Author's response to reviews

Title: The effectiveness of pelvic floor muscle training alone and with other physical therapies for the treatment of stress urinary incontinence in women: a systematic review

Authors:

Patricia B Neumann (cpneumann@ozemail.com.au)
Karen A Grimmer (Karen.grimmer@unisa.edu.au)
Yamini Deenadayalan (Yamini.deenadayalan@unisa.edu.au)

Version: 3 Date: 14 February 2006

Author's response to reviews: see over
We provide herewith our response to the reviewers’ reports of Neumann et al. “Pelvic floor muscle training and other physical therapies for stress urinary incontinence in women: a systematic review.”

A point-by-point response is detailed below. Amendments to the original text have been underlined.

Response to Dr Herbison’s report

Point 1.
“I am uncertain about the use of levels of evidence, that are supposed to apply to all evidence, to individual studies. For example, no study can have level I evidence as that is evidence from a systematic review. I am not sure how much this matters, as it could be worked around, by saying that this study would contribute to level x evidence. I would remove the levels from the abstract at least and be clear that these are the Australian ones, as there are many others.”

• we are not sure what point he is trying to make about the levels of evidence applying to all evidence. To our knowledge there has been no reference to ‘level I studies’ in this review since it is stated in Appendix 3 that the Australian National Health and Medical Research Council (1999) guidelines for assessing levels of evidence were used and the criteria strictly followed. As Dr Herbison states, level 1 is evidence from systematic reviews (SRs). According to the inclusion criteria, SRs were not considered and so have not been referred to in the review.

• reference to the levels of evidence has been removed from the abstract: It reads now as follows (p4, line 1-2):

“Twenty four studies, including 17 RCTs and seven non-RCTs, met the inclusion criteria.”

Point 2
“The authors say that this is a study looking at “effectiveness” rather than “efficacy” but then say that time scale is not being considered. I would find it hard to consider effectiveness without regard to the timeframe over which treatment may work. Just about all of the studies measure the outcome at the end of treatment. For effectiveness you would have to assume that the treatment would continue or that the effect would remain. I am not convinced that either of these will happen.”

• to our knowledge we have not stated that ‘time scale’ is not being considered. In fact, one of the stated objectives was to investigate the optimal treatment period and number of treatments (research question 4, page 8)
• as further evidence of our interest in the time scale, it is stated for each study in Tables 4-10.
• time scale (or duration of treatment) is considered as an important factor affecting outcome and is discussed at some length (p45, line 17-p46, line 9).
• the point is taken that the distinction between efficacy and effectiveness research may need to be more clearly expressed in the SR and the text has been amended accordingly (p6, line 17-22):

“A distinction is to be made between the terms ‘efficacy’ and ‘effectiveness’ (Brook 1985). Efficacy is defined as “the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use”. By contrast, effectiveness is considered to have all the attributes of efficacy but to reflect “performance under ordinary conditions by the average practitioner for the typical patient” (Brook 1985). “

• it is conceded that most of the studies are in fact efficacy studies ie RCTs but effectiveness studies ie lower level studies are included as well and this point has been made in the text eg p7-9.

• The Abstract Conclusion has been modified to incorporate the reviewer’s concerns about the distinction between effectiveness and efficacy: (page 4, lines 11-14):

“There is strong evidence for the efficacy of physical therapy for the treatment for SUI in women but further high quality studies are needed to evaluate the optimal treatment programs and training protocols in subgroups of women and their effectiveness in clinical practice.

• and on p 38, line 22:

“Question 1: What is the evidence of the efficacy or effectiveness of PFMT.,”

• other references to ‘effectiveness’, where both ‘efficacy’ and ‘effectiveness’ are implied, have been amended to read “efficacy or effectiveness” or simply ‘evidence’

• the point which the reviewer makes about effectiveness and having “to assume that the treatment would continue or that the effect would remain” is not clear and has not been raised by other authors [1-3]. In fact Jefford (2003) states that effectiveness studies are concerned with issues of health outcome “at the end of treatment”: 
"Outcomes assessment (health care) is defined as: research aimed at assessing the quality and effectiveness of health care as measured by the attainment of a specified end result or outcome."

