Abstract: Background: Although many studies have reported the effectiveness of pelvic floor muscle training for treating female urinary incontinence, the magnitude of the effect and the optimal configuration of the parameters of the training have not been clearly determined.

Objectives: The aims of this meta-analysis were (a) to calculate the effect size of pelvic floor muscle training compared to no-treatment on incontinent episodes, urine leakage amount, and perceived severity of urine loss, and (b) to identify parameters of PFMT and subjects’ characteristics influencing on the magnitude of the effects.

Method: The search for relevant literature published from 1980 to 2005 consisted of using several computerized databases, citation searching, and footnote chasing. Twelve studies met the inclusion criteria were reviewed and coded.

Results: The overall mean weighted effect size on incontinent episodes, urine leakage amount, and perceived severity were -0.68 (Z = 5.89, p = .000), -1.48 (Z = 2.64, p = .008), and -1.66 (Z = 1.68, p = .092) respectively. The studies with women having stress urinary incontinence showed a mean weighted effect size of -0.77 (Z = 7.03, p = .000), whereas studies with women having any type of urinary incontinence showed -0.47 with (Z = 4.40, p = .000). The mean weighted effect size for studies including subjects over 60
years mean age was -0.54 ($Z = 6.21$, $p = .000$), whereas that of studies in which the average age was younger than 60 years was -0.94 ($Z = 6.58$, $p = .000$).

Discussion: The treatment effect of pelvic floor muscle training on the incontinent episodes may be greater in younger women with only stress urinary incontinence. It appears that the number of daily contractions and the length of training period are not related to effect sizes on condition that training includes at least daily 24 contractions and keeps for at least six weeks.
August 26, 2006

Molly C. Dougherty, PhD., RN
Editor, Nursing Research
School of Nursing CB#7460
UNC-Chapel Hill
Chapel Hill, NC 27599-7460

Dear Dr. Dougherty:

I am submitting the manuscript titled “The Effects of Pelvic Floor Muscle Training on the Treatment of Women with Urinary Incontinence: A Meta-Analysis of Randomized Controlled Trials” to the Nursing Research which we believe will be of interest to Nursing Research readers. This manuscript conforms to the format requirements and includes 16 page-long text, one table and seven figures.

My coauthors and I do not have any interests that might be interpreted as influencing the results of the study, and have followed APA ethical standards in conduct of the study. All of the authors agreed to byline order and to submission of the manuscript in this form. We also agreed that I will be serving as the corresponding author for convenience, as I am located in United States and I mentor my coauthors. Dr Choi was a visiting scholar last year and Ms. Park is my doctoral student advisee.

If you have any questions regarding this manuscript, please do not hesitate to contact me by phone or e-mail. Thank you for your time and attention.
Sincerely,

Mary H. Palmer, PhD, RNC, FAAN
Helen W. & Thomas L. Umphlet Distinguished Professor in Aging
UNC - Chapel Hill CB #7460 Chapel Hill, NC 27599
919-966-7204 (voice) 919-966-7298 (fax)
The Effects of Pelvic Floor Muscle Training on the Treatment of Women with Urinary Incontinence: A Meta-Analysis of Randomized Controlled Trials

Heejung Choi, PhD, RN, Associate Professor
Konkuk University, Korea

Mary H. Palmer, PhD, RN, C, Professor
mhpalmer@email.unc.edu

and Jeongok Park, MS, RN, Doctoral Candidate
The University of North Carolina, Chapel Hill
Acknowledgement

This study was performed by financial support by Konkuk University, Chungju, Republic of Korea. And Konkuk University requested authors to identify that “This work was supported by the Konkuk University.”
Introduction

Urinary incontinence (UI), defined by the International Continence Society (ICS) as “the complaint of any involuntary leakage or urine,” is a common but underreported health problem among women (Abrams et al., 2002). The prevalence of female UI varies from 5% to 66% depending on UI definition, methods of measurement, age of women included in the study, and setting (e.g., general or institutionalized population). Evidence exists that UI may seriously affect the physical, psychological, and social well-being of women (Hunskaar et al., 2005). UI has substantial impact. The annual direct cost of urinary incontinence in women was estimated to be $12.4 billion (1995 dollars) as compared to $3.8 billion (1995 dollars) in men (Glazener et al., 2001). Because of its prevalence and impact on women’s well-being, and its high financial costs, UI is a major health problem that requires serious attention from health professionals.

