

Author's response to reviews

Title: The feasibility of whole body vibration in institutionalised elderly persons and its influence on muscle performance, balance and mobility: a randomised controlled trial.

Authors:

Ivan Bautmans (ivan.bautmans@vub.ac.be)

Ellen Van Hees (ellenvanhees@tiscali.be)

Jean-Claude Lemper (jclemp@pgc.be)

Tony Mets (tony.mets@az.vub.ac.be)

Version: 3 **Date:** 14 September 2005

Author's response to reviews: see over

RESPONSE TO THE REVIEWERS.

Dear Sir,

Thank you for being interested in our manuscript. We read carefully your remarks and suggestions. We tried to adapt our manuscript in the perspective of your review. Our responses to your remarks are listed below in order of appearance in your review (new pagination in the response, when necessary initial pagination in *italic*). We hope that our motivations and the changes we made to the manuscript correspond with your expectations. Nevertheless, feel free to make all necessary additional linguistic and editorial changes.

Reviewer Vance Berger

General

First, the use of nonparametric analyses is appropriate, although the Shapiro-Wilk test is a waste of time. It lacks power to detect non-normality, so even if it had not indicated non-normality, it would still be appropriate to use nonparametric analyses. The fact that non-normality was found indicates that it must have been gross non-normality.

Reference to the non-normality of the datasets and to the Shapiro-Wilk test have been removed from the Methods-section, page 5 'Statistical analysis'

Second, there were three post-randomization exclusions out of 24 randomized subjects, so 12.5%. The objective of a comparative study is to determine if the subjects treated with one treatment modality fare better than the subjects treated with the other treatment modality. Given the variability among subjects, it is not surprising that some subjects in each group will fare better than some subjects in the other group. This does not constitute a treatment effect. A treatment effect refers to an analysis which incorporates all subjects in one group and compares them to all subjects in the other group. This means that analyses which do not include all subjects randomized to each group are potentially misleading, and do not address the questions of interest. Only an analysis performed on the intent-to-treat population is reliable (Peduzzi, P., Detre, K., Wittes, J., Holford, T., "Intent-to-Treat Analysis and the Problem of Crossovers", *Journal of Thoracic Cardiovascular Surgery* 101, 481-487, 1991; Newell, D. J., "Intention-to-Treat Analysis: Implications for Quantitative and Qualitative Research", *International Journal of Epidemiology* 21, 5, 837-841, 1992). Without this, the comparison would be "de-randomized" (Denis et al., *Urology* 49 (4A), 15-26, 1997), and would reflect the experiences of only a subset of the randomized patients. Moreover, this subset was not selected at random. In fact, there is every reason to believe that those patients who contribute data differ systematically and substantially from those who do not.

As reported on page 5 (*Old manuscript Results, first paragraph, line 5-8*), 3 subjects in the WBV+ group dropped out. None of these three participants was available for end-evaluations: 1 female presented groin pain after the first exercise sessions and refused to continue the program, 1 female became afraid to go to the rehabilitation room, 1 male developed airway infection during the study apparently unrelated to the WBV (accompanied with severe decline in clinical condition and confined to bed rest making re-assessment impossible). Therefore, these dropouts must be considered as 'missing values' and not as 'exclusions'. Also, it is evident that these 3 subjects did not cross over from the intervention arm to the control arm of the study. Thus the

principle of intention-to-treat analysis was not violated and we feel that another analysis is not possible. This item is clarified for the reader on page 5 (Results, first paragraph, line 4-10)

The link between these dropouts and the WBV has been added to the discussion in the revised manuscript (page 8 third paragraph, first sentence)

However, we agree that it cannot be excluded that our results would have been different if these persons did not drop out. This remark and a reference have been added to the discussion section of the revised manuscript (page 8 third paragraph line 1-6).

