*British Journal of Psychiatry* (1995), **166**, 727–733

**Peer Review**

# Peer Review of "Risperidone in the Treatment of Patients with Chronic Schizophrenia: a Multi-National, Multi-Centre, Double-Blind, Parallel-Group Study versus Haloperidol"

A. L. JOHNSON and D. A. W. JOHNSON

The following assessments represent two reviewers' comments on both the original and revised versions of the paper by J. Peuskens on behalf of the Risperidone Study Group. The reviewers were originally anonymous.

The responses to the assessments, from J. Peuskens on behalf of the Risperidone Study Group, follow the reviewers comments.

## D. A. W. Johnson

### Assessment of original manuscript

I am afraid that this research is basically flawed and cannot, therefore, give a valid scientific result.

The study was carried out on a group of chronic schizophrenics (DSM–III–R diagnosis). However, nothing more is known about the chronicity or illness characteristics of this group.

The patients were on an unknown medication regime prior to the trial. No details are given about the dosage of the pre-trial drugs or, indeed, the number or type of drugs during the preceding number of months. It is, therefore, impossible to judge whether the entry into the trial represented a reduction, an increase or a maintenance of pre-trial medication.

The study was designed to be an eight week trial but in effect the patients did not achieve the target dose until the end of the first week and, therefore, the true trial was for only seven weeks. This is just about a long enough period to evaluate the immediate or acute response in a drug free patient. It tells us nothing about the continuation or maintenance effect of the medication.

The washout period of one week is of no value whatsoever in patients who have been on medication continuously for weeks or months. In particular, it should be noted that a significant proportion of patients had been on depot injections and in these patients the continuing effect would be for many months if they had been established on a maintenance dose prior to the trial. The trial was, therefore, a comparison of haloperidol plus a diminishing dose and effect of the previous drug regime, compared to risperidone on a mixture of dose ranges in combination with a diminishing therapeutic effect of previously prescribed drugs. Such a design would not allow an evaluation of the true effect of risperidone in any dose compared to haloperidol in any dose.

Haloperidol was prescribed in a dose of 10 mg daily to one comparison group. The value of this dose of haloperidol in any group of schizophrenics is debatable. It might be that it gives a low number of side-effects at the price of poor therapeutic control or, alternatively, a high level of side-effects. This is not just because of the dose but because the individual may have been changed from a neuroleptic with less risk of extrapyramidal symptomatology. We do not know, of course, whether we are looking at dose reductions or what the pre-trial levels may have been. We also do not know what proportion of patients in each group had their anti-Parkinsonian medication withdrawn on day one of the washout period.

In the results I note that the washout phase was reduced to six days in 25% of patients and the drop-out rate for the eight week period was also 25%.

My worst fears are confirmed by a clinical improvement in 20% of patients during this seven week period of active treatment with only 10 mg of haloperidol. This raises major questions. My anxieties are made even worse by the statement in the Discussion section (not mentioned in the Method section) that the patients were, in fact, poorly controlled.

No amount of statistical analysis, argument or rationalisation can overcome the basic deficiency in the methodology. It must be borne in mind that this is presented as a major trial supporting the use of a new drug to treat a major mental illness. It is, therefore, something that must be taken with great seriousness. I do not believe this study reaches the modern requirements of a scientific evaluation.

### Assessment of revision

Schizophrenia remains one of the most serious mental illnesses dealt with by psychiatrists. The history over the last 20 years is that of a number of false starts with claimed new treatments. Although

undoubtedly we have learnt to use our physical treatments to the better advantage of the patients, there has really only been one significant breakthrough since the introduction of chlorpromazine in 1952. The unequivocal advance is clozapine. Happily, we are now on the brink of a number of new atypical neuroleptics and it will be difficult for the prescribing clinicians to evaluate the relative advantages or disadvantages of these drugs unless the information presented is clear and unequivocal. There are many limitations placed upon us by ethics in carrying out these tests and it is right that we should live with these limitations. However, it must also be right that the limitations of the research carried out must be honestly acknowledged and any conclusions reached modified by the presence of such limitations.

