Venlafaxine had higher response and remission rates than paroxetine in non-chronic treatment resistant depression

*Poirier MF, Boyer P. **Venlafaxine and paroxetine in treatment-resistant depression. Double-blind, randomised comparison.** Br J Psychiatry 1999 Jul;**175**:12–6.*

QUESTION: In patients with non-chronic treatment resistant depression, is venlafaxine as efficacious as paroxetine in achieving treatment response and remission?

Design
Randomised (unclear allocation concealment*), blinded (patients and clinicians)*, placebo controlled trial with 4 weeks of follow up.

Setting
Clinical centres in France.

Patients
123 patients (mean age 43 y, 72% women) with major depression (*DSM-III-R*) of <8 months duration. Inclusion criteria were age 18–60 years, 17-item Hamilton Depression Rating Scale (HDRS) score ≥18 and resistance (Clinical Global Impression [CGI] scale improvement score of 3 at the second treatment) to 2 antidepressant treatments for the current depression (therapeutic dose for ≥4 wk, then an alternative antidepressant at an effective dose equivalent to clomipramine 100–150 mg for ≥4 or ≥2 weeks if discontinued because of safety problems). Exclusion criteria were use of study drugs during the current episode; use of anticoagulants, phenytoin, mood stabilisers, or electroconvulsive therapy; recent use of antipsychotics or monoamine oxidase inhibitors; nonaffective disorder; suicide ideation; treatment inhibiting organic disease; seizure disorders; psychoactive substance dependence; cardiac, renal, or hepatic disease; pregnancy; or lactation. Follow up was 87% (observed case analysis).

Intervention
61 patients were allocated to venlafaxine and 62 to paroxetine. Venlafaxine was started at 75 mg/day then titrated to and maintained at 200–300 mg/day by day 10 (mean dose 269 mg/d). Paroxetine was started at 20 mg/day, then increased to and maintained at 30 or 40 mg/d by day 8 (mean dose 36 mg/d).

Main outcome measures
HDRS score at day 28 for those who completed the trial. Treatment response was a HDRS score decrease >50% plus a CGI scale improvement score of 1 or 2. Remission was a HDRS score <10.

Main results
More patients who completed the trial in the venlafaxine group achieved therapeutic response (p = 0.04) and remission (p = 0.01) than patients in the paroxetine group (table). Using the last outcome carried forward, venlafaxine showed a trend toward achievement of response (45% v 36%, p = 0.07) and had higher rates of remission (37% v 18%, p = 0.02). Because of worsening scores for some patients who received venlafaxine, the groups did not differ for change in HDRS score (decrease of 11 v 10 points, p = 0.6), or improvement on the CGI severity (73% v 84%, p = 0.4) or improvement scale (64% v 66%, {p = 0.8}†). The groups did not differ for total adverse effects.

Conclusion
Venlafaxine was more effective than paroxetine for increasing the therapeutic response and remission rates after 4 weeks of treatment in non-chronic treatment resistant depression.

*See glossary.
†p Value calculated from data in article.

Venlafaxine v paroxetine for non-chronic treatment resistant depression‡

Outcomes at 28 d	Venlafaxine	Paroxetine	RBI (95% CI)	NNT (CI)
Therapeutic response	52%	33%	59% (11 to 154)	6 (3 to 254)
Remission	42%	20%	112% (17 to 294)	5 (3 to 21)

‡Abbreviations defined in glossary; RBI, NNT, and CI calculated from data in article.

COMMENTARY

Treatment of depression has increasingly focussed on achieving full remission, because patients with residual symptoms are at substantially higher risk of relapse.[1] No antidepressant has yet been shown to be consistently superior, however, in achieving remission in either new onset or resistant depression. Arguments have been made that altering both norepinephrine and serotonin levels may produce a more robust antidepressant response.[2]

Paroxetine is a highly specific selective serotonin reuptake inhibitor (SSRI); venlafaxine has dual noradrenergic and serotonergic activity at therapeutic doses. Previous trials of venlafaxine have suggested superior efficacy over fluoxetine in initial treatment of severe depression.[3]

The type of patients studied (moderately depressed, with poor response to 2 treatments) are very commonly seen in general practice and general psychiatry clinics; the study's finding, if validated, has direct relevance for day to day practice for both generalists and specialists. Uncertainty as to the trial setting (ie, primary, intermediary, or tertiary), however, will hinder clinicians' ability to compare the patients studied with their own.