Three major types of female UI treatment are: pharmacological, surgical, and conservative. It is beyond the scope of this paper to discuss pharmacological and surgical treatments and the reader is referred to the 3rd International Consultation on Incontinence for in depth discussions of these topics (Abrams et al., 2005).

Conservative treatment, also called behavioral therapy, includes physical therapies, lifestyle interventions, and scheduled voiding regimens (Wilson et al., 2005). A physical therapy intervention, pelvic floor muscle training (PFMT), is a frequently used treatment modality for female UI. Several studies have reported the effectiveness of PFMT for female UI (Berghmans et al., 1998; Hay-Smith et al., 2001; Hay-Smith & Dumoulin, 2006; Holroyd-Leduc & Straus, 2004). Their findings are consistent, that is PFMT is an effective treatment for many women with stress UI. Unfortunately almost every published clinical trial has not precisely described all the parameters of PFMT such as the number, frequency, and velocity of pelvic floor muscle (PFM)
contractions, the length of training period, and type of instruction provided. In addition, considerable variation exist in outcome measures and whether PFMT was applied as the sole intervention or in combination with other modalities such as biofeedback, electrical stimulation, and bladder training (Hay-Smith & Dumoulin, 2006). Moreover, previously published systemic reviews (Berghmans et al., 1998; Bo, 1996; Bo & Maanum, 1996; Holroyd-Leduc & Straus, 2004; Wilson et al., 2002) relied on qualitative synthesis, thus the effect size specific parameters were not reported. For these reasons, the magnitude of the effect for many PFMT protocols and the optimal configuration of the parameters of PFMT have not been clearly determined (Gormley, 2002; Hay-Smith et al., 2001; Hay-Smith, Herbison, & Morkved, 2004; Holroyd-Leduc & Straus, 2004).

In an attempt to remedy this situation, Hay-Smith et al. (2001) tried to calculate effect sizes and to compare the effectiveness of different PFMT protocols with studies published in from 1980 to 2000. These authors calculated effect sizes of PFMT compared with various control treatments on perceived cure/improvement and on incontinent episodes. They did not, however, compare the effect of specific parameters of PFMT protocols on outcomes, no doubt in part due to variability in PFMT protocols used and that relatively few published studies specified which PFMT protocol parameter or combination of parameters were most effective on outcomes.

Recognizing the need for this information, Wilson and colleagues (2002) extrapolated evidence from exercise science literature and made recommendation for specific parameters including: minimum daily number, velocity, and frequency of pelvic floor muscle (PFM) contraction and a minimum length of training. To further advance and refine evidence based practice for conservative treatment for female urinary incontinence, we conducted a meta-analysis with published PFMT clinical trials.
Purpose, Research Questions, and Variable Definition

The primary aim of this meta-analysis was to calculate the effect size of PFMT compared to no-treatment on two objective outcome measures: incontinent episodes and amount of urine leakage and on one subjective outcome measure: perceived severity of urine loss.

Because agreement about what constitutes an effective PFMT protocol does not exist (Berghmans et al., 1998; Hay-Smith et al., 2004; Wilson et al., 2002) a secondary aim was also defined. It is to identify the effect sizes for each parameter and then to determine the most effective configuration of these parameters in treating female UI.

The tertiary aim was to identify sample characteristics acting as moderators on the effect of PFMT. Potential sample characteristics moderators include type of UI (McDowell, Burgio, Dombrowski, Locher, & Rodriguez, 1992) because PFMT was originally designed to reduce urine loss from inadequate muscle tone and decreased urethral resistance (DeLancey, 1988). Age is also a very important sample characteristic as conflicting findings related to age appear in the literature. For example, some researchers (Theofrastous et al., 2002; Wyman, Fantl, McIlhish, Bump, & the Continence Program for Women Research Group, 1998) found that there was no significant relationship between response to PFMT and subject age. Other researchers (Wilson, Al Samarrai, Deakin, Kolbe, & Brown, 1987) reported successful treatment was more likely in younger women. In other study, treatment responders were older than borderline responders (Bo & Larsen, 1992). In recent studies, it was suggested that the effect of PFMT may be greater in younger women, however, no compelling evidence exists that older women benefit less from PFMT (Hay-Smith & Dumoulin, 2006; Wilson et al., 2005). This latter conclusion was based on only a few investigations and effect sizes were not reported, therefore more analysis is required to investigate the association between age and PFMT outcome.
The research questions guiding the meta-analysis are as follows: (a) to what extent is PFMT effective on incontinent episodes, urine leakage, and perceived severity when compared to no-treatment? (b) What is the effect of PFMT treatment parameters and which configuration of treatment parameters is most effective on outcomes outlined above? (c) Does the type of UI and subject’s age moderate the effect of PFMT on outcomes listed above?