The authors must accept and acknowledge that depot medication continues for many weeks and even months after the first missed injection. A number of discontinuation studies clearly identify no statistical disadvantage to patients until three months after the first missed injection. In this study 37% of patients were on depot neuroleptics. Not only must the authors acknowledge the considerable limitation this places upon their study, since this percentage of patients would have had dual medication of an unknown amount for the period of assessment, but they must also show the number of patients who were previously on depot injections who were allocated to each of the trial cells.

It must further be acknowledged that many patients under routine clinical treatment actually benefit from a reduction in their medication dose rather than an increase or, indeed, a continuation of their previous dose level. All patients in the haloperidol group were on 10 mg daily. We need to know whether this represents a reduction, an increase or the same dose for the majority of patients. We need to know similar information for the risperidone groups. The better results in the risperidone 4 and 8 mg group might be explicable purely on a change of dose rather than a change of drug.

It is generally accepted that the prescription of anticholinergic drugs has an influence upon the circulating levels of neuroleptics and/or the receptor activity. If the haloperidol patients received more anticholinergic drugs, it is possible this may have influenced the outcome score. Not least because anticholinergics themselves have side-effects which influence score ratings.

The third consensus conference on the methodology of clinical trials of antipsychotic drugs may have concluded that the optimal duration for short-term trials was between four and eight weeks but this can only be in demonstrating an initial response of a drug. It is clear that with a washout of seven days or less (for 25% of patients) an eight week trial must include some influence from the previous neuroleptics and their metabolites if those patients have been on chronic treatment. An eight week trial can only give a very preliminary and tentative outcome unless it is in individuals who have previously been drug free.

A number of patients had their washout periods shortened because of acute deterioration in their mental state. In these patients the eight week trial period becomes a trial of treatment for an acutely ill patient. In the remaining patients it is part of ongoing maintenance or continuation therapy for the eight weeks assessed. It is, therefore, necessary to know how these patients were distributed between the two principal groups (risperidone and haloperidol). This might also influence the fact that 4 to 8 mg of risperidone was more effective than a lower dose.

A further guiding principle is that the results of a particular research project can only refer to the sample studied unless there is clear evidence that it is possible to generalise from the sample studied. In this study we really have almost no information about the index group. I know only too well the problems of collecting a large sample and the many ethical considerations involved. Nevertheless, I do think that attention must be drawn to this further shortfall.

Another example of where the authors appear to be rather dogmatic rather than recognising the very complex area is their discussion about the use of haloperidol 10 mg and whether or not this is the most effective dose. The whole area of dose–response curves in the treatment of schizophrenia is controversial with conflicting results. Most treating clinicians would recognise that whereas there are guidelines and dose spectrums which are more appropriate in certain groups of patients, in reality, the prescribing of neuroleptics is a very individual process for particular patients and dose variations are dependent upon the target symptoms or other goals with respect to behaviour, drug tolerance etc.

I believe the authors are not being sufficiently cautious in the presentation of their results. In particular, they are not drawing attention to the many pitfalls of this survey. At the very most, this study is a very small step along the pathway of establishing risperidone as an effective alternative to other treatments. I believe it would be in everyone's interest to make much stronger acknowledgement of these considerations.

A. L. Johnson

**Assessment of original manuscript**

I was extremely pleased, and indeed, very surprised to see this paper. I had no idea that any randomised clinical trials (RCTs) in psychiatry had been designed with, or achieved, the sort of sample size attained in this study. Some of us have been advocating for many years that RCTs in psychiatry are two orders of magnitude too small, and that the only way to get sensible answers to important clinical questions is through large, simple, multi-centre studies. Unfortunately it is not easy to get psychiatrists to collaborate on this sort of scale, and the Study Group (especially the Organisers) must be congratulated on a major achievement in completing this important trial.

I was equally surprised to find that so large a trial does not have an attached statistician, unless there is one buried somewhere in the Appendix. Perhaps there wasn't one, which might explain some of the statistical deficiencies in the analysis and presentation; these must be addressed and the paper revised accordingly.

