

trained community care workers: part therapist, part housing officer, part prescriber. Just what such a real world mental health trooper might look like has been described in detail.<sup>4</sup> The battles that they face will be remote from the world depicted in the latest NICE guidance.

**Competing interests:** None.

1. **National Institute for Health and Clinical Excellence.** Core interventions in the treatment and management of schizophrenia in primary and secondary care (update). Clinical Guidelines CG82, issued March 2009. <http://www.nice.org.uk/guidance/index.jsp?action=byID&o=11786> (accessed 29 July 2009).

2. **Kendell R.** The distinction between personality disorder and mental illness. *Br J Psychiatry* 2002;**180**:110–15.
3. **Killaspay H,** Bebbington P, Blizard R, *et al.* The REACT study: randomised evaluation of assertive community treatment in north London. *BMJ* 2006;**332**:815–20.
4. **Salter M,** Turner T. *Community mental health care: a practical guide to outdoor psychiatry.* Edinburgh: Churchill Livingstone, 2008.

## Are all antidepressants equal?

G Gartlehner,<sup>1,2</sup> B N Gaynes<sup>3</sup>

Primary care physicians and psychiatrists manage the majority of patients suffering from acute phase major depressive disorder (MDD). For most patients, antidepressant treatment is the primary choice of care. Second generation antidepressants (SGAs)—developed following the first generation of tricyclic and monoamine oxidase agents—have become the preferred drug choice because of their greater tolerability, lower risk of lethality and similar efficacy compared with first generation agents.

Clinicians prescribing SGAs face a multitude of drug choices and are the target of extensive marketing campaigns by the pharmaceutical industry. In 2007, three of the 20 top selling drugs in the USA were antidepressants with annual sales ranging from \$2.3 billion (venlafaxine XR (Effexor XR)) to \$1.4 billion (duloxetine (Cymbalta)).<sup>1</sup> At the time of writing, 13 different SGAs have been approved for the treatment of major depression in the USA and Canada and two additional drugs (reboxetine, milnacipran) are available in some European countries. Some of these drugs are now available as generic medications, others are still patent protected. Economically, drug choice matters. The US Consumers Union found that in 2008 the average monthly costs of treatment with second generation antidepressants in the USA varied from \$20 to \$400 depending on the medication of choice.<sup>2</sup> The study

conducted by Cipriani and colleagues<sup>3</sup> addressed an ongoing challenge for clinicians: how to choose among antidepressant treatments and select the best drug for an individual patient suffering from an episode of major depression (*see page 107*).

### WHY SUCH A STUDY IS IMPORTANT

The burning question for patients and clinicians is whether differences in costs are substantiated by differences in benefits and harms. Evidence based data to answer this question and to guide selection of treatments has been limited. Because of the lack of available direct head to head comparisons, prior systematic reviews have been able to say little about differences among the medications.<sup>4,5</sup> The ideal studies to fill this knowledge gap would be large, pragmatic, randomised controlled trials (RCTs) that directly compare the benefits and harms of SGAs. Unfortunately, out of the more than 70 possible comparisons among SGAs, barely more than half have been investigated in RCTs.<sup>6</sup> Many of the available studies are small or have methodological problems that severely limit the ability to draw firm conclusions about the comparative efficacy and safety.

In a complex statistical analysis, Cipriani and colleagues<sup>3</sup> compared response rates and discontinuation rates of individual SGAs and offered a simple answer: sertraline (Zoloft) and escitalopram (Lexapro) are better than other SGAs followed by venlafaxine (Effexor) and mirtazapine (Remeron).<sup>3</sup> Is this the long sought for answer for clinicians who treat patients with MDD? An accompanying editorial in *The Lancet* “Antidepressants are not all created equal” was jubilant and asserted that “...a new gold standard of reliable information has been compiled...”<sup>7</sup> Subsequent letters to the editor, however,

were less enthusiastic but rather outright critical about the methods and conclusions of this study.<sup>8–13</sup> The underlying tenor of the critics was that Cipriani *et al* failed to acknowledge the methodological limitations of the approach and, by ranking drugs, evoked an unsubstantiated sense of precision based on evidence that is fraught with uncertainty. A closer look at the underlying methods is necessary to understand the controversy.

### STUDY DESIGN AND METHODS

The objective of the study was to provide a clinically useful summary of the effects of SGAs that can be used to guide treatment decisions.<sup>3</sup> To detect relevant studies, authors conducted systematic literature searches in the Cochrane Collaboration depression, anxiety and neurosis trials register. Furthermore, they contacted pharmaceutical companies, regulatory agencies and study investigators to acquire unpublished or missing data. The primary outcomes were response (defined as a 50% improvement of the baseline score) and treatment discontinuation rates at 8 weeks.

