The comparative efficacy of trazodone and imipramine in the treatment of depression

Scott B. Patten, MD, FRCPC

Objective: To review published clinical trials comparing the efficacy of trazodone with that of tricyclic antidepressant medication.
Data sources: MEDLINE was searched for relevant articles published from 1983 to 1991. The bibliography of a review article was searched for further references.
Study selection: In all, 25 clinical trials were found. Six of these met the methodologic assessment criteria (adapted from the McMaster guidelines for the evaluation of clinical trials), which included the stipulation of a score of 18 or more on the Hamilton depression rating scale and a 50% reduction in that score as an outcome measure.
Data extraction: All six studies compared trazodone with imipramine. Data describing response to the treatments were extracted, and post-hoc power estimates were calculated. The analysis also involved statistical tests of a modified null hypothesis, the generation of confidence intervals (CIs) and a meta-analysis.
Data synthesis: All the studies found no significant difference in the efficacy of trazodone and imipramine. However, the statistical power of most of them was less than 50% and often less than 10%; thus there was a low probability that differences would be detected. The results of statistical tests of the modified null hypothesis, inspection of the CIs and the results of the meta-analysis all suggested that trazodone and imipramine are equally efficacious.
Conclusion: The application of various techniques for the analysis of equivalence data suggests that trazodone and imipramine are of approximately equivalent efficacy. The data are compatible with small differences in efficacy, but the differences are of a magnitude such that they are unlikely to be of clinical significance.

Objectif : Examiner les essais cliniques publiés comparant l'efficacité de la trazodone à celle des antidépresseurs tricycliques.
Sources des données : L'auteur a fait des recherches dans MEDLINE pour extraire les articles pertinents publiés de 1983 à 1991. Il a examiné la bibliographie d'un article de recension pour trouver d'autres références.
Sélection des études : Au total, 25 essais cliniques ont été relevés, dont 6 répondaient aux critères d'évaluation méthodologiques (adaptés des lignes directrices de McMaster pour l'évaluation des essais cliniques) qui supposaient une cote de 18 ou plus sur l'échelle de dépression Hamilton et une réduction de 50 % de cette note en tant que mesure de résultat.
Extraction des données : Dans les six études, on comparait la trazodone à l'imipramine. L'auteur a extrait les données descriptives sur les réactions aux traitements et calculé a posteriori les estimations de force. L'analyse englobait en outre des tests statistiques d'une hypothèse nulle modifiée, la génération d'intervalles de confiance (IC), ainsi qu'une méta-analyse.

Dr. Patten is a clinical fellow, Alberta Heritage Foundation for Medical Research, University of Calgary, Calgary, Alta.

Reprint requests to: Dr. Scott B. Patten, Department of Community Health Sciences, Faculty of Medicine, University of Calgary, 3330 Hospital Dr. NW, Calgary, AB T2N 4N1

← For prescribing information see page 1235

Synthèse des données : Aucune étude n'a fait ressortir de différences importantes dans l'efficacité de la trazodone et de l'imipramine. Cependant, la force statistique, dans la plupart des cas, était inférieure à 50 % et souvent à 10 %; ainsi, la probabilité de déceler des différences était faible. Les résultats des tests statistiques de l'hypothèse nulle modifiée, l'inspection des IC et les résultats de la méta-analyse faisaient ressortir que la trazodone et l'imipramine ont la même efficacité.

Conclusion : Par diverses techniques d'analyse des données d'équivalence, on constate que la trazodone et l'imipramine ont une efficacité approximativement équivalente. Les données sont compatibles aux faibles différences d'efficacité, mais l'ampleur de ces différences fait qu'il est peu probable qu'elles aient quelque importance sur le plan clinique.