Page 3, Summary. (p. 712 this issue.) Contains no quantitative estimates of comparative efficacy with precision, for example in line 13, similarity of the haloperidol and risperidone 16 mg responses should be quantified with a mean difference and 95% CL.

Page 4, bottom and page 5, top. (p. 712.) Again in reporting results from previous studies, the emphasis is on statistical rather than clinical significance. It would be far more informative to quantify the differences between treatments.

Page 6, line 6. (p. 713, Study design.) Describe precisely how randomisation was carried out (see Altman & Dore. Randomisation and baseline comparisons in clinical trials, *Lancet*, **335**, 149–153 (1990)), including stratification. There is no description of the overall design of the trial, and no statement about determination of size.

Page 7, Efficacy evaluation. (p. 713.) I have major problems with the methods of analysis which are rather inefficient; these are pointed out below.

Page 8, line 5. (p. 714.) I do not understand the "comparison with the previous treatment".

Page 9, Statistical analysis. (p. 714.)

(1) para 1. Do not compare baseline distributions of randomised groups using tests of statistical significance; randomisation guarantees that they differ by chance ($P = 1.0$). If you are concerned about imbalance on possible prognostic factors then adjust for them by appropriate analysis, whether or not such factors differ significantly between treatment groups. (See Altman, Comparability of randomized groups, *Statistician*, **34**, 125–136 (1985)).

(2) para 2. Why analyse changes from baseline? This is inefficient, and you might do much better using repeated measures analysis of covariance of (perhaps transformed) scores; such methods also enable use of both baseline scores and avoid unnecessary selection of cutpoints, such as the arbitrary 20% reduction in PANSS. (See Frison & Pocock, Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design, *Statistics in Medicine*, **11**, 1685–1704 (1992)). An alternative would be to use overall summary measures as proposed by Matthews *et al* (Analysis of serial measurements in medical research, *BMJ*, **300**, 230–235 (1990)). The analysis must reflect design features of the trial; suddenly in line 5, we find stratification by country, the first hint that within-country randomisation might have been employed. If such stratification is relevant for one outcome variable, it is relevant for all. In line 3, I do not understand the connection between 95%CL and Fisher's LSD; exact CL are available for small groups, but you don't have this problem. The last sentence of this paragraph is incorrect.

(3) Page 10, line 3. (p. 714.) When you expect a dose–response relationship, insisting on significance of the overall test is too stringent, and Kaplan-Meier is rather weak. Why not use the Cox model with appropriate modelling of dose-effects?

(4) Page 10, line 6. (p. 714.) You should briefly describe what the Jonckheere-Terpstra test is intended for – I suspect that no readers of the *BJP*, and few statisticians will have heard of it! It would be interesting to know why you chose this particular test rather than one of the host of others which are available? (See for example, Hochberg & Tamhane, *Multiple Comparison Procedures*, Wiley, New York (1987)).

(5) Page 10, para 3. (p. 714.) Analysis of EPS must also reflect design features, so use logistic regression not chi-squared.

(6) Page 10, para 4. (p. 714.) This is very weak. The trial size is large, so you may get statistical significance for effects that are clinically unimportant. More sensible interpretation is required. (See Gardner & Altman, Confidence intervals rather than *P* values: estimation rather than hypothesis testing. *BMJ*, *i*, 747–750 (1986)).

Page 10, Results, para 1. Delete reference to statistical tests. Table 1 requires proper layout and

https://doi.org/10.1192/S0007125000075097 Published online by Cambridge University Press

summary statistics – females under males; 25th, 50th and 75th centiles for age and age at onset. Cannot the sex be determined for the missing patient? The table shows that patients were randomised in equal numbers to the six treatment groups; a more efficient design would allocate in the ratios 1:1:1:1:1:2.2. Is the mean a sensible summary statistic for duration of the washout period? Why not give the actual distribution?

Table 2. Rather than mean no. of days in study, tabulate the completed weeks, so that the distribution of drop-out times can be compared among groups.