Because not all antidepressants have been compared directly in RCTs, Cipriani and colleagues<sup>3</sup> used a statistical technique called multiple treatments meta-analysis to derive estimates of the comparative efficacy and safety for all possible comparisons among SGAs. Such an approach essentially combines results from trials directly comparing two or more SGAs with estimates of treatment effects based on common comparators. For example, if two drugs exhibit a similar treatment effect relative to a common comparator, the conclusion would be that these drugs have similar efficacy.

In the absence of direct comparisons, such an approach is legitimate and findings can allow inferences about the relative benefits (or harms) of drugs that have never been compared directly. Nevertheless, results have to be interpreted cautiously. Such a statistical approach is commonly viewed as observational evidence, even if the statistical

<sup>1</sup> Danube University, Department for Evidence-based Medicine and Clinical Epidemiology, Krems, Austria;

<sup>2</sup> Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, USA;

<sup>3</sup> Department of Psychiatry, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, USA

**Correspondence to:** Dr G Gartlehner, Danube University, Department for Evidence-based Medicine and Clinical Epidemiology, Krems, Austria; gerald.gartlehner@donau-uni.ac.at

model includes exclusively RCTs as component studies.<sup>14</sup> Results are not as valid and reliable as those from conventional meta-analyses of RCTs, which are generally regarded as the best available evidence to assess treatment effects.

Caution is warranted because the validity of results of indirect comparisons depends on various assumptions, some of which are unverifiable. The key assumption behind any indirect comparison is that the populations between the two sets of trials are similar with respect to prognosis, severity of disease and important confounders.<sup>14</sup> For instance, it would not be legitimate to compare the effect of paroxetine in young adults with severe acute phase MDD with the effects of sertraline in frail, elderly patients with minor depression. These two populations have different prognoses, different spontaneous remission rates (ie, control event rates) and different risks for adverse events. Concluding that differences in response and remission rates are entirely attributable to differences in efficacy between paroxetine and sertraline would not make sense.

The similarity of populations is only one factor to be considered when conducting indirect comparisons. In addition, study designs have to be similar, drug dosages have to be comparable, treatment effects have to be measured using the same outcomes and follow-up times of patients have to be alike. Differences in study designs are particularly relevant when newer agents are compared with existing agents.<sup>15</sup>

If these key assumptions are not met, just as with any observational study, bias and confounding will inevitably be introduced and distort results.

## RESULTS OF THE MIXED TREATMENT META-ANALYSIS

Overall, authors included 117 RCTs published between 1991 and 2007 with data involving almost 26 000 patients. About two-thirds of participants were women. Findings indicated that mirtazapine, venlafaxine and sertraline were statistically significantly more efficacious than duloxetine, fluoxetine, fluvoxamine, paroxetine and reboxetine (odds ratios varied from 1.22 to 2.03). Taking discontinuation rates into consideration, escitalopram and sertraline presented the best risk-benefit profiles, followed by venlafaxine and mirtazapine. Escitalopram and sertraline had statistically significantly fewer discontinuations than duloxetine, fluvoxamine, paroxetine, reboxetine and venlafaxine.

## METHODOLOGICAL SHORTCOMINGS OF THE APPROACH

Cipriani and colleagues<sup>3</sup> have invested a great amount of time and diligence into researching, extracting and collating data of this large body of evidence. Their willingness to provide online access to the dataset, which guarantees a degree of transparency that is rarely seen in systematic reviews, is also applaudable. They have also used a cutting edge statistical model to overcome the lack of head to head evidence. Nevertheless, we believe that various deficiencies in the inclusion of component studies and choices of outcome measures exist that limit the credibility of their results and conclusions.

For example, the primary outcome of their study was response to treatment, defined as the proportion of patients who had a reduction of at least 50% from baseline on the Hamilton Depression Rating Scale (HAM-D) or Montgomery-Asberg Depression Rating Scale (MADRS) or who scored much improved or very much improved on the Clinical Global Impression (CGI) at 8 weeks. To produce valid results in indirect comparisons of response rates, the essential assumption would be that a response on one scale equals a response on the other scales. We know, however, that the convergent validities among these scales are not perfect. Pearson's correlations between HAM-D and MADRS range from 0.68 to 0.88 and between HAM-D and CGI from 0.56 to 0.77 (a Pearson correlation of 1 would resemble perfect agreement).<sup>16</sup> Although these numbers are generally viewed as adequate for clinical use, they will introduce uncertainties when results of these scales are compared.