Table 1 has no p-values. While there is a vast literature (e.g., Senn, 1994, "Testing for Baseline Balance in Clinical Trials", *Statistics in Medicine* 13, 1715-1726) which indicates that hypothesis testing for baseline imbalance is illogical in randomized trials, this view is itself illogical. Consider the argument that any hypothesis test would be testing for balance in the population of all possible randomizations which could have occurred. By virtue of the random allocation of subjects to treatment groups, this is necessarily true. Thus, any time this null hypothesis is rejected we know that we have a Type I error. This argument would be compelling if all observed baseline imbalances in the context of randomized trials were necessarily random. Yet systematic baseline imbalances (selection bias) can occur in randomized trials (Berger, VW, "Selection Bias and Covariate Imbalances in Randomized Clinical Trials", John Wiley and Sons, Chichester, 2005); moreover, there is evidence that such selection bias DOES occur in randomized trials (Berger VW, Weinstein S, "Ensuring the Comparability of Comparison Groups: Is Randomization Enough?", *Controlled Clinical Trials* 25, 2004, 515-524). This being the case, the view that one ought not test formally for baseline imbalance in randomized trials confuses the sufficiency of randomization to eliminate systematic imbalances with necessity (Berger VW, Selection Bias and Baseline Imbalances in Randomized Trials, *Drug Information Journal* 38, 2004, 1-2). That is, randomization may be necessary to ensure that any observed baseline imbalances are random, but it certainly is not sufficient, as selection bias can occur even in randomized trials (Berger VW, Christophi CA, "Randomization Technique, Allocation Concealment, Masking, and Susceptibility of Trials to Selection Bias", *Journal of Modern Applied Statistical Methods* 2003; 2(1):80-86). While not generally cast specifically as a test of this type of selection bias in randomized trials, tests of baseline imbalances do, in fact, constitute tests for selection bias, although they are crude ones. Given the potential for selection bias in this trial, better methods to detect selection bias (see Berger VW and Exner DV, Detecting selection bias in randomized clinical trials. *Controlled Clinical Trials* 1999;20(4):319-327 for the specifics of how such a test should be performed) should be employed. However, in the absence of such analyses (and it is hard for me to imagine a good reason why these better analyses would not be presented), the p-values for the baseline imbalances should be presented, even if they are not significant.

All remarks regarding table 1 (*old manuscript*) have been integrated in the new manuscript. Since this table became too large for comfortable reading, it has been split in two components (table 2 and 3). Exact p-values have been added in table 2 & 3. We have read with great interest all the mentioned articles concerning statistical procedures and have considered computing the proposed calculation. Given the randomisation procedure in our study (for all participants at the same moment by lottery, see below), without possible deferred enrolment, the proposed analysis [1] was not performed.

Table 3 needs numerical p-values, and not footnotes.

Numerical p-values have been added to table 4 (*table 3 old manuscript*) in the revised manuscript.

The first reviewer was concerned over the randomization to 11 and 13 instead of 12 and 12. At issue here is the set of restrictions used for the randomization. We know that there was stratification, but the authors did not elaborate on the specific methods by which the randomization was accomplished. One of the primary reasons for going through the added effort and expense of randomization is to minimize baseline imbalances across the comparison groups, and to ensure that whatever imbalances remain are random, and not biases. However, there are many ways to randomize a study (Berger VW and Bears JD, "When Can a Clinical Trial Be Called 'Randomized'?", *Vaccine* 2003; 21; 468-472), so simply stating that it was randomized is not sufficient to convey exactly how the randomization procedure was implemented. Moreover, the specifics of how randomization is performed determine the extent to which randomization can minimize baseline imbalances. That is, the design (specific restrictions placed on the randomization) is a key to determining the extent to which the study was susceptible to selection bias (Berger VW and Exner DV, *Detecting selection bias in randomized clinical trials. Controlled Clinical Trials* 1999;20(4):319-327). Also, the design has implications for which types of analysis are valid (Berger, VW, "Pros and Cons of Permutation Tests in Clinical Trials", *Statistics in Medicine* 19, 1319-1328, 2000). It is for these reasons that Item #8B of the CONSORT Statement (*Annals of Internal Medicine* 134, 8, 4/17/01, 663-694) suggests that information on the specific restrictions placed on the randomization be reported routinely. So the authors need to provide more details concerning the randomization within strata. For example, was blocking used? If so, then what were the block sizes?