Page 11, Efficacy. (p. 715.) Delete first sentence. I have already suggested that the method of analysis is inefficient and arbitrary – why 20% reduction (as opposed to any other) from baseline (day 0?). What are the denominators for the %s – should be total number randomised in each group. Figure 1 is misleading because of the non-zeroed scale. What is regarded as "clinical improvement on the total BPRS"?

Tables 3 and 4. All this can be improved by using ANCOVA, including checks that the distributions of the PANSS and BPRS scores satisfy requirements of approximate Normality. Does the analysis of subscale scores really enhance that of the total scores? Test for trend (both linear and quadratic) across the doses?

Page 12. (p. 716.) Lots of significance tests without clinical interpretation.

Page 13. Delete the post hoc analysis – it just clutters the main outcomes.

Figure 4. (Fig. 3, p. 721.) These data could be analysed very nicely by logistic regression analysis, which would show a linear trend across the five doses.

**Assessment of revision**

The revised paper is an improvement on the original, but I have a few outstanding points which require attention. Whether or not the name of the statisticians should appear amongst the authors is an issue for the principal investigators rather than the reviewers. However given the size of this trial, and the amount of work in handling the data and in conducting the analyses, I would hope that the principal investigators would want the statisticians to be up front. I feel that an acknowledgement is not sufficient, and that you should consider the standard presentation for multicentre, collaborative trials, i.e. "The Risperidone Study Group" under the main title, then everyone listed in the Appendix under appropriate headings, e.g. Principal Investigators/Clinical Coordinators, Working Party/ Steering Group Members, Statisticians, Authors, Collaborating Physicians, etc. Everyone then feels part of the study and may want to take part in another one.

Page 3, Summary. (p. 712.) I would like 95%CL on % response rates in risperidone and haloperidol treated patients, as well as difference between these %s, again with 95%CL. Also 95%CL on differences in CGI scores at endpoint.

Page 10, Statistical analysis, para 1. (p. 714.) You have misunderstood this point. Significance tests are not useful for detecting imbalances in the distributions of baseline variables. Delete this paragraph, and the third sentence of the Results section.

Page 10, bottom. (p. 714.) I suggest that using an overall significance level of 10% is inadequate. Given the large number of significance tests conducted during the analysis, you will in fact be operating with a much higher (closer to 50%) overall significance level (and your power calculation (under Study Design on page 6) does not take account of the multiple testing). It might be reasonable to be more stringent and operate at an overall level of 1%, certainly not above 5%. So delete everything after "significance" on the line above "Results".

Table 1. For the last four variables instead of means give median (25th and 75th centiles). The same applies to number of days in study near the bottom of Table 2.

Figure 1. This has a non-zeroed scale which exaggerates the differences between treatments. I suggest reduce the size of the blocks, and use a vertical scale anchored on zero.

Tables 3, 4 and 5. These three tables highlight one of the main problems with the paper – so much information is presented that it is difficult to focus on the important results, and to avoid getting overwhelmed by the data. The authors need to think very carefully about the presentation in these tables, with a view to reducing it by more than 50% using a better subdivision of information between tables and text. Remember that in clinical trials we are interested mainly in comparative statistics between the randomised groups, rather than within-group changes, so all 95%CL which relate to within-group changes should be deleted. In Table 3 I suggest report the mean change from baseline only and the number of observations on which it is based, together with the differences in changes between each risperidone dose and haloperidol with a 95%CL

and its significance. Would it not be sufficient to present the information for the Total PANSS, and then to comment on the subscales in the text. The same comment applies to Table 4 which could be restricted to Total BPRS. Table 5 also restricted to Total Score.

**Comment on final version**

Still longer than I hoped but, on balance, OK to publish – it is a large and important trial. The authors have responded to many of my comments over the two revisions.

**A. L. Johnson**, MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 2SR; **D. A. W. Johnson**, Withington Hospital, West Didsbury, Manchester M20 8LR.

