Electromyographic Biofeedback to Improve Lower Extremity Function After Stroke: A Meta-Analysis

Julie D. Moreland, MSc, Mary Ann Thomson, MSc, Angela R. Fuoco, BHSc(PT)

ABSTRACT. Moreland JD, Thomson MA, Fuoco AR. Electromyographic biofeedback to improve lower extremity function after stroke: a meta-analysis. Arch Phys Med Rehabil 1998;79:134-40.

Objective: To examine the efficacy of electromyographic (EMG) biofeedback compared with conventional physiotherapy for improving lower extremity function in stroke patients.

Data Sources: A literature search covering the years 1976 to 1995 in MEDLINE, CINAHL, and EXCERPTA MEDICA.

Study Selection: Studies of adults after stroke, in which the treatment group received biofeedback alone or with conventional physical therapy and the control group received conventional physical therapy. Outcomes included functional measures related to the lower extremity. The study design criterion was that all must be randomized controlled trials.

Data Extraction: Study quality was assessed independently by two observers using eight criteria. Data for analysis were extracted by two observers to ensure accuracy.

Data Synthesis: For outcomes that were analyzed in more than one study, meta-analyses were done. Seventy-nine studies were identified as potentially relevant and eight studies met the selection criteria. The mean effect sizes were: for ankle dorsiflexion muscle strength, 1.17 (95% CI, .50-1.85; p = .0006); for gait quality, .48 (95% CI, −.06-1.01; p = .08); for ankle range of motion, .07 (95% CI, −.42-0.57; p = .78); for ankle angle during gait, .52 (95% CI, −.18-1.21; p = .14); for stride length, .09 (95% CI, −.56-.73; p = .80); and for gait speed, .31 (95% CI, −.16-.78; p = .20).

Conclusions: The results indicate that EMG biofeedback is superior to conventional therapy alone for improving ankle dorsiflexion muscle strength.

IMPORTANT GOALS in the rehabilitation of stroke patients include improvements in walking, in gross motor function, and in lower extremity motor control. Electromyographic (EMG) biofeedback is one method directed at improving lower extremity impairments and the associated disabilities, and several studies and reviews have examined the effectiveness of this technique. Periodic synthesis of this literature is important to update clinical decision-making about the selection of techniques for lower extremity training and gait rehabilitation.

This review addresses the following question: after stroke, is EMG biofeedback more efficacious for improving lower extremity motor control and walking than conventional therapy? Since this question is clinically driven, overt motor control and walking changes were selected as outcomes rather than changes at the physiologic level. As in a previous review of the EMG biofeedback literature for upper extremity rehabilitation following stroke,[1] EMG biofeedback was defined as the use of instrumentation applied to the patient's muscle(s) with external electrodes to capture motor unit electrical potentials. The biofeedback unit converts the potentials into visual or audio information for the patient and the therapist. The patient is instructed to activate or decrease the activity of the muscle(s). Usually the method is used to augment desired muscle action or to decrease unwanted muscle activity. A standardized approach to biofeedback training for rehabilitation after stroke has not been put forth.

RATIONALE FOR THIS REVIEW

A search of the literature for reviews of EMG biofeedback for stroke rehabilitation, initially completed in 1992,[1] was updated for this review. We published a summary of these reviews in 1994.[1] We included reviews that contained both upper extremity and lower extremity studies. The criteria developed by Oxman and Guyatt[2] were used to critically appraise these review articles. Nonquantitative reviews suggested that the early uncontrolled studies demonstrated that EMG feedback was efficacious. Later studies found no difference between EMG and control groups; however, type II error was not addressed by these literature reviews. Subsequently, three meta-analyses were identified: Schleenbaker and Mainous,[3] Moreland and Thomson,[1] and Glanz and coworkers.[4]

Schleenbaker and Mainous[3] performed a meta-analysis and reported statistically significant results in favor of biofeedback. The outcomes that were analyzed were a broad range of functional outcomes. Although the studies they included were randomized or matched control designs, their internal validity was not evaluated. For example, studies with inadequate follow-up could bias the results. Blinding of the outcome observer may also be an important methodologic feature. Two of the studies had no treatment in the control groups, and consequently, conclusions can be drawn only about whether EMG biofeedback is better than no therapy at all. It may have been clinically important to do a sensitivity analysis on the basis of whether the control group received treatment or did not. Combining all studies yielded a positive result; however, the combined effect size was not converted to a clinical measure to interpret the relevance of the effect size.