The results in this study showed greater response and remission rates in the venlafaxine group. 2 problems, however, limit the robustness of the findings. Firstly, more melancholic patients and those who were severely ill were in the paroxetine group (the difference was not statistically significant). Secondly, the cut off point for declaring remission (HDRS score <10) was higher than that usually used (HDRS score ≤7 or 8).

This study is important because it further supports the strategy of using high dose venlafaxine as a preferred antidepressant in treating refractory patients.

Lawrence Martin, MD
St Joseph's Hospital
Hamilton, Ontario, Canada

1 Fawcett J. Antidepressants: partial response in chronic depression. *Br J Psychiatry Suppl* 1994;**165**:37–41.
2 Ferrier IN. Treatment of major depression: is improvement enough? *J Clin Psychiatry* 1999;**60** Suppl 6:10–4.
3 Schatzberg AF. Antidepressant effectiveness in severe depression and melancholia. *J Clin Psychiatry* 1999;**60** Suppl 4:14–21;discussion 22.

Cluster randomisation trials in mental health research

The unit of allocation in most clinical trials is the individual patient. However, experimental trials in which the unit of allocation is an intact cluster of participants (eg, families, schools, medical practices, communities) are becoming increasingly widespread in the evaluation of healthcare and educational interventions. For example, Avorn *et al* describe the results of a health education trial (aimed at staff personnel in nursing homes) that was designed to reduce the use of psychoactive drugs by residents.[1] Six pair matched facilities were included in this trial, with 1 facility in each matched pair randomly assigned to the educational programme and the other facility serving as a control.

Several reasons exist for favouring cluster randomisation in this trial. A principal one would be to avoid the experimental contamination which could occur when the same personnel are asked to give both interventions to different participants and when knowledge of the intervention may influence the responses of participants in the control group. A second reason is that the assignment of a new educational programme to some individuals within a nursing home but not to others might be regarded as unacceptable, or even unethical, by some practitioners. Finally, having administratively set up such a programme within a facility, it would seem much more likely to function effectively from a practical perspective if all staff members, and not just some, were involved.

A notable design feature of this trial is that the 12 nursing homes recruited were pair matched on the basis of size, type of ownership, and level of drug use. The purpose of such matching was to ensure that the facilities in each pair were similar with respect to baseline drug use, but geographically distant enough to minimise the risk of experimental contamination that could arise through the sharing of knowledge. Such matching or stratification by selected baseline risk factors is a common feature of cluster randomisation trials, particularly when the total number of clusters to be randomised is small.

A second example of a recent cluster randomisation trial is given by Kinmonth *et al* who assessed the effect of additional training in a patient centred approach as directed to practice nurses and general practitioners.[2] The main outcome measures for this study included the quality of life and psychological wellbeing of patients with newly diagnosed type 2 diabetes. The trial randomised 41 practices (21 to the intervention group, 20 to the control group) in a health region in southern England, with the number of patients per practice ranging from 250–360. Because the intervention was aimed at and delivered by healthcare professionals, the most natural unit of randomisation was at the practice level. None the less, the effectiveness of the intervention was measured by recording outcomes at the level of individual patient. This disparity between the unit of randomisation and the unit of analysis is a characteristic feature of cluster randomisation trials.

Other published examples exist in which randomisation at the cluster level is the most natural choice, or even a clear necessity. This was arguably the case in the HIV prevention trial described by Grosskuth *et al*.[3] As the authors note, randomisation was necessary at the community level in this trial because the intervention involved the provision of improved services at designated health facilities, with these services available to the entire population served by each facility.

Many of these reasons would seem to apply in a natural way to trials evaluating diagnostic or therapeutic interventions that are directed to psychiatrists and other mental health professionals. In every instance, however, the rationale for adopting cluster randomisation inevitably must rest on very practical considerations. This is because cluster randomisation designs tend to be less efficient, in a statistical sense, than designs that randomise individuals to intervention groups. The loss of efficiency arises because the responses of individuals in the same cluster tend to be more similar than the responses of individuals in different clusters.