Point 3

“I find it hard to be sure that they have a comprehensive group of trials. They have not followed the QUORUM statement in this, as that says you must say how many trials were screened, how many excluded and the reasons for this. For example they exclude Glazener et al 2001 and Glazener et al 2005. I am sure there would be others that people might wonder why they are not included”

- the Joanna Briggs Institute guidelines for systematic reviews were used as a basis for this review [4].
- the protocol followed the principles of the QUORUM statement as it was stated that 7760 possible papers were identified and considered for inclusion according to the criteria. The reason that the final 24 studies were included was based on the fact that they fulfilled all the inclusion criteria, as stated in Appendix 2.
- the other 7736 studies, including Glazener 2001 and 2005 did not fulfil the criteria and so were excluded. These two studies were of populations with mixed incontinence and only populations with stress urinary incontinence were included, as clearly stated in the inclusion criteria (p10 line 22).
- However, to clarify the inclusion criteria the following has been added (p11, line 7):

  Studies were excluded if they included subjects with mixed UI or detrusor overactivity because of the assumption of a different underlying pathology and thus rationale of treatment, even if outcomes for subgroups of women with SUI were reported.

- the statement “I am sure there would be others that people might wonder why they are not included” seems a subjective comment which is not substantiated by the reviewer. We believe that the inclusion criteria are clearly stated and that others can deduce why excluded papers were properly excluded.

Point 4

“There is an overemphasis on the quality score of the studies. The one thing that is certain about quality scores is that they can give almost any answer when used in a systematic review. (see Juni et al: The hazards of scoring the quality of clinical trials for meta-analysis JAMA 1999). The particular quality score chosen has some elements that are highly unlikely to be associated with bias in the outcome, or the difference in outcomes between the groups in the studies. For example, whether a study reports
“ethical processes”, “consideration of sample size” or “literature review relevant” or not will not affect the internal or external validity. This is not to say that these things are not worth doing or reporting, just that they will not introduce bias. Thus there is the potential for a study that has a high quality score to still produce a more biased results than a study with a lower quality score. Studies in a meta-analysis should be assessed for quality, but this should not be by way of a score.”

- The quality review process was particularly important in this SR because trials other than RCTs were included. It was deemed important to assess whether non-RCTs were necessarily of poorer quality than RCTs in order to justify their inclusion. This point was discussed in the review (p38) and it was found that indeed some non-RCTs were of superior quality, based on this rating scale, than some of the RCTs. The emphasis on quality scores was intentional to provide a better understanding of quality in this SR as we may have been criticised for including non-RCTs.

- The quality score used was closely based on the tool developed by the Occupational Therapy School, McMaster’s University [5] and with reference to the guidelines produced by the University of York [6] as stated in the methods. Use of the tool developed by Law et al (1998) is a requirement for systematic reviewers for the Joanna Briggs Institute and so is in common use, particularly by Allied Health reviewers [7].

- the quality score used was closely based on the tool developed by the Occupational Therapy School, McMaster’s University [5] and with reference to the guidelines produced by the University of York [6] as stated in the methods. Use of the tool developed by Law et al (1998) is a requirement for systematic reviewers for the Joanna Briggs Institute and so is in common use, particularly by Allied Health reviewers [7].

- the AHRQ (Agency for Healthcare Research and Quality) report #47 states that “more than 100 sources of information on systems for assessing study quality and strength of evidence for systematic reviews” were summarised and that “after applying evaluative criteria…they identified 19 study quality and seven strength of evidence grading systems that those conducting systematic reviews…can use as starting points”. Thus Dr Herbison’s criticisms of the quality scale fail to acknowledge that there are many scales for rating quality and that there is currently no agreed gold standard.

- it is acknowledged that some of the items on the quality scoring scale, as detailed by the reviewer, do not contribute to assessment of bias. However, as there is no agreed gold standard instrument [8] and as other items such as “ethical processes”, “consideration of sample size” or “literature review relevant” were included in other scales eg Law’s, they were deemed suitable for inclusion in this systematic review.

- studies of high methodological quality and which had little evidence of bias were discussed by the authors of this review and were highlighted for the strength of their evidence, as discussed in the AHRQ report #47.

- Dr Herbison states that “there is the potential for a study that has a high quality score to still produce a more biased result than a study with a lower
quality score”. This point is acknowledged but has been addressed by the comprehensive nature of our rating scale which assessed issues of bias as well as other issues such as “ethical processes”, “consideration of sample size” or “literature review relevant”, which we acknowledge do not contribute to the assessment of bias. Thus readers were provided with details considered by some authors to be important in quality assessment (eg Law et al 1998) and not only with issues relating to assessment of bias. Thus a high score in this review assumed good control of bias as well as the presence of other issues of quality reporting eg generalizability.

- Herbert and Bo 2005 (BMJ) state that “the QUORUM statement on reporting of meta-analyses of randomised trials exhorts reviewers to describe details of the intervention provided in each trial”. They continue that they believe that in addition “reviewers should explicitly assess the quality of interventions”.