Trazodone is an antidepressant medication that differs chemically and pharmacologically from the tricyclic antidepressants. It is a triazolopyridine derivative and has a complex heterocyclic structure not shared by other antidepressants. Trazodone's side effects are generally more favourable than those of the tricyclic agents: it lacks anticholinergic activity and therefore tends to cause fewer instances of dry mouth, blurred vision, constipation and urinary retention than do tricyclic antidepressants. The most prominent side effect of trazodone is sedation. However, because depressed patients frequently suffer from insomnia this side effect can be beneficial. Trazodone is usually well tolerated; however, it blocks α-adrenergic receptors and may cause orthostatic hypotension. Also, it has been associated with priapism, a rare but very serious side effect.[1]

The generally favourable side effects of trazodone make it an attractive alternative for the pharmacotherapy of depression. However, many authors have expressed doubt that trazodone is as effective as the traditional tricyclic antidepressants.[2-5] Various clinical trials have compared trazodone's efficacy with that of tricyclic antidepressants. These studies have generally used statistical techniques oriented toward hypothesis testing and have reported no significant differences in efficacy between trazodone and the comparison medications. Such studies may be termed "active control equivalence studies," because their goal is to demonstrate that a new agent (trazodone) is as efficacious as a standard treatment (tricyclic antidepressants).[6] However, for methodologic reasons such studies may be difficult to interpret.

Active control equivalence studies are conceptually different from placebo-controlled clinical trials in that their goal is to "prove" the null hypothesis that there is no difference between treatments.[7] This is problematic, because by their nature clinical trials cannot "prove" a null hypothesis; they can only find or fail to find evidence against it. If a study reports no significant difference in the proportions of patients responding to two treatments, this means that what differences there are could have arisen by chance more than 5% of the time if the two treatments were equally effective. Clearly, this does not prove that the two treatments are equally effective.

Problems related to the interpretation of data from active control equivalence studies of trazodone have led to confusion. Almost all comparative trials have reported no differences in efficacy; however, many have had small samples, so that it is difficult to judge whether the efficacy is equivalent or whether the studies have merely failed to detect clinically important differences. This concern has been raised by previous reviewers concerned with the efficacy of trazodone: "These studies were unlikely to detect statistically significant differences in efficacy of the active drugs because of the relatively small number of patients in each group."[2]

The issue is of considerable clinical importance, because trazodone is often better tolerated than tricyclic antidepressants. If it is equally effective it may be a preferable treatment for many patients. I undertook a formal review to resolve the question of the comparative efficacy of trazodone and tricyclic antidepressants. The strategy was to use, whenever possible, methods for assessing equivalent efficacy that might supplement those reported in the original studies.

Methods for assessing equivalent efficacy

Several methods have been proposed for addressing the problems posed by equivalence trials. One is to modify the traditional approach to hypothesis testing by reformulating the null hypothesis so that its rejection by an appropriate statistical test would imply equivalent efficacy. Another is to use power estimates and confidence intervals (CIs) rather than hypothesis-testing techniques. A third is to conduct a meta-analysis.

The conventional null hypothesis is that the two compared treatments are of equivalent efficacy. A two-sided statistical test is employed to test the evidence that the data are inconsistent with this assumption. The reformulated hypothesis could state, for example, that the standard treatment is as good as or more effective (by a specified proportion) than the new treatment. The proportion would be chosen to represent the amount by which the new

treatment could be less effective than the standard treatment and still be considered equivalent for practical purposes. Rejection of this null hypothesis by means of a one-sided test would imply that the new treatment is equivalent to the existing one. One procedure for the generation of the relevant test is described by Blackwelder.[7]

Another widely supported approach to the analysis of equivalence data involves the use of CIs.[6,8] If treatment outcome is measured as the proportion of patients responding to new and established treatments, then 95% CIs can be generated for the difference in proportions responding to the drugs. Such intervals yield more information than do statistical tests of significance by providing a range of plausible values for the population parameter about which the study attempts to make inferences. In the case of differences in proportions of patients responding to two antidepressant treatments, CIs provide an estimate of how different the response rates could be and still produce the data observed in the study with a reasonable likelihood. This clearly adds a dimension to the information provided by the results of a statistical test and may provide helpful insights into the comparative efficacy of two agents.

The advantage of CIs over hypothesis testing is that whereas p values are influenced by the sample size of the study, the width of the CI provides information about the influence of sample size on the precision of the findings.

The third approach is meta-analysis. Pooling data from several studies may increase the chance of differences in efficacy between treatments being detected, particularly when the individual studies had small samples and may have lacked the power to detect real differences. Thus, failure to find evidence of such differences in a meta-analysis would provide better support for the equivalent efficacy of two drugs than would the negative findings of any one trial.