PFMT has been defined as a program of repeated PFM contraction taught by a health care professional (Wilson et al., 2002). This terminology is important as training implies repeated exercise over time whereas the word, exercise, refers to one episode of training (Wilson et al., 2002). Outcome variables in this meta-analysis are defined as follows: incontinent episodes are the number of times urine is lost; urine leakage is defined as amount of urine lost with each episode; and perceived severity defined as subjective severity of urine loss which means how much severe women felt their symptom of urine loss.

Methods

Searching for Relevant Literature

The search for relevant literature published from 1980 to 2005 consisted of using several computerized databases, citation searching of the Social Sciences Citation Index (SSCI) and the Science Citation Index (SCI), and footnote chasing.

The computerized database searches used in this analysis were Medline, Cumulative Index for Nursing and Allied Health Literature (CINAHL), University Microfilms International (UMI) Dissertation Information Service, and KoreaMed. In the case of search system that has the function of selecting options to limit retrieval, the investigator set limits to studies published in the English or Korean language and to only randomized controlled trials. In addition, the following key words were used: “pelvic floor muscle,” “Kegel,” “urinary incontinence,”
“conservative treatment,” “behavioral therapy,” and “women.” These computerized searches revealed 122 studies including 25 dissertations. Two authors reviewed the abstracts of all 122 studies independently with following inclusion criteria: (a) study design was randomized controlled trial (RCT), (b) PFMT was used as an UI intervention, (c) control groups limited to only no-treatment groups or placebo PFMT groups, (d) at least one of the selected outcome variables, incontinent episodes, urine leakage amount, and perceived severity was measured. The two authors agreed that 33 abstracts met inclusion criteria. Additional citation searching and footnote chasing added three more original research articles, thus 36 research articles were pulled for complete review.

**Selection of Studies**

The first and third authors reviewed the 36 studies to independently confirm that they met inclusion criteria. Twenty four studies were excluded with following reasons: (a) five studies applied PFMT as a preventive measure instead of treatment measure, (b) one study (Miller, Ashton-Miller, & DeLancey, 1998) used “The Knack” which was defined as intentionally contracting the PFM before and during a cough instead of usual PFMT, (c) seven studies had control groups that were neither no-treatment groups nor placebo PFMT groups, (d) five studies reported outcomes not using mean and standard deviation which could be used for calculating mean difference effect size and mean weighted effect size (MWES), (e) one study (Glazener, Herbison, MacArthur, Grant, & Wilson, 2005) reported on only six-year follow-up effect of PFMT, (f) one study (Lagro-Janssen, Debruyne, Smits, & van Weel, 1992) was excluded because the sample and reported results substantially overlapped with a study by the same authors already included for this meta-analysis, (g) finally, four studies were excluded because quasi-experimental rather than randomized controlled trial design was used.
Data-Collection Methods

The authors created a codebook on selected variables. Each study was coded for five dimensions: (a) report identification including author, publication year, study setting (country) and coder of study; (b) subjects parameters including sample size, mean and range of age, types of UI, and baseline perceived severity of UI; (c) methodology including assignment of subjects, blindedness, and attrition; and (d) treatment parameters including length of PFM training period, the number of daily PFM contractions, velocity of contractions, holding time in contractions, combined treatment measure with PFMT, and time to outcome measure; and (e) information for effect size including mean and standard deviation. In addition to objectively assess study quality, the criteria for quality of randomized controlled trials developed by Jadad and colleagues (1996) were adopted. These criteria had three scoring items including description of randomization, double blinding and withdrawals/dropouts. Scores for study quality range from 0 to 5.

At the beginning of the coding process, the first author and the third author coded two studies concurrently to improve agreement on coding decisions. The next ten studies were coded independently. To measure inter-rater reliability, coder agreement was calculated by the percentage of agreement between the two coders on subjects’ characteristics, study methodology, and treatment parameters. The coder agreement was 91% and discrepancies between coders were resolved by discussion.