**Author's response to the initial assessments**

Thank you for giving me the opportunity to respond to your reviewers' comments. I was astonished that the reviews were so conflicting in nature. I have revised the manuscript in an attempt to deal with all of their points, and would like to mention that the Risperidone Study Group aimed to perform the study according to general guidelines on the design of such trials and in line with the current literature, for example, Awad.

**Review 1**

*Introduction*: The two sentences concerning mean shifts versus baseline have been removed from the text.

*Patient characteristics*. More details concerning the patients' age at first hospitalisation, mean number of previous hospitalisations and mean duration of current hospitalisation have been added to Table 1, demonstrating the chronicity of their illness.

*Previous medication*. Information concerning pre-trial medication has been added to the Results section.

Because of the diversity of the treatments administered it was not possible to calculate the dose in haloperidol equivalents for all patients. However, it is worth mentioning that 298 patients (22%) had been previously treated with haloperidol, the median dose being 10 mg.

*Study duration*. The aim of the study was to evaluate the short-term response to risperidone; this fact has been incorporated in page 5. A study duration of 4 to 8 weeks has been recommended for this purpose (Third Consensus Conference on the Methodology of Clinical Trials of Antipsychotic Drugs, October 14–16 1990).

Indeed, the majority of trials concerning short-term treatment in the literature are of 4 to 8 weeks duration. Patients who participated in this trial were given the opportunity to continue treatment in an open label long-term trial. The results of maintenance treatment will be described in another paper which is in preparation.

*Wash-out period*. Discussion of the duration of the wash-out period has been added (see page 17). As you will see, attempts were made to overcome the methodological problems highlighted by the reviewer, but these could not be totally eliminated.

*Haloperidol dose*. Discussion points concerning the fixed doses of risperidone and 10 mg dose of haloperidol have been expanded. Although it is not mentioned in the manuscript, 10 mg is considered a normal dose in Europe.

*Antiparkinson medication*. Page 6 of the manuscript states that all patients were withdrawn from antiparkinson medication on the first day of the wash-out period. Thirty-three per cent of patients had been previously treated with antiparkinson agents; this has been added on page 11.

*Drop-Outs*. I do not understand the statement "In the results I note that the wash-out phase was reduced to six days in 25% of patients and the drop-out rate for the eight week period was also 25%". If this statement refers to a comparison of drop-out rates between the wash-out and the treatment periods, it should be mentioned that these are basically different situations which are not comparable.

Alternatively, if the statement refers to the fact that only 50% of patients were available for evaluation, it should be noted that the 25% of patients for whom the wash-out phase was reduced to 6 days were still evaluated. The 8-week evaluations were performed on 75% of patients.

*Clinical improvement*. It seems that clinical improvement was misunderstood by the reviewer. The primary measure of efficacy was the percentage of patients showing clinical improvement, defined as a reduction in total PANSS score of at least 20% from baseline (see efficacy evaluation, page 7 of the manuscript); 58.7% of haloperidol-treated patients improved as assessed by this parameter, not 20%.

**Review 2**

This reviewer comments on the absence of a statistician attached to this trial. There were indeed statisticians involved, but, as the appendix lists only

https://doi.org/10.1192/S0007125000075097 Published online by Cambridge University Press

the psychiatrists taking part in the investigations, they were not listed. We have, however, made an acknowledgement to the lead statistician; if you advise that his name should be added to the appendix please let me know.

*Page 3, Summary; page 4–5*. The sentence has been deleted; this was not a major efficacy parameter and has therefore been excluded in order to achieve the required summary length.

*Page 6, line 6*. Details of the overall trial design, determination of size and randomisation procedure have been added to the methodology (page 6, 7).

*Page 8, line 5*. "Comparison with previous treatment" has been clarified (page 8).

*Page 9, statistical analysis*. I agree that the significance of the tests that compare demographic and baseline distributions are meaningless (given the randomisation) and uninformative (given the multiplicity of such tests). These tests traditionally serve no purpose other than to provide a descriptive idea of the possible imbalances between the treatment groups.