Several other factors contribute to the uncertainty of estimates in the study of Cipriani and colleagues.<sup>3</sup> For instance, they included studies with very different populations such as frail elderly, patients with accompanying anxiety and inpatients as well as outpatients. Presumably, these populations differed substantially in severity of disease and prognosis. Moreover, studies with rigid methods were mixed with others characterised by less internal validity such as single blinded trials or studies with high dropout rates. As is well known, these studies tend to overestimate treatment effects.<sup>17</sup> Although these issues should not preclude indirect comparisons, they do indicate that findings should be interpreted carefully.

Finally, the effect measure of choice was odds ratios rather than relative risks. Odds ratios have mathematical advantages that statisticians value.

Practitioners, however, frequently overestimate their clinical importance because they tend to interpret odds ratios just as relative risks. Given the high event rates for response in antidepressant trials, odds ratios provide substantially larger values than relative risks.<sup>18, 19</sup> This factor thus risks overstatement and overinterpretation of the differences reported by Cipriani and colleagues.<sup>3</sup>

The ranking of drugs was not based on the comparative efficacy alone. This study also assessed subjects' discontinuation rates and used them as proxies for acceptability. Overall discontinuation rates are not an adequate measure of tolerability or safety because remission or lack of efficacy can also be causes of dropout and mask important differences in adverse events. For example, a recent meta-analysis has shown that the overall discontinuation rates between selective serotonin reuptake inhibitors (SSRIs) as a class and venlafaxine were similar (relative risk (RR) 1.10; 95% confidence interval (CI) 0.96 to 1.26).<sup>20</sup> A more detailed look, however, revealed that discontinuation rates because of adverse events were statistically significantly higher in patients treated with venlafaxine than with SSRIs (RR 1.42; 95% CI 1.15 to 1.75). This difference was attributable primarily to higher rates of nausea and vomiting caused by venlafaxine compared with SSRIs (34% vs 22%). Thus overall discontinuation rates are a questionable choice for determining the balance between benefits and harms, especially harms that matter most to patients. Given that specific harms differ demonstrably across drugs (eg, sertraline has substantially higher rates of diarrhoea than other second generation antidepressants<sup>21</sup>), the benefit-harm trade-offs are far more complex than *The Lancet* article conveys.

## AUTHORS' CONCLUSIONS

The authors concluded that clinically important differences exist among SGAs favouring escitalopram and sertraline over other antidepressants. They deduced that sertraline might be the best choice when starting treatment for moderate to severe depression because it has the most favourable balance between benefits and acceptability, and costs.

## COMMENTS ON AUTHORS' CONCLUSIONS

The conclusions and endorsement of sertraline as the best choice to start antidepressive therapy are the most problematic parts of this paper. Although

Cipriani and colleagues<sup>3</sup> have conducted a comprehensive and thorough analysis of SGAs, in their conclusions they have vastly overstated the underlying evidence. Ranking sertraline and escitalopram higher than other drugs conveys a precision and existence of clinically important differences that is not reflected in the body of evidence. The point estimates of differences for most comparisons are clinically irrelevant and, more importantly, are fraught with uncertainties. As the authors graphically display in the paper, for sertraline and escitalopram the range of probabilities actually extends from the first to the eighth rank for both efficacy and acceptability.

Pfizer (the maker of Zoloft) and Forest Laboratories (the maker of Lexapro) must have been delighted by these conclusions. These results mean millions of pounds, dollars, and euros in additional sales and their marketing machines are probably distributing this article at full speed. Many clinicians will take the results at face value because the manuscript was published in *The Lancet*, a highly respected, high impact journal. Furthermore, it was accompanied by a strangely uncritical editorial<sup>7</sup> singing the praises of these findings.

What does all this mean for practising clinicians? To date, the general consensus has been that the efficacy and effectiveness of second generation antidepressants are very similar. The American College of Physicians Guideline for the treatment of MDD suggests that physicians and patients select among treatment options based on known differences of adverse events and costs.<sup>22</sup> After reviewing the study by Cipriani and colleagues<sup>3</sup> and considering the methodological limitations, the American College of Physicians decided not to change this assessment.