Randomisation was done for all 24 participants at the same moment by lottery. Stratification for gender, dependence category for basic activities of daily living (ADL) according to Katz et al [2] and age was applied. A-priori, we considered subjects in ADL-category O or A possibly different from those in ADL-category B; and the oldest old possibly different from younger ones (cut-off at age 84 years). Exact distribution of 12/12 in intervention and control group was not an a-priori criterion.

For each participant a card was made containing identification number, gender (male or female), dependency level (O/A or B) and age (old or oldest old). Next, the cards were put in different baskets dividing the population into 8 subgroups according to the stratification criteria (table 1). From each basket separately, alternatively cards were assigned blindly to the intervention or control group by means of lottery. For each basket separately, the starting sequence for lottery was determined by tossing a coin. The results are displayed in table 2.

Table 1. Distribution of the participants according to stratification criteria

Stratification	N
male, ADL=O/A, old	6
male, ADL=O/A, oldest old	0
male, ADL=B, old	3
male, ADL=B, oldest old	0
female, ADL=O/A, old	7
female, ADL=O/A, oldest old	5
female, ADL=B, old	2
female, ADL=B, oldest old	1

Table 2. Results of randomisation.

Stratification	Intervention (N)	Control (N)
male, ADL=O/A, old	3	3
male, ADL=O/A, oldest old	0	0
male, ADL=B, old	2	1
male, ADL=B, oldest old	0	0
female, ADL=O/A, old	4	3
female, ADL=O/A, oldest old	3	2
female, ADL=B, old	1	1
female, ADL=B, oldest old	0	1
Total	13	11

More detailed information concerning the randomisation procedure has been added to the methods section of the revised manuscript (Randomisation, page 3-4)

The first reviewer is also correct that there is too much data compression in Table 3. More raw data needs to be shown.

Data at baseline and 6 weeks for both intervention and control groups are presented in table 4 (*table 3 old manuscript*) in the revised manuscript.

The second reviewer also raises some good points, and I would like to see the author responses to these questions.

Please see our response to reviewer Olivier Bruyere.

And I agree that the figures do not help.

The figures, as well as all reference to them, have been removed from the revised manuscript

Reviewer Christophe Delecluse

**General
none**

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. I have some major concerns regarding the interpretation of the results and the statistics: One should be very cautious in the interpretation of the results for the following reasons:

- the number of subjects is limited**
- the inter-individual variability in most of the outcome parameters (changes) is large**
- statistical analysis was not performed on the absolute values but on the differences (changes) between pre-post.**

For this reason I advice that an expert statistician revises this manuscript.

Please see our response to Reviewer Vance Berger.

In two outcome parameters a significant difference in effect was found between WBV and control condition (main findings of the study): Tinetti-test and Timed get-up-and-go test. For both parameters one should consider some facts (some limitations) before making further conclusions:

a. Tinetti-test: it is not clear why the control group showed a significant decrease in performance???. However this fact determines mainly the conclusion that WBV had a positive effect on the result of the Tinetti tests.

More detailed discussion concerning the changes in scores on the Tinetti-test has been added to the revised manuscript (Discussion, page 6, first paragraph, line 8-12).

b. Timed get-up-ad-go test: if we consider the absolute values, we notice that the pretest values of the WBV-group were 3.1 seconds slower compared to the controls. Although this difference is not significant, it can not be denied that in the post test the absolute performance of the control group was still better compared the WBV-group. It may be that the changes in performance was larger in the WBV-group due to the lower starting value in that group?