One can indeed apply a lot of possible tests or procedures to compare treatments. Some of them, like repeated measures analysis, are more advanced. The drawback of such models are that they are based on assumptions about the data that are not always true (distribution of the data, missing data at random etc), and that they are complex and not easy to interpret for everybody. Another way of handling repeated measures is to derive summary statistics that summarise the repeated observation into one or a few statistics. The shift versus baseline at endpoints is one possibility, the area under the curve is another. The former summary statistic has a clear interpretation; it represents the absolute improvement or regression at endpoint. Various statisticians like Stuart J Pocock and Matthews *et al* have confirmed the value of these procedures and presented them as a first choice.

Several methods were performed, but all analyses confirmed the conclusion obtained by the method described in the manuscript (sensitivity analysis).

*Page 12*. I was not clear of the reviewer's meaning re 'lots of significance tests without clinical interpretation'. Actual values for the mean change in scores versus baseline are listed in Table 3. To allow greater interpretation of the results 95% confidence intervals have been added. Please could you clarify further if this is not what the reviewer required.

*Page 13*. Post hoc analysis deleted.

*Figure 4*. The main purpose here was to compare the different risperidone doses with haloperidol 10 mg. Logistic regression is indeed a valuable tool to investigate a linear trend across different dosages, but it is not straightforward to include the haloperidol treatment group. Logistic regression analysis across the risperidone doses had not been performed for this reason and because the figure clearly shows the trend. Results of this analysis have now been added to the text (page 15).

I hope these changes meet with your approval. I look forward to hearing your response.

### Author's response to the second assessments

I have revised the manuscript in light of the reviewers' comments wherever possible, and hope that the explanatory points listed below will clarify some issues.

#### Review 1

*Depot neuroleptics*. The percentages of patients previously treated with depot neuroleptics in each group has been added. The limitations of the prior administration of depot preparations was also discussed in the manuscript (now page 16). However, in this large number of patients (>1000) an effect of previous depot neuroleptic treatment in only one treatment group would not be expected.

*Dose reduction*. For the 22% of patients previously treated with haloperidol, median doses for each treatment group have been added. These demonstrate that haloperidol use and dose were well-distributed among all treatment groups. If we can assume that these 22% of patients provide a representative measure of the previous antipsychotic load in all treatment groups, if the treatment effect recorded in the study was related to an increase or decrease in antipsychotic dose, the effect should be well-distributed over all treatment groups.

*Anticholinergics*. We have been unable to find evidence that it is generally accepted that the prescription of anticholinergic drugs influences the circulating levels of neuroleptics and/or receptor activity. I would be interested in any definitive references. For the purpose of minimising the influence of anticholinergic treatment of ESRS scores, results were expressed as "shifts to the maximum score". This is explained on page 8 of the manuscript.

*Wash-out period*. The percentages of patients in each group whose wash-out period was reduced because of acute deterioration have been added.

*General comments*. I do not believe that our statements, assumptions and conclusions are any more, or less, dogmatic or incautious than those of other publications on other drugs. Indeed, the very large number of patients (>1000) included in this trial should make the results much more reliable than those of other publications, the majority of which have included a far smaller number

of patients. It should be mentioned that, to our knowledge, this is one of the first very large multinational trials evaluating the dose responses in different fixed dose regimens against an active control. In addition, valuable evaluation scales such as the PANSS were used. All this should allow us to say that the results from the study are at least as good as those of other trials with other compounds.

**Review 2**

*General comments*

With respect to the authors, we would prefer to keep the Risperidone Study Group members listed as they are.

Points concerning the summary statistical analysis and Table 1 and 2 have been added. As in our experience physicians are interested in baseline results and results on all subscales, and as it is not easy to include all these results in the text, we prefer to keep Tables 3, 4 and 5. The 95% confidence intervals have been altered to show differences between risperidone and haloperidol. Figure 1 has been deleted as the zeroed scale is no longer a good illustration of the result.

I do hope that you are able to accept our manuscript in its current form, and thank you for the work and time which has been put into the reviews. I look forward to hearing from you.

**J. Peuskens**, University Psychiatric Hospital St Jozef, B-3070 Kortenberg, Belgium