Statistical Analyses

An effect size and 95% confidence interval (CI) for each outcome variable in each study was calculated using Hedges’s g defined as the difference between the group means divided by the pooled standard deviation (Rosenthal, 1984).

In this meta-analysis, overall MWES for incontinent episodes, urine leakage amount, and
perceived severity were calculated weighing for study variance. To test the assumption that the studies share a common population effect size, Q test for homogeneity was used.

Sensitivity analysis was performed to identify whether the results of a research synthesis were sensitive to the inclusion or exclusion of low-quality studies because quality can influence the outcome in a clinical trial (Halvorsen, 1984). Subgroup analyses also were performed to examine whether the results varied by parameters of PFMT and subject characteristics.

To investigate the possibility of publication bias affecting the results of this meta-analysis, funnel graph and rank correlation test such as Kendall’s tau were performed based on standardized effect sizes and sampling variances of all studies reviewed (Begg, 1994). The fail-safe N, another method related to publication bias based on $p$ value, was computed. Statistical analyses were performed using StatsDirect Statistical Software.

### Results

#### Characteristics of Studies and Overall Mean Weighted Effect Sizes

The studies were conducted in different countries: five in the United States (41.7%), two in two European countries (16.7%), two in Korea (16.7%), two elsewhere (16.7%), and one not reported (8.3%), see Table 1. Although many parameters of PFMT may have an impact on UI outcomes, only the length of training period and the number of contractions were clearly described in most of the studies. The length of training period ranged from 6 to 24 weeks, and the number of pelvic floor muscle contractions per day ranged from 24 to 200 and varied in each study. Seven of the 12 studies (58.3%) were scored as 3 or higher for study quality (Table 1).

In this meta-analysis, a negative effect size meant that the subjects in the intervention groups improved more on each outcome measure than the subjects of the control groups. That is, subjects in the intervention groups showed fewer incontinent episodes, less urine leakage, and
Incontinent Episodes

Ten studies (Aksac et al., 2003; Bo, Talseth, & Holme, 1999; Burgio et al., 1998; Burns et al., 1993; Dougherty et al., 2002; Ghoneim et al., 2005; Lagro-Janssen, Debruyne, Smits, & van Weel, 1991; McDowell et al., 1999; Subak, Quesenberry, Posner, Cattolica, & Soghikian, 2002; Sung, Hong, Choi, Baik, & Yoon, 2000) were analyzed for incontinent episodes. One study (Glazener et al., 2001) was excluded from incontinent episodes analysis because the outcome variables reported was the percentage of women who had at least one weekly incontinent episode. The means and standard deviations (SD) of experimental and control groups of Lagro-Janssen et al. (1991) and Bo et al. (1999) were taken from the meta-analysis performed by Hay-Smith and Dumoulin (2006) as the original articles reported only confidence interval (CI), rather than SD. With Ghoneim et al. (2005) study, we estimated the SD from the published CI. All of the other studies with the exception of Sung and colleagues’ (2000) measured incontinent episodes using bladder records, and time intervals varied from 24 hours to two weeks.

The sample sizes of the included studies ranged from 30 to 147, with a mean of 85 (SD = 35.20), and a total of 849 subjects. The subjects’ age ranged from 18 to over 90 years. Effect sizes ranged from -1.71 to -0.42. The overall MWES was -0.68 with a 95% confidence interval of -0.91 to -0.46 (Z = 5.89, p = .000) (Figure 1). The homogeneity test showed Q = 22.59 (df = 9, p = .006), indicating that the studies varied in effect size.

Urine leakage amount

Five studies (Aksac et al., 2003; Bo et al., 1999; Dougherty et al., 2002; Sung et al., 2000; Yoon, Song, & Ro, 2003) were reviewed for this outcome. Six leakage outcome measures
were included in calculating overall MWES as Bo and her colleagues (1999) measured urine leakage amount with two different methods: stress pad test and 24-hour pad test. Most of the measurements used for urine leakage amount involved various types of pad tests, such as 24 hours, one-hour, 30 minutes, and stress-pad test. The sample sizes of these studies ranged from 25 to 147, with a mean sample size of 62 ($SD = 44.09$) and a total of 372 subjects. The subjects’ age ranged from 18 to over 80 years.