In table 3 (*table 1 old manuscript*) of our revised manuscript we have added the baseline data of the participants in the intervention group who were reassessed at 6 weeks. From these data it can be seen that the baseline scores on the timed get-up-and-go test were more similar in both the intervention and control groups (15.3 ± 5.5 and 14.8 ± 6.3 seconds respectively).

2. We notice that there is a large variability in the leg extension parameters (in table 3:changes). It is clear that this variability reduces the chance to find a significant difference. Did some of the elderly had problems in performing these tests? The authors refer to high single session reproducibility and high intra-observer reliability coefficients reported in the literature. Where these coefficients recorded for frail elderly or for another group of subjects?

Important age-related variability is a characteristic of all studies involving geriatric populations, especially in those dealing with frail elderly persons. None of the participants showed difficulties in performing the linear isokinetic leg extension tests. Our experience in previous studies involving elderly subjects indicate that this is a useful test procedure in elderly persons [3, 4]. Discussion concerning the possible influence of this variability on our results has been added in the revised manuscript (Discussion, page 7, second paragraph, line 3-10).

The study of Lenaerts et al [5] reports high reliability of the test procedure in young individuals. This has been clarified in the methods section of the revised manuscript (Methods, page 5 Functional performance, paragraph 4, line 5-6).

3. Background pg 3 §3: I don't think that the volume (3 sets of 10 reps) of the traditional resistance training programs is a problem for institutionalized elderly. If one performs 3 strength exercises for the lower limbs, he will only need 10 minutes (rest periods included). I think that rather the intensity (70-80% of 1RM) is a problem for frail elderly. please comment.

We agree that a limited number of exercises can be performed at high intensity within short exercise duration. By the term 'exercise volume' we consider not only the duration of the exercise, but the combination of total exercise duration, exercise intensity and number of repetitions, thus reflecting the total weight moved over time in relation to rest periods. There is sufficient evidence that frail elderly subjects are able to exercise at high intensity (up to 80% of maximal resistance) [6, 7]. In our experience fatigue, a symptom reported by 98% residents of a long-term care facility [8], as well as a lack of motivation discourage frail elderly patients in performing a certain volume of physical exercise (whether it is a limited number of repetitions at high intensity or a large number of repetitions at low intensity). The paragraph has been adapted in the revised manuscript (page 3 Background, third paragraph). We have also adapted the use of terminology in table 1 and in the discussion section (page 7 third paragraph lines 16 and 17).

4.Methods -randomisation: If you randomise 24 subjects in two groups, one would expect two groups of 12 persons, however N=13 in the WBV-group and N=11 in the control group??

Please see our response to reviewer Vance Berger.

5. The duration of the study was limited to 6 weeks only. Why?

The rationale for the duration of our study has been added to the discussion section of the revised manuscript (Discussion page 7, third paragraph, line 1-5).

6. The attendance of the training sessions was 10 % higher in the WBV-group (96%) compared to the Control group (86%): may this have affected your results?

Discussion concerning the difference in attendance between both groups has been added to the revised manuscript (Discussion, page 8, second paragraph).

7. There is no reflection on the potential mechanisms that may be activated by the vibration stimulus. Referring to potential mechanisms may help the reader to understand why:

- longer WBV-interventions may or may not be more successful
 - in this study effects on upper limb flexibility are recorded, while performing lower body exercise

We agree that more information regarding potential mechanisms of WBV might improve the readability of our manuscript. Brief information and a reference in have been added to the discussion of the revised manuscript. (Discussion, page 7 third paragraph, two last sentences and page 7 last paragraph line 5 - first paragraph page 8).

Reviewer Olivier Bruyere

General

This manuscript presents additional, and in some cases new, evidences of the positive effects of WBV in institutionalised elderly persons. The manuscript is well written and provides important information. The message could be made even clearer if the authors considered the comments listed below.

Because of the nice design of this study (RCT), the most useful information relates to the differences between groups. Unfortunately, it does not clearly appear in the results section, or in table 1. If not significant differences were observed, the authors should try to explain it. From my point of view, the authors should not focus on within-group analysis (at least not in the abstract, nor in table 1), which is less interesting.