We[1] performed an overview of EMG biofeedback for the upper extremity. One of the inclusion criteria was that the control group received conventional therapy. The combined odds ratios for improved versus not improved were not statistically significant for the function outcome and not statistically significant for the impairment outcome. The size of the effect, based on number needed to treat, was small. In addition, a post hoc analysis of those who did improve showed an effect size of −.02 corresponding to one point on the Upper Extremity Function Test in favor of conventional therapy. We concluded that

From St. Joseph's Hospital and St. Peter's Hospital (Ms. Moreland); Chedoke-McMaster Hospitals (Ms. Thomson, Ms. Fuoco); and the School of Rehabilitation Science, McMaster University (Ms. Moreland, Ms. Thomson, Ms. Fuoco), Hamilton, Ontario, Canada.

Submitted for publication August 9, 1996. Accepted in revised form June 9, 1997.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Reprints are not available from the authors.

0003-9993/98/7902-4143$3.00/0

therapists should consider factors such as cost, ease of application, and patient preference when selecting between these therapies.[1] Since the inclusion criteria and, hence, the inference space differed from the meta-analysis of Schleenbaker and Mainous,[3] a comparison of the two meta-analyses is not indicated.

Glanz and colleagues[4] performed a meta-analysis to determine if biofeedback increased range of motion on the affected side after stroke. Inclusion criteria included randomized controlled trials and the availability of extractable data such that an effect size could be calculated. Meta-analyses were performed separately for the lower extremity and upper extremity studies. Of six lower extremity studies, one (Mandel)[5] did not provide any treatment to the control group. This may have influenced the findings in a positive direction. The pooled effect size for the lower extremity was 1.58 with a 95% confidence interval of −.58 to 3.59 (not statistically significant). Sensitivity analyses also yielded results that were not statistically significant. Since the effect size was not converted into the units of range of motion, it is difficult to determine if the estimate would represent a clinically important effect if it had been statistically significant. These methodologic problems limit the interpretation of this review.

The purpose of this overview was to examine the effect of EMG biofeedback in comparison with conventional therapy for lower extremity function and gait rehabilitation after stroke.

METHODS

Study Identification and Selection

The search strategies and selection criteria were described in detail elsewhere.[1] MEDLINE was searched using the key words "electromyography," "biofeedback," and "cerebrovascular disorders." The keywords "biofeedback" and "cerebral vascular accident" were used to search the CINAHL database. EXCERPTA MEDICA was reviewed manually. We updated the search to include December 1995, and we excluded the Dissertation Abstracts International database, which was not useful in the previous review. First authors of studies meeting the selection criteria were contacted by mail to obtain any further published or nonpublished studies. The search was limited to English-language publications. The following selection criteria were applied independently by two authors (MAT, ARF) to assess all relevant studies.

1. Population: adults poststroke.
2. Intervention: in the treatment group, EMG biofeedback alone or with conventional physical therapy; in the control group, conventional physical therapy excluding alternate feedback devices and functional electrical stimulation.
3. Outcomes: functional measures of the lower extremity, including lower extremity function tests and walking, stage of motor recovery, range of motion, and muscle strength.
4. Study design: randomized controlled trials.

Agreement on whether each study met the selection criteria was calculated using the weighted Kappa statistic. Data were extracted from the studies by two investigators (MAT, ARF) to ensure accuracy. For cross-over trials, data were taken only from the first period.

Quality Assessment of Included Studies

The quality of a trial may be defined as "the extent to which its design and conduct are likely to have prevented systematic errors." [2] Variation in the quality of included trials can lead to variation in the results. We used eight indicators of study quality: (1) at least 95% follow-up of subjects (excluding deaths); (2) comparability of treatment and control groups for age, time poststroke, receptive communication, sensation, and baseline measures of outcome variables; (3) provision of equal time and attention to both groups; (4) random allocation of therapists to subjects; (5) monitoring of treatment protocols for accuracy and consistency; (6) provision of placebo biofeedback to the control group; (7) avoidance of contamination and cointervention; (8) analysis of withdrawals in the group to which they were randomized. Two of the authors applied these criteria independently (MAT, ARF). Interobserver agreement was calculated using the Kappa statistic and any disagreements were resolved by discussion.