The effect sizes on urine leakage amount ranged from -12.29 to 0.09. The MWES was -1.48 with a 95% CI of -2.58 to -0.38 ($Z = 2.64$, $p = .008$) (Figure 2). Subjects in the intervention groups, therefore, showed significantly smaller amounts of urine leakage than the subjects in the control groups after finishing PMFT. The result of $Q$ statistics was 95.63 ($df = 5$, $p = .000$) and was significantly heterogeneous among studies.

**Perceived severity**

Perceived severity related to urine loss was examined in four studies (Dougherty et al., 2002; Glazener et al., 2001; Sung et al., 2000; Yoon et al., 2003). Measurement of this outcome varied greatly in the reviewed studies. For example, Yoon and colleagues (2003) measured perceived severity with eighteen 5-point scales while the authors of the other 3 studies used different single-items. The sample sizes ranged from 25 to 284, with a mean of 129 ($SD = 115.36$) and total of 516 subjects. The participants’ ages for calculating overall MWES of PFMT on perceived severity also ranged from 18 to over 80 years.

The four included studies had effect sizes of perceived severity ranging from -4.43 to -0.33. The MWES was -1.66 and a 95% CI was from -3.59 to 0.27 with $Z = 1.68$ and $p = .092$ (Figure 3). The heterogeneity test had a result of $Q = 191.79$ ($df = 3$, $p = .000$).
Result of Sensitivity Analysis

Sensitivity analysis based on the study quality rating was limited only to incontinent episodes due to the low number of studies available to calculate MWES for urine leakage amount and for perceived severity.

Among 10 studies measured incontinent episodes, six studies scoring 3 or higher were assigned a high quality rating whereas the remaining four studies were assigned a low quality rating. The MWES on incontinent episodes of high quality studies were -0.68, with a 95% CI of -0.85 to -0.50 (Z = 7.54, $p = .000$), and MWES of the studies rated less than 3 was -0.55 with a 95% CI of -0.77 to -0.32 (Z = 4.71, $p = .000$). The MWES on incontinent episodes for both groups were statistically significant, however, there were no significant differences between the MWESs of two study groups ($Q_{BET} = 0.81$, $df = 1$, $p = .368$).

Results of Subgroup Analysis

This meta-analysis also examined the effectiveness of PFMT on incontinent episodes by the number or magnitude of two specific PFMT parameters and two subject characteristics: (a) the length of training period; (b) the number of PFM contractions; (c) subjects’ type of UI, and (d) subjects’ age (Figure 4). Sub-analyses were not possible for urine leakage amount and perceived severity because so few studies were available.

Length of training period

The same 10 studies used in analyzing incontinent episodes reported training-period durations with a range of 6 to 24 weeks. The studies were divided based on eight weeks of training period because muscle hypertrophy begins only after regular and intense strength training for more than eight weeks (Dinubile, 1991). The overall MWES for the four studies that used PFMT for more than eight weeks was -0.66 with a 95% CI of -0.88 to -0.44 (Z = 5.91, $p$
The overall MWES for the six studies that used PFMT for eight weeks or less was -0.61 with a 95% CI of -0.79 to -0.43 (Z = 6.60, p = .000). However, there was no heterogeneity in these two study groups ($Q_{BET} = 0.13, df = 1, p = .718$).

**Number of PFM contractions**

Eight studies reported the number of daily contractions from a minimum of 24 to a maximum of 200. These studies were then divided into two subgroups on the basis of Wilson and colleagues’ (2002) suggestion that the PFMT should include three sets of 8 to 12 slow-velocity and maximal-strength pelvic floor muscle contractions daily. The overall MWES of three studies that used a minimum of 24-36 daily contractions was -0.68 with a 95% CI of -0.98 to -0.37 (Z = 4.32, p = .000). The overall MWES of five studies that used at least 45 daily PFM contractions was -0.68 with a 95% CI of -0.87 to -0.50 (Z = 7.26, p = .000). The MWESs of two study groups were not heterogeneous ($Q_{BET} = 0.00, df = 1, p = .995$).

**Subjects’ type of UI**

The ten studies involved identified UI type are as follows: six had women with stress UI and mixed UI with dominant stress symptoms; three had women with any kind of UI; and one had women with urge UI and mixed UI with dominant with urge symptoms. The studies with women having stress UI and mixed UI with dominant stress symptoms showed an overall MWES of -0.77 with a 95% CI of -0.98 to -0.55 (Z = 7.03, p = .000). Studies of subjects with any type of UI showed an overall MWES of -0.47 with a 95% CI of -0.68 to -0.26 (Z = 4.40, p = .000). As a result of the heterogeneity test, $Q_{BET}$ was 3.99 ($df = 1, p = .046$).

**Subject’s age**

The participants of five studies used for these analyses (Burgio et al., 1998; Burns et al., 1993; Dougherty et al., 2002; McDowell et al., 1999; Subak et al., 2002) were women aged 55
years and older. The mean age was 63.37 years. The overall MWES of studies with an average age over 60 years was -0.54 with a 95% CI of -0.71 to -0.37 ($Z = 6.21, p = .000$). The overall MWES of studies in which the average age was younger than 60 years was -0.94 with a 95% CI of -1.21 to -0.66 ($Z = 6.58, p = .000$). The heterogeneity test showed that $Q_{BET}$ was 5.98 ($df = 1, p = .015$).

Publication Bias

The results of Funnel graph on incontinent episodes, urine leakage amount and perceived severity were showed at Figure 5, 6, 7 respectively. The results of Kendall’s tau on each outcome were -0.42 ($p = .073$), -0.73 ($p = .017$) and 0 ($p = .75$) respectively. The fail-safe N ($N_{FS}$) for incontinent episodes was 210 and much greater than the reasonable guideline 60 computed from the formula with 10 studies. Inadequate number of studies prohibited testing fail-safe N for urine leakage amount, and perceived severity.

Discussion

Calculating the MWES of PFMT compared to no treatment was the one of the goals of this meta-analysis. Ten RCT for incontinent episodes, five RCT for urine leakage amount, and four RCT for perceived severity were reviewed.

The results of Q test for homogeneity on each outcome measure showed that the variances in MWES were not homogeneous. Hence, inclusion of other variables may be necessary to more fully explain the variance in these effect sizes. Therefore, the MWESs of the random effects model on each outcome were reported because the model is more conservative in the presence of unexplained heterogeneity than the fixed model (Shadish & Haddock, 1984).

The overall MWES of PFMT on the incontinent episodes was -0.68 indicating that PFMT was effective in reducing incontinent episodes compared to no treatment. This effect size
is considered a medium effect size (Cohen, 1988). Both the MWESs on urine leakage amount and perceived severity, -1.44 and -1.66 respectively, are considered very large (Cohen, 1988). However, the MWESs on urine leakage amount and perceived severity must be interpreted with caution. In the case of the MWES on urine leakage amount, funnel graph and the results of Kendall’s tau showed that the publication bias is likely to be present. The MWES on perceived severity that was calculated from only four studies had the statistic Z which did not exceed the 1.96 critical value at $a = .05$, therefore, it is possible to conclude PFMT has no significant effect on subjects’ perceived severity.

Had most studies precisely reported on all PFMT parameters and sample characteristics, more subgroup analyses could have been performed. In this case, the only subgroup analyses we were able to perform were length of training period, number of daily contraction, type of UI, and age. Because the number of studies and the variation in each parameter were very limited, an analysis of variance for effect sizes with one factor model was used instead of multiple regression analysis for effect sizes (Hedges, 1984).

The result of subgroup analysis by length of training period showed that the MWESs of both study groups, one with more than eight weeks training period and the other with eight weeks or less, were statistically significant in reducing incontinent episodes and that the difference between MWESs of the two groups was not statistically significant.

Three possible reasons for the lack of difference between these groups might be considered. First, the small variation among training periods could be a reason. In this meta-analysis, two subgroups were used. One subgroup consisted of two studies with 6-week training periods and four studies with 8-week training periods. Another subgroup consisted of three studies that used 12-week training periods and one study that used a 6-month training period.
A second possible reason might be that the outcome variable for this analysis was incontinent episodes rather than muscle strength. Empirical studies (Burgio et al., 1998; Burns et al., 1993; Subak et al., 2002; Wyman et al., 1998), which analyzed the timing of response to training, showed that reduction in incontinent episodes begins immediately after starting PFMT and the degree of this reduction decreased with time. It is necessary, however, for women to attain PFM hypertrophy to maintain reduced incontinent episodes or to reach a continent state, and they attain PFM hypertrophy by repeated contractions over time. Thus maintaining a regimen of performing PFM contractions regularly and repeatedly until at least achieving PFM hypertrophy is important. Future studies that measure both visible PFM hypertrophy and incontinent episodes over time are needed to determine the ideal length of training period.