Results of within group changes after 6 weeks have been removed from the results section of the abstract in the revised manuscript. We hope that our modifications (see also reviewer Vance Berger) to table 1 and table 3 improved their interpretation for the readers.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. Out of the 62 subjects eligible for this study, I understand that 29 (46%) had exclusion criteria. This could limit the feasibility of WBV in institutionalised subject. The authors should give a comment in the discussion section.

The discussion concerning the feasibility of WBV in institutionalised elderly persons has been expanded in the revised manuscript (Discussion, Last paragraph, page 8 line 8-15).

2. What is the primary outcome in this study? Feasibility or influence of WBV? This is not clear in the title and in the manuscript.

During the conception of the study, both feasibility of WBV and improvement due to the intervention were considered as linked to each other and in that sense as related primary outcome measures. The study protocol has been registered as such in the Current Controlled Trials database.

<http://www.controlled-trials.com/isrctn/trial/ISRCTN62535013/0/62535013.html>

3. Was the outcomes assessment (Get up and go, Tinetti test, ...) performed blindly (by someone who was unaware of the group assignment of the participant)?

Yes, this item has been clarified for the reader in the methods section of the revised manuscript (Page 5, Methods, Measurements, Functional Performance, last sentence).

4. Table 1 in not very clear. Please provide exact p value for between-groups differences. Delete within-group comparison from this table.

Table 1 (*old manuscript*) has been adapted and expanded to table 2 & 3 in the revised manuscript.

5. The authors could also provide body balance and gait results of the Tinetti-test.

Data concerning body balance and gait have been added to the revised manuscript (table 2 & 3; page 6 results, second paragraph).

6. This is surprising that, contrary to what has been showed in other studies (may be with another type of WBV platform?), no change occurred in the Tinetti-test, in the WBV+ group. Another surprising information relates to the decrease, in the control group, of the Tinetti-test. The authors should try to explain this.

More detailed discussion concerning the changes in scores on the Tinetti-test has been added to the revised manuscript (page 6 Discussion, first paragraph line 8-12).

7. I do not believe that the two figures are useful for the readers. Please delete them.

The figures, as well as all reference to them, have been removed from the revised manuscript

REFERENCES

1. Berger VW, Exner DV: **Detecting selection bias in randomized clinical trials.** *Control Clin Trials* 1999, **20**:319-327.
2. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW: **Studies of Illness in the Aged. The Index of ADL: A Standardized Measure of Biological and Psychosocial Function.** *Jama* 1963, **185**:914-919.
3. Bautmans I, Njemini R, Vasseur S, Chabert H, Demanet C, Mets T: **Biochemical changes in response to intensive resistance exercise training in the elderly.** *Gerontology* 2005, **51**:253-265.
4. Verbruggen LA, Lenaerts A, Duquet W, Pauwels C, Mets T: **Relationship between muscle strength in functional movements, activities of daily living and bone mineral density in osteoporotic women.** *Osteoporosis International* 2000, **11**:S109.
5. Lenaerts A, Verbruggen LA, Duquet W: **Reproducibility and reliability of measurements using a linear isokinetic dynamometer, Aristokin.** *J Sports Med Phys Fitness* 2001, **41**:362-370.
6. Fiatarone MA, Marks EC, Ryan ND, Meredith CN, Lipsitz LA, Evans WJ: **High-intensity strength training in nonagenarians. Effects on skeletal muscle.** *Jama* 1990, **263**:3029-3034.
7. Fiatarone MA, O'Neill EF, Ryan ND, Clements KM, Solares GR, Nelson ME, Roberts SB, Kehayias JJ, Lipsitz LA, Evans WJ: **Exercise training and nutritional supplementation for physical frailty in very elderly people.** *N Engl J Med* 1994, **330**:1769-1775.
8. Liao S, Ferrell BA: **Fatigue in an older population.** *J Am Geriatr Soc* 2000, **48**:426-430.