- the authors state on p38: “It is acknowledged that the methodological quality of the critical review tools themselves may have incorrectly reflected the quality and ranking of the included studies.” While every effort was made to describe the assessment tool and the possible shortcomings of the review process, it was not the intention of the reviewers to critically appraise the assessment process per se. This would seem to be beyond the scope of this review and not common practice when comparisons with other systematic reviews were made (Bialecerkowski et al 2005)

- the reviewer states that “Studies in a meta-analysis should be assessed for quality, but this should not be by way of a score” although others have described checklists and scoring systems which involve numbering eg PEDro, Downs and Black et al 1998. Black et al developed a scoring system (The Quality Index) precisely to provide “an overall score for study quality and a profile of scores not only for the quality of reporting, internal validity ..but also for external validity”. Other systematic reviews have been published with the quality scores for each study providing the informed reader with a concise overview of fulfilment of each quality item [7].

Point 5

“In the introduction, the authors say they want peer-reviewed reports, but this is not mentioned in the inclusion/exclusion criteria. Some of the abstracts will have undergone a cursory peer review.”

- this criticism is acknowledged and has been addressed by amending the inclusion/exclusion criteria (P10, line 16).

“Only peer-reviewed studies published in English in the last decade (1995-2005) were included in this review.”
Point 6
“In the methods of the review, they say about assessment of quality but then don’t say how they use this.”
• this criticism is acknowledged and has been addressed by amending the methods to state: (p15, line 13-p16,line 2)

“Details of the quality assessment are provided in Appendices 4 and 5 with studies ranked according to their quality assessment score to provide readers with an overview of their methodological quality. All the studies were then considered for the strength of their evidence based on the quality score and with particular consideration of the factors which were concerned with control of bias. Studies with a high quality score were considered to show evidence of good control of bias (e.g., attention to random allocation processes, baseline similarity of groups, reliable outcome measures) as well as other factors concerning quality reporting, such as consideration of ethical processes and relevance of the literature review. Studies with a high quality score are identified and highlighted by the reviewers in the text for their contribution to evidence about treatment outcomes.”

Point 7
“There is an element of study counting when they report number of outcomes reported and number significant. Many studies selectively report outcomes that are statistically significant and this should be discussed”
• analysis of studies in this SR was a narrative summary, as stated in the methods, and as such, a descriptive analysis of the included studies was appropriate
• strength of evidence is assessed by three factors according to the AHRQ report #47 and we quote:
  o “quality: aggregate of quality ratings for individual studies, predicted on the extent to which bias was minimised
  o quantity: magnitude of the effect, number of studies and sample size or power
  o consistency: for any given topic, the extent to which similar findings are reported using similar and different study designs.”

Thus our method of ‘study counting’ would not seem inappropriate.
• this SR has reported the percentage of outcomes in each study which were statistically significant and thus we infer that a percentage were not statistically significant. We considered all the outcomes presented in the methodologies and cannot concur with Dr Herbison that the studies included in this review have selectively reported the statistically significant findings.
• the percentage of outcomes which were statistically significant has been consistently reported eg p24 and p25. We found that a high number of outcomes were statistically
significant and have discussed this as evidence of strength of evidence, as mentioned above.

Point 8

“I have some issues with the section on compliance. Of course PFMT is unlikely to be effective if people don’t do it, but the effect may well be different in those who do it out of choice and those who are ‘encouraged’ to do it, so it is not just a simple matter of saying that if they did all the training then this would be the effect.”

- it was not the aim of this systematic review to provide a comprehensive evaluation of the issue of compliance with PFMT as this has been the subject of a previous review [9] but primarily to identify possible factors affecting outcome. Compliance was described by some authors of the papers reviewed and so was included in the discussion with particular reference to the poor quality of reporting. No statement was made to imply that it was simply a matter of saying that if they all did the training then it would be effective. We made the comment that where outcomes were poor and compliance was not measured, that lack of compliance may have contributed to the poor result.
- In order to address the reviewers concern that compliance with PFMT should be given more weight, the discussion has been amended (page 44, line 19) (as stated below) and ‘compliance’ has been added to the ‘points for further research’.