Patients’ adherence to PFM contractions may also confound the effect of PFMT because PFMT requires at least several weeks of maintaining regular contractions. Most of the included studies in this meta-analysis did not measure or report the level of adherence, even though some studies applied specific methods for increasing adherence to PFMT programs. As Wilson et al. (2002) noted, the studies investigating the effects of PFMT should precisely describe specific parameters of PFMT such as the number, the velocity, and the strength of the contractions, and length of training period. Most studies, however, have poorly reported those parameters of the PFMT. Only two studies (Bo, et al., 1999; Lagro-Janssen, et al., 1991) in this meta-analysis reported using close to the maximum strength contractions recommended to increase muscle volume (Miller, Kasper, & Sampselle, 1994). Two other studies (Burns et al., 1993; Ghoniem et al., 2005) reported on the velocity of contractions.

The number of contractions was used in subgroup analyses as eight studies reported the number of daily contractions. The studies were divided into two groups, one including studies
using daily minimum contractions within the range of 24 to 36 as suggested by Wilson and colleagues (2002) and the other subgroup included studies with at least 45 contractions daily. The result indicated that the PFMT has significant effect on incontinent episodes in both groups and no variation in effect size across studies with different number of daily PFM contractions. Thus no statistically significant relationship between the number of contractions and effect size on incontinent episodes was found.

Other contraction characteristics may have an impact on incontinent episodes; therefore, more research is needed to test combinations of number, strength, and velocity of contractions to determine the ideal configuration of a PFMT protocol for female UI.

In recent years, PFMT has been applied for the behavioral management of urge UI with the understanding that PFMT can inhibit unstable bladder contractions (Burgio et al., 1998; Wyman et al, 1998). Generally PFMT has been used to manage stress UI based on the rationale that PFMT enhances urethral resistance and PFM strength, thus to prevent leaking urine during times of sudden, episodic increased intravesical pressure (DeLancey, 1988).

In this meta-analysis, studies were divided two UI type groups. One group included studies with only women with stress UI alone and the other group included studies with incontinent women regardless of the UI type. Bladder training (BT) was used in the latter group as an additional intervention. Significant treatment effects on reducing incontinent episodes were reported although no analyses were conducted on the effectiveness of PFMT and BT for each type of UI. The result of this subgroup analysis, however, showed that the PFMT is more effective on women with stress UI alone, although the PFMT was effective to both groups on reducing incontinent episodes. Research is still needed to explore the specific mechanism by which PFMT effects stress and urge UI.
Age is often viewed as a risk factor for incontinence due to increased prevalence of UI with aging. No causal relationship has been found. For subgroup analysis, the studies were divided into two study groups including studies with mean ages younger than and older than 60 years. The result of this subgroup analysis implies that the effect of PFMT in the younger group was significantly greater than that for the older group, although PFMT was significantly effective in reducing incontinent episodes in both groups. This result confirms earlier findings that although the effect is greater in younger women, older women can benefit from PFMT. Future studies should explore different methods of training such as individualized and graded PFMT regimen and determine the effect of age may play.

One limitation of this meta-analysis is the limited number of eligible studies especially, those reporting measured urine leakage amount and perceived severity as outcomes. This small number of studies also prevented us from using multifactor models in subgroup analysis. We did not contact authors of unpublished studies to include in this meta-analysis. In addition, the authors of included studies were not contacted to provide supplemental information to conduct analyses thus it is possible more information may have been available.

In conclusion, the treatment effect of PFMT on the incontinent episodes may be greater in younger women with stress UI alone, though PFMT is also effective for all women with UI. It appears that the number of daily PFM contractions and the length of training period as parameters of PFMT treatment are not related to effect sizes on incontinent episodes on the condition that PFMT includes at least daily 24 PFM contractions and the regime is maintained for at least six weeks. To advance evidence based interventions for incontinent women, researchers must include detailed reports of parameters used in PFMT regimes and to test configuration of these parameters to devise the ideal protocol to effectively treat incontinence.
References


Meta-Analysis of Pelvic Floor Muscle Training


Wyman, J., Fantl, J., McClish, D., Bump, R., & the Continence Program for Women Research

Figure Caption

*Figure 1.* Effect Size of Each Study (study number 1, 2, 3, 4, 5, 6, 8, 9, 10, 11) and Overall Mean Weighted Effect Size on Incontinent Episodes
Figure Caption