Study design issues in cluster randomisation

The degree of similarity among responses within a cluster is typically measured by a value known as the intracluster (intraclass) correlation coefficient, which usually must be estimated from the sample data. Denoted by ρ, this value may be interpreted as the standard Pearson correlation coefficient between any two responses in the same cluster. If $\rho = 0$, participants are completely independent of each other, while if $\rho = 1$ (perfect correlation), the information in a cluster is totally summarised by the response on a single cluster member. In most cluster randomisation trials, however, the value of ρ is small and positive, which is equivalent to stating that the variation among observations in different clusters exceeds the variation within clusters. Under these conditions, it is sometimes stated that the design is characterised by "between cluster variation." The underlying reasons for between cluster variation will differ from trial to trial, but in practice include the following:

(a) Participant selection, where individuals are in a position to choose the cluster to which they belong. For example, in a trial randomising physician practices, the characteristics of patients belonging to a practice could be related to age or sex differences among practitioners. To the extent that these characteristics are also related to patient response, a clustering effect will be induced within practices. In addition, as noted by Rhee *et al*, the outcomes on 2 or more patients treated by the same physician could share the influence of that physician's style of practice.[4]

(b) The influence of covariates at the cluster level, where all individuals in a cluster are affected in a similar manner as a result of sharing exposure to a common environment. As discussed by Rice and Leyland, patients attending the same hospital may share several common influences, including, for example, the same pressure to shorten length of stay.[5] In community based studies, differences in bylaws between municipalities could influence the success of smoking cessation programmes. In other studies, where intact families or households are randomised, the combined effect of both environmental and genetic factors may contribute to the observed between cluster variation.

(c) The effect of personal interactions among cluster members who receive the same intervention. For example, treatments or educational interventions provided in a group setting could lead to a sharing of information among group members that creates a clustering effect. More generally, as noted by Koepsell, just as infectious agents can be spread from person to person, the transmission of attitudes, norms, and behaviours among people who are in regular contact can result in similar responses.[6]

Without extensive empirical data, it is usually impossible to distinguish among the potential reasons for between cluster variation. Regardless of the specific cause, however, such variation invariably leads to a reduction in the effective sample size for the trial, where the size of the reduction increases with both the magnitude of ρ and the average cluster size. This in turn leads to a loss of precision in estimating the effect of intervention.

These effects of clustering can be easily expressed quantitatively. Consider an experimental trial in which k clusters, each consisting of m individuals, are randomly assigned to either an experimental or control group. We suppose that the primary aim of the trial is to compare the groups with respect to their mean values on a normally distributed response variable Y having a common but unknown variance σ^2. Estimates of the population means μ_1 and μ_2 are given by the usual sample means $\bar{Y}_1$ and $\bar{Y}_2$ for the experimental and control groups respectively. As shown by Donner et al,[7] the variance of each of these means is given by

$$Var(\bar{Y}_i) = \frac{\sigma^2}{km}[1 + (m-1)\rho] \qquad (1)$$

where ρ is the intracluster correlation coefficient. If σ^2 is replaced by P(1-P) where P denotes the probability of a success, equation (1) also provides an expression for the variance of a sample proportion under clustering.

For sample size determination, equation (1) implies that the usual estimate of the required number of individuals in each group should be multiplied by the variance inflation factor (or design effect) $IF = 1 + (m-1)\rho$ to provide the same statistical power as would be obtained by randomising km individuals to each group when there is no clustering effect. It should also be noted that small values of ρ accompanied by large values of m can considerably inflate the required sample size for a trial. For example, if $\rho = 0.01$ and medical practices of average size 1000 are randomised to each of 2 intervention groups, then the total number of participants required under cluster randomisation is more than 10 times that required under individual randomisation.