“A. Compliance

Another factor which may influence outcome is the degree to which subjects actually comply with the treatment program prescribed. Compliance with PFMT is a complex issue and has been the subject of a previous review [9]. The terminology is not agreed as some authors consider ‘adherence’ to be a more appropriate term implying voluntary co-operation rather than coercion. [9, 10]. Subject compliance or adherence was infrequently and generally poorly reported with no standardised, validated or reliable approach to its assessment. However it would appear to be of considerable importance in any PFMT program which depends on subjects performing exercise in order to effect physiological changes. There are complex psycho-social issues involved in interventions which demand that women commit time and effort on a regular basis to training [9, 10]. It is likely in the high quality studies with good outcomes that subjects adhered to the treatment protocol. However, in studies which reported poorer outcomes and also did not report subjects’ compliance, it is not possible to say whether an ineffective intervention or the subjects’ lack of compliance was responsible for the poor result.”
“B. research is needed into:

- (page 48, line 23)...“the factors which influence a subject's likelihood of attending appointments, continuing with treatment and complying with the home training program”

Point 9

“In the implications for practice section, second bullet point it should be made clear that they are not saying that BF, ES etc may still have a place in clinical practice as stand alone therapies, but on the evidence in this review only as adjuncts to PFMT.”

- the reviewers concerns have been addressed and the point amended (page 47, line 15):

“no benefit was found in this review in adding BF, ES and abdominal muscle training to a PFMT protocol but the external validity of these results yet to be established. Thus these interventions may still have a place in clinical practice as adjuncts to PFMT in particular populations of women with SUI.”

Point 10

“Implications for research, bullet point 2. Do they mean effectiveness or efficacy?”

- the reviewer’s comment is acknowledged and the point amended to read (Page 48, line 20):

“Research is needed into:

  o ..the longer term efficacy and effectiveness of physical therapies..”

Point 11

“In Table 1, why is age presented as the only participant information. For stress incontinence it is likely that parity is even more important – and I am sure there are other patient characteristics related to incontinence that could be presented (eg BMI).”

- it is acknowledged that many more patient characteristics could have been presented but due to word/room constraints, we based our report on the previous SR on urinary incontinence by Hay-Smith et al (2001) published in the Cochrane Library which also provided details of age as the only patient characteristic.

Point 12

“Figure one should be replaced by a 2_D plot. These are much easier for the reader to interpret”

- the 3_D plot has been replaced with a 2_D plot as suggested.
Point 13
“In the appendix with the quality score the abbreviation OM is not explained.”

- this point has been addressed by the addition of an explanatory key for both relevant appendices 4&5.

Point 14
“While most of the provisions of the QUORUM statement are met there are some that could be met without making large changes eg including the words ‘systematic review’ in the title.”

- this point has been addressed by amending the title as follows:

“Pelvic floor muscle training and other physical therapies for the treatment of stress urinary incontinence in women: a systematic review”

- note that the word “effectiveness” has been been deleted from the title.

Point 15
“I think I should say that I am involved in the Cochrane collaboration and think that as many systematic reviews as possible should be published in the Cochrane Library. I think that this review would require some tidying up before being acceptable for the Cochrane Library”

- While we acknowledge that the Cochrane Library is an important forum for publishing SR of efficacy studies, it was explicitly the authors’ intention NOT to publish this review in the Cochrane Library because of their narrow interpretation of SR ie to only include RCTs. We intended the review to be broader and more inclusive of other possible evidence, as discussed at some length in the paper.

Additional changes have been made in accordance with the recommendations for Minor Essential Revisions by the other two reviewers:

eg page 48, line 1 :the Implications for Practice dot point no 5 has been amended to eliminate a statement conflicting with the main text:

“changes in incontinence outcomes were demonstrated after treatment duration of one week to six months. Improvements in PFM strength may be demonstrable after six weeks of strength training but up to six months or more may be needed to optimise physiological changes.

1. Page 9, line 12: the paragraph has been amended to be more precise and to clarify the terminology with efficacy and effectiveness:
In order to investigate both the efficacy and effectiveness of PFMT and other physical therapies for female SUI, RCTs and prospective research designs other than RCTs were also considered in this review.

Any PFMT program i.e. pelvic floor muscle exercises, with application of a specific PFMT protocol together with any combination of adjunctive therapies: biofeedback (BF), electrical stimulation (ES), vaginal weights or cones (VW) were included.

2. **page 11, line 5:** minor correction to text:

   “Any PFMT program i.e. pelvic floor muscle exercises, with application of a specific PFMT protocol together with any combination of adjunctive therapies: biofeedback (BF), electrical stimulation (ES), vaginal weights or cones (VW) were included.”

3. **page 42, line 5:** inclusion of 2 sentences in the discussion to clarify the recommendations for strength training, which appeared in the dot points for clinical implications but not in the text:

   “Although strength changes were demonstrable after four weeks of PFMT, the most rigorous method of measuring PFM strength failed to show any change in strength after eight weeks of training [11]. Longer training times should therefore be considered to optimise physiological muscle changes [12].”

References: