Author's response to reviews

Title: Pilot comparative effectiveness study of surface perturbation treadmill training to prevent falls in older adults

Authors:

Jon D. Lurie (tamara.s.morgan@dartmouth.edu)
Alexandra B. Zagaria (alexandra.b.zagaria@dartmouth.edu)
Dawna M. Pidgeon (dawna.m.pidgeon@hitchcock.org)
Judith L. Forman (judith.l.forman@dartmouth.edu)
Kevin F. Spratt (kevin.f.spratt@dartmouth.edu)

Version: 3 Date: 2 April 2013

Author's response to reviews: see over
Dear editor,

Please find enclosed our revised manuscript. We wish to thank the reviewers for their thoughtful comments, which we feel have allowed us to substantially improve the manuscript. Below we catalog a point-by-point response to the issues raised along with a summary of the corresponding changes made to the manuscript.

Reviewer #1:

This manuscript describes an interesting and important pilot study to determine the feasibility of a randomised trial of perturbation training for the purpose of falls prevention.

Major compulsory revisions

1. The aims described in the abstract need to be made clearer. The nature of the trial/training and the setting is not mentioned. Also, I am not sure that ‘clinical relevance’ is appropriate, as no definition of clinical relevance regarding faller status has been provided. Perhaps the aim is to determine the effect size that might warrant further study?

We have altered the language of both the Abstract and the Background section of the main paper in order to try to present greater clarity and consistency of the study aims:

“The purpose of this pilot study was to determine the feasibility of conducting a large pragmatic randomized trial comparing a multidimensional exercise program inclusive of the surface perturbation treadmill training (SPTT) to multidimensional exercise alone (Standard PT) and 2) to assess fall outcomes between the two groups to determine whether an effect size large enough to warrant further study might be present.

We realize that “an effect size large enough to warrant further study” is not defined and largely subjective, but we feel that this is true of all pilot studies.

2. Provide an indication of the sample size in the abstract

We have added information on the number of participants to the abstract.

3. Describing the difference in proportion of fallers as ‘substantially reduced’ needs to carefully considered and given some perspective. I suggest comparing the size of this effect to that of other falls prevention interventions reported in the literature.

We have altered the description to maintain focus on the actual magnitude of the change without subjective description; although we continue to believe that a 33-50% reduction in event occurrence would reasonably be considered substantial by the majority of individuals, we agree that substantial is an irreducibly subjective term and we have therefore removed it.

Minor essential revisions
1. Was a 3 month retrospective question regarding falls events an appropriate method of falls surveillance? What is the reliability of this method?

We added the following paragraph to our discussion:

“A second major limitation was the reliance on patient recall at 3-month follow-up for identifying falls due to limited resources for this pilot study. While not entirely analogous to our method, 12-month recall compared to weekly post-cards to assess falls showed a sensitivity of 89% and specificity of 95%; sensitivity of 12-month recall for an injurious fall was 100% (Ganz, 2005 #42). Therefore, we likely missed some falls using a 3-month recall method but there is no reason to believe that these missed falls would have been systematically biased between groups. False positive all reports were unlikely to be a significant issue given the high specificity of recall. A more definitive study should try to improve ascertainment using diaries, more frequent monitoring, or both.”

2. Provide results of chi-square tests.

Now provided in the results section.

3. Were other PT objectives met, since the SPTT group had fewer sessions (presumably with time taken from normal therapy for perturbation training)?

We have added the following paragraph to the discussion:

“Another notable feature is that the average fall-risk scores in both treatment groups went from above the “at-risk” threshold levels at enrollment, to below these levels at the end of treatment. This indicates that PT goals were met in both the intervention and control groups; this is not unexpected as the therapists involved were experienced in gait/balance training. This finding does underscore, however, that the improvements in fall outcomes being evaluated were over and above those found from a robust, standard of care intervention.”

4. If relative risks are to be reported, the confidence intervals would be useful.

We have provided confidence intervals for RR, which now reflect analyses adjusted for baseline fall risk scores across groups.

5. Typo on line 3 of the Conclusion paragraph.

This has been corrected.

6. Define acronyms of outcome measures in Table 4.

This has been added to the bottom of the table.

Reviewer #2:

Overall comment: do not place the full stop in front of the references, but it should be placed after the last bracket (MER)

Thank you for pointing out this mistake, the citation formatting has been corrected.

ABSTRACT
1. It is stated that the second purpose is “to compare fall outcomes to determine whether a clinically relevant effect warranting further study might be present”. This purpose is not explicitly stated in the main text, neither is it clear how such a clinical relevance is evaluated (it is only stated that the study was underpowered to find differences). The authors should make sure that the purpose in the abstract and the main text is the same (MCR)

We have altered the language of both the Abstract and the Background section of the main paper in order to try to present greater clarity and consistency of the study aims:

“The purpose of this pilot study was to determine the feasibility of conducting a large pragmatic randomized trial comparing a multidimensional exercise program inclusive of the surface perturbation treadmill training (SPTT) to multidimensional exercise alone (Standard PT) and 2) to assess fall outcomes between the two groups to determine whether an effect size large enough to warrant further study might be present.

We realize that “an effect size large enough to warrant further study” is not defined and largely subjective, but we feel that this is true of all pilot studies.

2. “Falls were assessed by telephone interview at 3 months”. Unclear when this was, and when the falls would have happened. Is it better with “Falls during the 3 months intervention period were assessed by telephone interview after the intervention”? This must be described clearer in the text and described similarly here (MCR)

Patients were enrolled after an initial evaluation and start of PT. Follow-up was 3 months after enrollment in order to ensure equal follow-up time across the 2 groups from the time of randomization. The assessment period thus did include the period of treatment and we have tried to make that clearer in the revision.

We have also added the following to the methods:

“We assessed the proportion of patients in each group with any fall and the proportion with any injurious fall via phone interview at 3 months after enrollment. The early part of the assessment period thus was inclusive of the intervention which began at the time of enrollment. Thus, differences in fall outcomes between the groups could represent differences in effectiveness for preventing falls or more rapid attainment of similar effectiveness.”

4. “These results were not statistically insignificant but this pilot study was not powered for hypothesis testing”. Typing error? (MER)

This typographical error has been corrected.
4. The introduction is overall easy to read and gives a relevant background. However, there are several important terms that are not defined or explained properly. Examples are tripping and slipping, motor control learning, task-specific adaptive training, dynamic stability training and surface perturbation treadmill training. This makes it difficult to fully understand the Surface Perturbation Treadmill Training (SPTT) device and the theoretical context important to it. The terms must be described better (MCR)

We have altered the language in the introduction to try to improve clarity and included the following definitions:

- **dynamic stability** (the ability to maintain balance following a postural perturbation)
- **task-specific adaptive training** (practicing the actual motor skill of a defined task such as avoiding a fall after loss of balance)
- **tripping** (an unexpected deceleration of the foot causing truncal rotation in the direction of ambulation)
- **slipping** (an unexpected acceleration of the foot causing truncal rotation in the opposite direction of ambulation) are a major contributor to falls.

We hope these are useful, though it is not clear to use that these will be helpful to the majority of readers who are more likely to understand the word trip than its definition. It is worth noting that in our own review of over a dozen articles specifically on tripping and slipping (including several in BMC Geriatrics), these terms were not given specific definition within any of those articles.

5. What is meant by the statement "dynamic stability has emerged as a biomechanical paradigm....". This statement can be explained better or written differently (MER)

This sentence was re-written for improved clarity as follows:

“Recently, dynamic stability (the ability to maintain balance following a postural perturbation) has emerged as a potentially important aspect of defining and modifying fall risk (Mansfield, Phys Ther 2010).”

6. Hypothesis 3 is difficult to test in a pilot study? Also it does not reflect the purpose stated above (MCR)

We have reworded with for greater clarity as follows:

We hypothesized that: 1) SPTT could be incorporated into a multidimensional exercise program within a clinical PT practice; 2) patients would be generally accepting of randomization between the two treatment arms. We also sought to compare fall outcomes between the two treatment groups to determine whether an effect size large enough to warrant further study might be present.

**METHODS**
7. I miss some information about the therapists doing the intervention. How many, how experienced are they etc. Since the treatment seems heavily dependent on clinical reasoning and judgment, the effectiveness may rely on how good the therapists are. Especially in a larger study this should be explained and accounted for (MCR)

We have added the following additional information on the sites and the specific therapists involved.

“The study was conducted in 2008 and 2009 at two outpatient PT clinics. One was a physical therapy department at a small community hospital with 2 PTs general physical therapists; this department was an early adopter of the SPTT and had been using it clinically prior to the start of the pilot. The second site was a large academic medical centers and involved 2 therapies with specific training, expertise and practice focused specifically on fall prevention including both gait/balance training and vestibular rehab; this site had a well-established gait/balance training program but had not used the SPTT device prior to the start of the study.”

8. After reading about the SPTT in the methods section, I still feel I need to know more to fully understand it. The training needs to be described better and preferably with a picture to give a visual understanding of it as well (MCR)

We have expanded the explanation as follows:

“Treatment for the SPTT cohort included the same general protocol as for the Standard PT cohort, but with the addition of the surface perturbation treadmill training. Each treadmill training session was incorporated into standard visits at the therapist’s discretion. The SPTT sessions consisted of the subject, who was in a safety harness that would catch them before hitting the machine if they fell, standing or walking at comfortable speed on the microprocessor-controlled, stepper motor-driven treadmill. The training protocol consisted of postural disturbances that simulated a trip (sudden motion of the treadmill belts that resulted in forward-directed rotation of the subjects torso) or a slip (sudden motion of the treadmill belts that resulted in forward-directed rotation of the subjects torso). The magnitude of the disturbance for each trial is defined by four parameters: peak velocity, elapsed time to peak velocity, elapsed time during which the peak velocity is maintained, and time required to decelerate the treadmill belt to zero velocity. The magnitude of perturbation (amount and speed of sudden treadmill belt motion) was set for each trial by the therapist from level 1 (mildest) to level 5 (most vigorous) based on the patients day-to-day and trial-to-trial performance. In general, the magnitude of the disturbances during training sessions progressed from less to more challenging; however, the exact sequence was not established a priori.”

We do not have a picture that we think would be particularly useful, and a video of the training (which likely would be helpful) is beyond the scope of this manuscript.
9. The BBS and the TUG lacks references for the cut-off points used. Include references here (like you have done on the DGI and the ABC) (MER)

We have added the additional references

10. Is it right that the cut-off for fall risk on the BBS is 50? Is it not 45? Check this and reference the choice (MER)

This was a typographical error and was corrected.

11. Is the falls registered during the intervention period? In that case does it necessarily reflect the effect of the intervention? The period for the falls must be described more precisely (MCR)

We have also added the following to the methods:

“We assessed the proportion of patients in each group with any fall and the proportion with any injurious fall via phone interview at 3 months after enrollment. The early part of the assessment period thus was inclusive of the intervention, which began at the time of enrollment. Thus, differences in fall outcomes between the groups could represent differences in effectiveness for preventing falls or more rapid attainment of similar effectiveness.”

12. What about power analysis? In the abstract and discussion it is stated that this study was "underpowered" to find differences (even though one of the hypothesis was to test this). Describe on what grounds the power was evaluated? (MER)

We have added the following paragraph to the discussion:

“One major limitation of this study was that, as a pilot, in was not adequately powered to detect important changes in fall outcomes. A sample size calculation done for planning a larger more definitive trial showed that with a baseline fall risk in the control population of 33%, similar to what we found in this pilot, to obtain 80% power with a type one error rate of 0.05, 106 subjects would be needed to detect a 50% reduction in fall risk, 320 to detect a 30% reduction in fall risk, and 750 to detect a 20% reduction in fall risk. Thus a definitive study would need to be designed for between 500 and 1000 subjects.”

RESULTS

13. Why was the 64 participants not divided equally in the two groups (50/50). Why was 33 allocated to one group and 31 to the other? (MER)

The groups were not completely balanced because with random allocation, some degree of imbalance is expected, (random processes rarely yield completely equal groups). We sought to limit the degree of imbalance using permuted block randomization which will yield completely equal groups if enrollment is ended precisely at the end of a block. However as this was a pilot we enrolled as many participants as possible within the time and resources available and did not seek to arbitrarily cut off enrollment at the end of a block as there would be no benefit to the analysis from enrolling fewer patients in order to have completely equal groups.
14. An important (and impressive) finding is that both treatment groups went from being at fall risk to non-fall risk (based on the cut-offs of the measures). This can be highlighted. (MER)

We agree, thank you for the suggestion. We have added the following paragraph to the discussion:

“Another notable feature is that the average fall-risk scores in both treatment groups went from above the “at-risk” threshold levels at enrollment, to below these levels at the end of treatment. This indicates that PT goals were met in both the intervention and control groups; this is not unexpected as the therapists involved were experienced in gait/balance training. This finding does underscore, however, that the improvements in fall outcomes being evaluated were over and above those found from a robust, standard of care intervention.”

15. Table 1: Ok. Would it be an idea to include the content of the home-exercise here? (DR)

The home exercise component incorporated the listed features and consisted of continuations of the various strengthening and flexibility, static and dynamic balance, mobility training interventions at the therapists discretion and were not a separate component per se.

16. Table 2: Ok. Gender should maybe come as the actual number of females with percentage in brackets? (DR)

We have adjusted this and found a typo in the % female in the prior table, which was also corrected.

17. Table 3 lack units. Why is the N in this table different from the other tables? This is also only stated and not explained in the text. What about the home-exercise? Is the time and content of these exercises described somewhere? Was “dosage” of home-exercises equal in both groups? (MCR)

We have added the units of time (minutes) to the table, the units for number of session is the just that and we are unclear how to label it any differently. We have expanded the explanation of the different “N’s” in the text

“To this end, the number of treatment sessions and session length in each group were evaluated through review of medical records. Medical records were available for 25/26 of the SPTT and 26/33 of the Standard PT subjects.”

We also added further explanation of the role of the home-exercise component across groups.

“The components of the intervention that were prescribed for home exercise between visits were individualized by the therapist but did not systematically differ across groups. We had no mechanism for monitoring compliance with the home exercise component of the intervention.”
18. Table 4: There seem to be an error in the results from the BBS for the Standard PT group. The difference should be -0.54 and not 4.46? (MCR)

Thank you for picking up this error, it has been corrected.

19. Figure 1: You now use follow-up and analysis. Would it be better with Pre-analyses, intervention, post-analyses? It is for me still unclear in what 3 months period falls were assessed (MCR)

The categories of follow-up and analysis are not ours, they are the recommended structure from the CONSORT statement and these refer to numbers of patients with data available for follow-up and the number included in the analysis. Since this is standardized according to the CONSORT statement we think using other categories would add rather than decrease confusion.

As discussed above, we have tried to make the timing of follow-up clearer in the text which hopefully has addressed the reviewers concerns.

DISCUSSION

20. The groups were quite different in relation to gender (62.8% in the standard group vs. 37.1% in the SPTT group). Since female gender is a risk factor for falls (especially for injurious falls) this is important to discuss in relation to the finding that there were more falls in the group with the highest number of females. In a future RCT, it is possibly important to have similar groups in terms of gender. Also, if I haven’t misunderstood the falls that happened during the intervention period were registered, and then it should really be discussed whether it is actually an effect of the intervention we are looking at. The use of retrospective fall registration and its limitation should also be described (MCR)

We apologize for the typo in table 2; the actual percentages were 50% and 67% and these were not above the differences expected by randomization in groups of this size (p=0.20) The percentages will be more similar in a larger trial just based on the mathematics of randomization in larger groups.

Also, we added the following paragraph to our discussion:

“A second major limitation was the reliance on patient recall at 3-month follow-up for identifying falls due to limited resources for this pilot study. While not entirely analogous to our method, 12-month recall compared to weekly post-cards to assess falls showed a sensitivity of 89% and specificity of 95%; sensitivity of 12-month recall for an injurious fall was 100% [Ganz, 2005 #42]. Therefore, we likely missed some falls using a 3-month recall method but there is no reason to believe that these missed falls would have been systematically biased between groups. False positive all reports were unlikely to be a significant issue given the high specificity of recall. A more definitive study should try to improve ascertainment using diaries, more frequent monitoring, or both.”
21. The contribution from the therapists in the study should be discussed (see also comment number 7) (MCR)
We have added the following paragraph to the discussion:

“Another notable feature is that the average fall-risk scores in both treatment groups went from above the “at-risk” threshold levels at enrollment, to below these levels at the end of treatment. This indicates that PT goals were met in both the intervention and control groups; this is not unexpected as the therapists involved were experienced in gait/balance training. This finding does underscore, however, that the improvements in fall outcomes being evaluated were over and above those found from a robust, standard of care intervention.”

22. Since this is a pilot study in advance of a larger RCT study, number of participants needed in the RCT should be discussed (preferably based on a power calculation) (MER)
We have added the following paragraph to the discussion:

“One major limitation of this study was that, as a pilot, it was not adequately powered to detect important changes in fall outcomes. A sample size calculation done for planning a larger more definitive trial showed that with a baseline fall risk in the control population of 33%, similar to what we found in this pilot, to obtain 80% power with a type one error rate of 0.05, 106 subjects would be needed to detect a 50% reduction in fall risk, 320 to detect a 30% reduction in fall risk, and 750 to detect a 20% reduction in fall risk. Thus a definitive study would need to be designed for between 500 and 1000 subjects.”

ACKNOWLEDGEMENT AND COMPETING INTEREST

23. What is now written under acknowledgement should possibly be under competing interest instead? (DR)
We have made this change as requested.