Methods

Literature review

MEDLINE was searched for articles published from 1983 to 1991. The key words "trazodone and clinical trial" and "imipramine or amitriptyline or doxepin or nortriptyline or desipramine" were used. All articles that appeared to be clinical trials were reviewed. The bibliography of the review article that was found was used to trace additional articles. Studies were included in the review if they met the following criteria, adapted from those suggested for the methodologic assessment of clinical trials by the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton:[9] (a) the study must have explicitly stated that there was random assignment to the treatment groups; (b) the study must have measured the outcome of treatment categorically, a reduction by 50% in the score on the Hamilton depression rating scale[10] (a widely used and clinically relevant measure) being taken as evidence of a response to treatment; (c) all patients had to qualify for a diagnosis of depression supported by the Feighner or research diagnostic criteria,[11] the *Diagnostic and Statistical Manual of Mental Disorders*[12] or the Symptom Profile for Depression[13] and had to have a score of 18 or more on the Hamilton scale; (d) the doses of medication had to be feasible for clinical practice (no more than 300 mg daily of imipramine or its equivalent and 600 mg daily of trazodone); and (e) all patients enrolled in the studies had to be accounted for in the analysis. Some studies did not include all the enrolled subjects in the analysis. These studies were acceptable if it was clear that such subjects dropped out for reasons unrelated to treatment. However, if the reasons were treatment related, then the studies were included only if enough information was reported to allow a repeat analysis that included the dropouts as nonresponders.

Statistical analysis

Post-hoc power estimates were calculated for the statistical tests used in each study.

The null hypothesis was redefined and statistical tests generated according to the method described by Blackwelder.[7] The reformulated hypothesis stated that the proportion responding to imipramine was equal to or greater than the proportion responding to trazodone plus 0.05. Rejection of this hypothesis would mean that the data were incompatible with differences in response rates of 0.05 or more, implying at least approximate equivalence of efficacy. For example, if the proportion of patients responding to imipramine were 0.75, then rejection of the modified hypothesis would imply that the proportion responding to trazodone was less than 0.70. Differences of 5% or less are unlikely to be of clinical relevance, so the test provides a valid assessment of equivalence.

The data on the proportions of patients responding to each drug were extracted from the studies reviewed, and 95% CIs were calculated by means of the method described by Fleiss.[14] The intervals were calculated as the rate of response to trazodone minus that to imipramine; a value of zero corresponded to no difference, a positive value corresponded to a superior rate of response to trazodone and a negative value to a superior rate of response to imipramine.

In addition, a meta-analysis was performed on

the six studies identified in the literature search. The data were arranged into two-by-two contingency tables that were combined with the use of the Mantel-Haenszel technique. A method originally described by Robins, Breslow and Greenland[15] was used to generate CIs for the Mantel-Haenszel pooled odds ratio.

Results

The MEDLINE search uncovered references to 12 clinical trials. One review article was also found that cited 15 more clinical trials, for a total of 27. Two studies appeared to be published twice in different journals, reducing the yield to 25. Six of the studies failed to state explicitly that subjects had been randomly assigned to groups, 16 failed to use a 50% reduction in the Hamilton D score as a categoric outcome measure, and 11 failed to document the diagnosis according to standard criteria or to document sufficient severity of depression. All the studies used clinically realistic doses of the medications, as defined. Eleven studies failed to account for all subjects in the analysis. In all, four of the studies were excluded from the review because of failure to meet one of the quality review criteria, seven were excluded because they failed to meet two of the criteria, and eight were excluded because they failed to meet three or more of the criteria. Thus, six studies fulfilled the criteria for this analysis.[13,16-20] All the studies compared trazodone with imipramine. The sample sizes, duration of treatment and doses of drugs used are summarized in Table 1.

One surprising aspect of the quality review was that so many articles failed to report 50% reductions in Hamilton-D scores as an outcome measure. However, only one paper was excluded from the review solely because it did not use this outcome measure.