In fact, Cipriani and colleagues<sup>3</sup> were not the first investigators to determine the comparative efficacy of second generation antidepressants with the means of indirect comparisons. In November 2008 we published a study with similar but not

identical methods in the *Annals of Internal Medicine* and drew a different conclusion: benefits do not differ materially across the drugs but differences in adverse events exist.<sup>21</sup> For most comparisons, differences in treatment effects were similar between the two studies; in both studies some of the comparisons rendered statistically significant differences in response rates. We simply took underlying uncertainties into greater consideration and interpreted findings more cautiously than Cipriani and colleagues.<sup>3</sup>

The challenge for primary care physicians and psychiatrists of how to select the best drug for an individual patient suffering from acute phase MDD has yet to be resolved. Given findings from the STAR\*D (Sequenced Treatment Alternatives to Relieve Depression) trial, a large pragmatic study of treatment of major depressive disorder, the desire to determine one antidepressant as the drug of first choice might be misguided after all. STAR\*D suggests that particular drugs are less important than monitoring the patient's symptoms and side effects and adjusting the regimen accordingly, including switching drugs or adding new drugs to the regimen.<sup>23</sup> Basing the choice of an antidepressant on known differences in side effects and costs is, therefore, probably still the best approach.

**Competing interests:** BNG has received consulting fees or research support from GlaxoSmithKline, Shire Pharmaceuticals, Ovation Pharmaceuticals and Pfizer, Inc. GG has no competing interests.

## REFERENCES

1. **Drugs.com.** Top 200 Drugs for 2008 by Sales. 2009 (updated 2009). Available from: [www.drugs.com/top200.html](http://www.drugs.com/top200.html) (accessed 20 July 2009).
2. **Consumer Reports.** Best Buy Drugs: Antidepressants. Available from: <http://www.consumerreports.org/health/best-buy-drugs/antidepressants.htm> (accessed 23 July 2009).
3. **Cipriani A, Furukawa TA, Salanti G, et al.** Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009;**373**:746–58.
4. **Hansen RA, Gartlehner G, Lohr KN, et al.** Efficacy and safety of second-generation antidepressants in

- the treatment of major depressive disorder. *Ann Intern Med* 2005;**143**:415–26.
5. **Williams JW Jr, Mulrow CD, Chiquette E, et al.** A systematic review of newer pharmacotherapies for depression in adults: evidence report summary. *Ann Intern Med* 2000;**132**:743–56.
  6. **Gartlehner GHR, Hansen RA, Thieda P, et al.** Comparative effectiveness of second-generation antidepressants in the pharmacologic treatment of adult depression. Rockville, MD: Agency for Healthcare research and Quality 2007 (updated 2007). Available from: [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm) (accessed 25 July 2009).
  7. **Parikh SV.** Antidepressants are not all created equal. *Lancet* 2009;**373**:700–1.
  8. **Jefferson T.** Ranking antidepressants. *Lancet* 2009;**373**:1759.
  9. **Ioannidis JP.** Ranking antidepressants. *Lancet* 2009;**373**:1759–60.
  10. **Turner E, Moreno SG, Sutton AJ.** Ranking antidepressants. *Lancet* 2009;**373**:1760.
  11. **Gartlehner G, Gaynes BN, Hansen RA, et al.** Ranking antidepressants. *Lancet* 2009;**373**:1761.
  12. **Seyringer ME, Kasper S.** Ranking antidepressants. *Lancet* 2009;**373**:1760–1.
  13. **Schwan S, Hallberg P.** Ranking antidepressants. *Lancet* 2009;**373**:1761.
  14. **Glenny AM, Altman DG, Song F, et al.** Indirect comparisons of competing interventions. *Health Technol Assess* 2005;**9**:1–134, iii–iv.
  15. **Barbui C, Cipriani A, Brambilla P, et al.** "Wish bias" in antidepressant drug trials? *J Clin Psychopharmacol* 2004;**24**:126–30.
  16. **Bagby RM, Ryder AG, Schuller DR, et al.** The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* 2004;**161**:2163–77.
  17. **Moher D, Pham B, Jones A, et al.** Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609–13.
  18. **Altman DG, Deeks JJ, Sackett DL.** Odds ratios should be avoided when events are common. *BMJ* 1998;**317**:1318.
  19. **Deeks JJ.** Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;**21**:1575–600.
  20. **Gartlehner G, Thieda P, Hansen RA, et al.** Comparative risk for harms of second-generation antidepressants: a systematic review and meta-analysis. *Drug Saf* 2008;**31**:851–65.
  21. **Gartlehner G, Gaynes BN, Hansen RA, et al.** Comparative benefits and harms of second-generation antidepressants: background paper for the American College of Physicians. *Ann Intern Med* 2008;**149**:734–50.
  22. **Qaseem A, Snow V, Denberg TD, et al.** Using second-generation antidepressants to treat depressive disorders: a clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2008;**149**:725–33.
  23. **Gaynes BN, Rush AJ, Trivedi MH, et al.** The STAR\*D study: treating depression in the real world. *Cleve Clin J Med* 2008;**75**:57–66.