Challenges of applying cluster randomisation

Many of the methodological challenges of cluster randomisation arise because inferences are usually intended to apply at the level of the individual participants, while randomisation is at the cluster level. Application of standard statistical methods to the analysis, which invariably assume no between cluster variation, will tend to bias observed p values downward, thus risking a spurious claim of statistical significance and producing an artificially precise estimate of the intervention effect. The extent of this bias for clusters of fixed size m is proportional to the magnitude of the inflation factor IF which is analogous to the bias associated with the application of standard sample size formulas. This problem has led to the famous quote in the epidemiological literature by Cornfield that "randomisation by cluster accompanied by an analysis appropriate by individual is an exercise in self deception and should be discouraged."[8] As stated in other words by Wood and Freemantle,[9] "fitting a model that ignores (between cluster variability) is akin to expecting a free lunch—and there is no such thing as a free lunch." Methods of analysis for binary, quantitative, and time to event outcomes are discussed in detail by Donner and Klar.[10]

There are also ethical challenges unique to cluster randomisation trials to which attention has been given only recently.[11 12] For example, it is often necessary to distinguish 2 distinct levels of informed consent in such trials: (1) informed consent for randomisation (usually provided by a "decision maker," such as a physician or clinic director, and (2) informed consent for the individual participants, given that randomisation has occurred. By analogy to current ethical requirements for clinical trials, it would be unethical not to obtain informed consent from every cluster member before random assignment. It appears to be an unresolved question as to whether such a strict analogy is required for cluster randomisation trials, particularly when relatively large clusters, such as medical practices or entire communities, are randomised. This issue is likely to be the subject of considerable debate over the coming years as cluster randomisation designs become more widely used in the health research community.

ALLAN DONNER, PhD
University of Western Ontario, London, Ontario, Canada

1   Avorn J, Soumerai SB, Everitt DE, et al. A randomized trial of a program to reduce the use of psychoactive drugs in nursing homes. *N Engl J Med.* 1992;**327**:168–73.
2   Kinmonth AL, Woodcock A, Griffin S, et al. Randomised controlled trial of patient centred care of diabetes in general practice: impact on current wellbeing and future disease risk. *BMJ* 1998;**317**:1202–8.
3   Grosskurth H, Mosha F, Todd J, et al. Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial. *Lancet* 1995;**346**:530–6.
4   Rhee SO, Luke RD, Culverwell MB. Influence of client/colleague dependence on physician performance in patient care. *Med Care* 1980; **18**:829–41.
5   RiceN, Leyland A. Multilevel models: applications to health data. *J Health Serv Res Policy* 1996;**3**:154–64.
6   Koepsell TD. Epidemiologic issues in the design of community intervention trials. In: Brownson RC, Petitti DB, editors. *Applied epidemiology, theory to practice.* New York, NY: Oxford University Press, 1998;177–211.
7   Donner A, Birkett, N, Buck C. Randomization by cluster. Sample size requirements and analysis. *Am J Epidemiol* 1981:**114**:906–14.
8   Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol* 1978;**108**:100–2.
9   Wood J, Freemantle N. Choosing an appropriate unit of analysis in trials of interventions that attempt to influence practice. *J Health Serv Res Policy* 1999;**4**:44–8.
10 Donner A, Klar N. *Design and analysis of cluster randomization trials in health research.* London, UK: Arnold Publishing Co, 2000.
11 Edwards SJ, Braunholtz DA, Lilford RJ, et al. Ethical issues in the design and conduct of cluster randomised controlled trials. *BMJ* 1999;**318**:1407–9.
12 Glanz K, Rimer BK, Lerman C. Ethical issues in the design and conduct of community-based intervention studies. In: Coughlin SS, Beauchamp TL, editors. *Ethics and epidemiology.* New York, NY: Oxford University Press, 1996;156–77.

Correction

In the Poirier *et al* abstract on page 51 of the *Evidence-Based Mental Health* May 2000 issue, the table was not labelled as using per protocol analysis data. The editors acknowledge, with appreciation, Dr Joe Reilly and colleagues of Middlesbrough, UK, for drawing our attention to this omission. Using last observation carried forward data, venlafaxine had a higher rate of remission compared with paroxetine at 28 days (37% *v* 18%, p = 0.02; number needed to treat 6, 95% CI 3 to 32).