid* 2002(7):883–895.
18. Faraone SV, Biederman J, Spencer TJ, Aleardi M. Comparing the efficacy of medications for ADHD using meta-analysis. *Medscape General Medicine e Journal* 2006;8(4):4.
19. Faraone SV. Understanding the effect size of ADHD medications: Implications for clinical care. *Medscape Psychiatry Mental Health* 2003;8(2).
20. Geddes J, Butler R. Depressive disorders. *Clin Evid* 2002(7):867–882.
21. Soomro GM. Obsessive compulsive disorder. *Clin Evid* 2002(7):896–905.
22. Leucht S, Pitschel-Walz G, Abraham D, Kissling W. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo: A meta-analysis of randomized controlled trials. *Schizophr Res* 1999;35(1):51–68. ■