*Figure 2.* Effect Size of Each Study (study number: 1, 2, 5, 11, 12) and Overall Mean Weighted Effect Size on Amount of Urine Leakage.
Figure Caption

*Figure 3.* Effect Size of Each Study (study number 5, 7, 11, 12) and Overall Mean Weighted Effect Size on Perceived Severity
Figure Caption

*Figure 4.* Subgroup Analysis on Incontinent Episodes by Study Quality, Length of PFMT, Number of Daily Contraction, Type of UI, and Mean Age of Subjects
Figure Caption

Figure 5. Funnel Graph of Ten Studies for Incontinent Episodes
Figure Caption

*Figure 6.* Funnel Graph of Six Studies for Urine Leakage Amount
Figure Caption

*Figure 7. Funnel Graph of Four Studies for Perceived Severity*
### Table 1

*Characteristics of the Randomized Clinical Trials Included in the Meta-analysis*

<table>
<thead>
<tr>
<th>Authors (Year)</th>
<th>Study setting</th>
<th>Sample size</th>
<th>Age Range</th>
<th>Age Mean</th>
<th>Type of UI</th>
<th>Number of PFM contractions (daily)</th>
<th>Length of training (weeks)</th>
<th>Study quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aksac et al. (2003)</td>
<td>N/R</td>
<td>30</td>
<td>N/R</td>
<td>53</td>
<td>Stress</td>
<td>30</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Bo et al. (1999)</td>
<td>Norway</td>
<td>55</td>
<td>24 ~ 70</td>
<td>51</td>
<td>Stress</td>
<td>24 ~ 36</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>Burgio et al. (1998)</td>
<td>US</td>
<td>114</td>
<td>55 ~ 92</td>
<td>67</td>
<td>Urge</td>
<td>45</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Burns et al. (1993)</td>
<td>US</td>
<td>82</td>
<td>55 and older</td>
<td>63</td>
<td>Stress</td>
<td>200$^c$</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Dougherty et al. (2002)</td>
<td>US</td>
<td>147</td>
<td>55 ~ 95</td>
<td>68</td>
<td>All</td>
<td>45$^d$</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Ghoneim et al. (2005)</td>
<td>Netherlands, UK, US</td>
<td>79</td>
<td>18 ~ 75</td>
<td>53</td>
<td>Stress</td>
<td>50</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Glazener et al. (2001)</td>
<td>UK, NZ</td>
<td>284</td>
<td>N/R</td>
<td>30</td>
<td>All</td>
<td>80 ~ 100</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>Lagro-Janssen et al. (1991)</td>
<td>Netherlands</td>
<td>66</td>
<td>20 ~ 65</td>
<td>45</td>
<td>Stress</td>
<td>50 ~ 100</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>McDowell et al. (1999)</td>
<td>US</td>
<td>93$^b$</td>
<td>60 ~ 97</td>
<td>77</td>
<td>All</td>
<td>30 ~ 45</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Subak et al. (2002)</td>
<td>US</td>
<td>123</td>
<td>55 and older</td>
<td>69</td>
<td>All</td>
<td>N/R</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Age</td>
<td>Treatment</td>
<td>Stress</td>
<td>N/R</td>
<td>Sample Size</td>
<td>Data Included</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
<td>-----</td>
<td>-------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Sung et al. (2000)</td>
<td>KOREA</td>
<td>60</td>
<td>18 and older</td>
<td>N/R</td>
<td>Stress</td>
<td>N/R</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>Yoon et al. (2003)</td>
<td>KOREA</td>
<td>25</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>60</td>
<td>8</td>
</tr>
</tbody>
</table>

*Note: N/R = not reported.*

- a Sample size was based on data extracted for meta-analysis; therefore, it may be different from sample size of original study.
- b Less than 10 male participants were included in this sample.
- c Beginning with 4 set of 20 contractions and increased by 10 per set over 4 weeks until daily maximum of 200 contractions
- d Beginning with 15 contractions per day and increased by 15 every 3 weeks to 45 contractions per day

Incontinent episodes analysis: studies 1, 2, 3, 4, 5, 6, 8, 9, 10, 11 were included

Urine leakage amount analysis: studies 1, 2, 5, 11, 12 were included

Perceived severity analysis: studies 5, 7, 11, 12 were included
Figure 6

Bias assessment plot

Standard error

Effect size