Table 2 shows post-hoc power estimates for the categoric statistical tests comparing the proportions responding to trazodone and imipramine in the six studies. Trazodone was hypothetically assumed to be 10%, 20% or 30% less effective than imipramine. The calculations were performed by assuming that the proportion of imipramine responders in each sample was representative of the proportion of responders in the population from which that study sample was drawn. This proportion was multiplied

Table 1: Features of six clinical trials comparing the efficacy of trazodone and imipramine

Study	Year	n*	Daily dose (mg)		Treatment duration (d)
			Trazodone	Imipramine	
Gershon et al[13]	1981	263	200–600; mean 370	100–300; mean 190	21–28
Fabre et al[16]	1979	28	200–300; mean 287	Mean 140	28
Kellams et al[17]	1979	28	200–600; mean 515	100–300; mean 185	28
Trapp et al[18]	1979	30	Maximum 600	Maximum 300	21–28
Escobar et al[19]	1980	40	200–600; mean 550	100–300; mean 262.5	21–28
Feighner[20]	1980	45	200–600; mean 313	100–300; mean 159	28

*Sample size included some patients randomly assigned to a placebo group.

Table 2: Power estimates for clinical trials comparing the efficacy of trazodone and imipramine

Study	Proportion responding		Power (%) if trazodone were less effective than imipramine		
	Trazodone	Imipramine	10% less	20% less	30% less
Gershon et al[13]	43/91	57/100	12.1	34.8	65.9
Fabre et al[16]	5/9	6/10	4.5	7.5	11.5
Kellams et al[17]	5/9	2/10	3.2	4.1	5.2
Trapp et al[18]	4–6/10*	6/10	4.5	7.6	12.1
Escobar et al[19]	8/13	14/15	12.3	27.8	46.0
Feighner[20]	9/19†	5/20†	3.7	5.6	8.2

*Depending on whether a 40%, 50% or 60% reduction in Hamilton D score was used as the response criterion.
†Includes (as nonresponders) two dropouts from the group treated with trazodone and two from the group treated with imipramine.

by 0.9, 0.8 or 0.7 to generate a hypothetic value for the proportion responding to trazodone, given that it was 10%, 20% or 30% less effective than imipramine. A power formula was then applied to generate an estimate of statistical power.[21] In most cases the power of the individual studies was less than 50% (and in many cases less than 10%), implying that they had a very low probability of detecting real differences in efficacy between these two drugs. Thus, although the studies found no significant difference between the response to trazodone and that to imipramine no individual study had adequate power to validly claim that trazodone's efficacy was comparable to that of imipramine. In fact, the power estimates suggest that most of the studies may have been unlikely to find statistically significant differences in efficacy even if trazodone were 10%, 20% or 30% less likely to produce a response than imipramine.

When a z statistic was generated to test the reformulated null hypothesis[7] it was found that two studies[17,20] rejected the modified hypothesis (p = 0.026 and p = 0.034 respectively) that the proportion responding to trazodone was 0.05 (or less) smaller than that responding to imipramine. CIs for the differences in response rates are displayed in Fig. 1. The findings in each study are consistent with large differences in the proportions responding to the two drugs. However, from the group results it is apparent that all the studies are consistent with differences in efficacy of between 5% in favour of trazodone and 15% in favour of imipramine. Values outside this narrow range lie beyond the 95% confidence limits for some studies; thus the data are not compatible with large differences in efficacy but with small differences that may simply represent "noise" or random variation — in fact, no difference in efficacy may exist.

Inspection of the CIs does not suggest obvious departures from homogeneity across the data from the six studies. This impression was confirmed by use of a χ^2 test for homogeneity.[14] The χ^2 value was 3.12 (five degrees of freedom [df], p = 0.68), which is consistent with the data being homogeneous across the six studies. The studies were therefore considered suitable for a meta-analysis. From the Mantel-Haenszel test[15] there was no evidence against the null hypothesis of equivalent efficacy (Mantel-Haenszel test statistic = 0.074, df = 1, p = 0.79). The Mantel-Haenszel estimate of the pooled odds ratio was 0.93, indicating that the odds of a response to trazodone were very close to those of a response to imipramine. The 95% confidence limits associated with this ratio were 0.64 and 1.36; thus, the pooled data were consistent with an odds ratio of 0.64 to 1.36.

Even when pooled, the data from the relevant studies do not provide evidence against the null hypothesis that imipramine and trazodone are equally efficacious. The pooled odds ratio is very close to 1, which supports the assertion that the rate of response to these two agents is nearly equivalent.

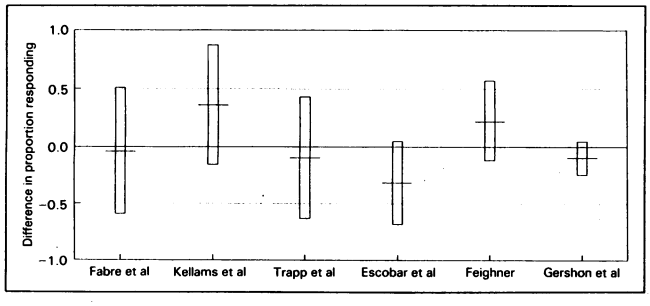

Fig. 1: Confidence intervals (95%) for proportion of patients responding to trazodone minus proportion responding to imipramine in six clinical trials.[13,16-20] The bars represent point estimates.

Conclusion

To demonstrate that two treatments are equally efficacious is more difficult than it may seem. Generally, it should not be regarded as sufficient to find "no significant difference" using a hypothesis-testing approach in a comparative clinical trial.

There are several methodologically sound studies, but the data from each one could not exclude, with reasonable probability, the possibility of a sizeable difference in efficacy, because no study had sufficient statistical power to detect such differences if they existed.

The results of a reanalysis of the published data with the use of statistical tests oriented toward equivalence data and of estimation techniques were consistent with trazodone and imipramine being approximately equivalent in efficacy. The results of a meta-analysis of the pooled data were also consistent with equivalent efficacy.

References

1. Scher M, Krieger JN, Juergens S: Trazodone and priapism. *Am J Psychiatry* 1983; 140: 1362-1363
2. Rudorfer MV, Golden RN, Potter WZ: Second-generation antidepressants. *Psychiatr Clin North Am* 1984; 7: 519-534
3. Shopsin B: Second generation antidepressants. *J Clin Psychiatry* 1980; 41: 45-56
4. Arana GW, Hyman SE: *Handbook of Psychiatric Drug Therapy*, Little, Boston, 1991: 67
5. Drugs for psychiatric disorders. *Medical Lett Drugs Ther* 1989; 31: 13-20
6. Makuch R, Johnson M: Issues in planning and interpreting active control equivalence studies. *J Clin Epidemiol* 1989; 42: 503-511
7. Blackwelder WC: "Proving the null hypothesis" in clinical trials. *Controlled Clin Trials* 1982; 3: 345-353
8. Makuch RW, Johnson MF: Some issues in the design and interpretation of "negative" clinical studies. *Arch Intern Med*

1986; 146: 986-989
9. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre: How to read clinical journals: V. To distinguish useful from useless or even harmful therapy. *Can Med Assoc J* 1981; 124: 1156-1162
10. Hamilton M: A depression rating scale. *Neurol Neurosurg Psychiatry* 1960; 23: 56-62
11. Spitzer RL, Endicott J, Robins L: Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry* 1978; 35: 773-782
12. *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed, rev, Am Psychiatr Assoc, Washington, 1987
13. Gershon S, Mann J, Newton R et al: Evaluation of trazodone in the treatment of endogenous depression: results of a multicenter double blind study. *J Clin Psychopharmacol* 1981; 1 (suppl 6): 39S-44S
14. Fleiss JL: *Statistical Methods for Rates and Proportions*, 2nd ed, Wiley, New York, 1981: 29-30
15. Robins J, Breslow N, Greenland S: Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics* 1986; 42: 311-323
16. Fabre LF, McLendon DM, Gainy A: Trazodone efficacy in depression: a double blind comparison with imipramine and placebo in day-hospital type patients. *Curr Ther Res* 1979; 25: 827-834
17. Kellams JJ, Marietta H, Klapper AB et al: Trazodone, a new antidepressant: efficacy and safety in endogenous depression. *J Clin Psychiatry* 1979; 40: 390-395
18. Trapp GA, Handorf CR, Larach V: Trazodone in the treatment of depressed inpatients. *Psychopharmacol Bull* 1979; 15: 25-27
19. Escobar JI, Gomez MSJ, Constain C et al: Controlled clinical trial with trazodone, a novel antidepressant: a South American experience. *J Clin Pharmacol* 1980; 20: 124-130
20. Feighner JP: Trazodone, a triazolopyridine derivative, in primary depressive disorder. *J Clin Psychiatry* 1980; 41: 250-255
21. Rosner B: *Fundamentals of Biostatistics*, 3rd ed, PWS Pubs, Boston, 1990: 350

Conferences

continued from page 1161

July 15-18, 1992: Society for Research into Hydrocephalus and Spina Bifida 36th Annual Scientific Meeting
Mainz, Germany
Dr. R. Bayston, Institute of Child Health, 30 Guilford St., London WC1N 1EH, England; telephone 011-44-1-071-242-9789, fax 011-44-1-071-831-0488

July 21-25, 1992: 6th International Conference on Human-Animal Interactions (hosted by the Human-Animal Bond Association of Canada and cosponsored by the International Association of Human Animal Interaction Organizations)
Palais des Congrès, Montreal
Human-Animal Bond Association of Canada, PO Box 313, Stn. B, Ottawa, ON K1P 6C4; (613) 747-0262, fax (613) 745-1846

Aug. 3-7, 1992: European Intensive Bioethics Seminar
Official language: English
Nijmegen, The Netherlands
Dr. J.V.M. Welie, Department of Ethics, Philosophy and History of Medicine, Catholic University of Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands; telephone 011-31-0-80-615320, fax 011-31-0-80-541862

Aug. 5-9, 1992: International Doctors in Alcoholics Anonymous 43rd Annual Meeting
Amway Grand Plaza Hotel, Grand Rapids, Mich.
Connie Hyde, 3311 Brookhill Circle, Lexington, KY 40502; (606) 233-0000, fax (606) 253-0862

Aug. 19, 1992: Canadian Medical Protective Association Annual Meeting (in conjunction with the 125th Annual Meeting of the CMA)
Hotel Newfoundland, St. John's
Canadian Medical Protective Association, Carling Square, 560 Rochester St., Ottawa, ON K1G 5K7; (613) 236-2100

Aug. 20-25, 1992: Canadian Society of Forensic Science Annual Conference — Truth Through Science and Integrity
Citadel Inn, Halifax
Abstract deadline is May 1, 1992.
Fredricka Monti, executive secretary, Canadian Society of Forensic Science, 215-2660 Southvale Cres., Ottawa, ON K1B 4W5; (613) 731-2096

Aug. 30-Sept. 3, 1992: European Federation of Immunological Societies John Humphrey Course on Tumour Immunology (a satellite conference to the 8th International Congress of Immunology organized by the Romanian Society for Immunology and with the sponsorship of the German Society of Immunology)
Iasi, Romania
Dr. Eugen Carasevici, Clinica de Oncologie, Laboratorul de Imunologie Tumorala, Spitalul Universitar "Sf. Spiridon," b-dul Independentei nr. 1, 6600 Iasi, Romania

Aug. 31-Sept. 2, 1992: IgA Nephropathy: the 25th Year — International Symposium
Nancy, France
Meeting Secretariat, Laboratoire d'Immunologie, BP 184, Avenue de la Forêt de Haye, 54500 Vandoeuvre-lès-Nancy, France; telephone 011-33-83-59-28-56, fax 011-33-83-44-60-22

Du 2 au 4 sept. 1992 : Conférence internationale sur l'entraide (parrainée par le Conseil canadien de développement social)
Centre des conférences du gouvernement, Ottawa
Golden Planners, 404-126, rue York, Ottawa, ON K1N 5T5

Sept. 2-4, 1992: International Conference on Self-Help/Mutual Aid (sponsored by the Canadian Council on Social Development)
Government Conference Centre, Ottawa
Golden Planners, 404-126 York St., Ottawa, ON K1N 5T5

continued on page 1202

For prescribing information see page 1253 →