Data Analysis

A meta-analysis is a systematic overview that incorporates specific statistical strategies to assemble and summarize the results of a number of studies into a single numerical estimate. It provides an estimate of the magnitude of the relation between independent and dependent variables and tests whether the relation or effect is statistically significant. In our study, these variables were the effects of EMG biofeedback and conventional physiotherapy on specific lower extremity functional outcomes. A common estimator of effect magnitude is the "effect size" or standardized mean difference. It is most commonly calculated by subtracting the posttreatment mean of the control group from that of the experimental group and dividing by either the posttreatment control or pooled group standard deviation.[6] This method assumes that the distributions are such that the means and standard deviations are meaningful. Where raw data were available, we examined the distributions to ensure that they had an approximate normal distribution. In some of the selected studies, the posttreatment control or experimental means and standard deviations were not published, nor could they be obtained from the authors. Change scores (posttreatment minus pretreatment scores) could be calculated in these studies. We used the change score in the numerator and the postintervention standard deviation in the denominator (where available) as described by Lund.[6] When the postintervention standard deviation was not available, we used the standard deviation of the change scores in the denominator (using the same method for each outcome). When standard deviations were unobtainable, they were imputed from another study that used the same unit of measurement and had the largest standard deviation. Study results were combined if more than one study examined the same outcome. The specific methods for calculating effect sizes from individual studies, determined by the available data, were as follows: to calculate *ankle muscle strength and ankle range of motion*, the difference in mean change score of treatment and control groups was divided by the pooled standard deviation of change scores; to calculate *gait quality, ankle angle during gait, stride length, and gait speed*, the difference in mean change score of treatment and control groups was divided by the pooled standard deviation of the posttest scores.

We calculated combined effect sizes using a random effects model for the outcome variables. A random effects model was chosen since it is a method that includes between-study variances as well as within-study variances.[7] It is an appropriate method when heterogeneity is expected. Some heterogeneity was expected because it was anticipated that studies would have some differences in their samples and applications of treatment. Meta-analysis Programs (5.1)[8] was the software used to determine mean (pooled) effect size, the 95% confidence interval for the effect size, the significance level (Z test), the amount of variance explained by sampling error, and the test for homogeneity (Q) of effect sizes. Since the observed variability in sample estimates of effect size is partly due to the variability in the

Table 1: Characteristics of Selected Studies

Characteristic	Basmajian et al,[9] 1975	Binder et al,[10] 1981	Burnside et al,[11] 1982	John,[12] 1986	Mulder et al,[17] 1986	Cozean et al,[14] 1988	Colborne et al,[13] 1993	Intiso et al,[16] 1994
Country	US	US	UK	UK	US	US	Canada	Italy
Facility	Rehab center (inpatient and outpatient)	Outpatient	Hospital outpatient	Hospital (inpatient and outpatient)	Rehab center and research lab	Rehab center	Stroke rehab unit	Rehab center
Mean months poststroke*	33.6 (4-120)	>15	57.84 (6-144)	2.5 (.4-9.3)		?	17	9.8
Mean age (yrs)*	51 (30-63)	? (?)	? (53-84)	49 (35-63)	? (34-68)	56 (?)	? (?)	? (40-85)
Receptive aphasia	Excluded	Excluded	Included	Excluded	Excluded	?	Excluded	Excluded
Sensory loss	?	NA	?	?	?	?	?	?
Conventional physical therapy with biofeedback	Yes	Yes	No	No	No	Yes	No	Yes
Placebo in control group	No	No	Yes	No	No	No	No	No
Frequency and duration of treatment	15 sessions, 5 weeks	12 sessions, 4 weeks	12 sessions, 6 weeks	12-15 sessions, 3 weeks	15 sessions, 5 weeks	18 sessions, 6 weeks	8 sessions, 4 weeks	40 sessions, 8 weeks
Outcome measures	Ankle AROM Ankle strength Gait quality	Ankle AROM Ankle angle Knee angle Unilateral weight-bearing Timed ambulation	Ankle AROM (dorsiflex) Dorsiflex strength (0-5) Gait quality	Ashburn Scale Time to walk 15m and climb 7 stairs Active knee extension Knee angle of 5 second static knee contraction	Ankle AROM Gait velocity	Ankle angle Knee angle Stride length Gait cycle time	Velocity Stride length Stride time Ankle and knee angles Push-off impulse Swing ratio Weight ratio	Canadian Neurological Scale Adams Scale Ashworth Scale Barthel Index Gait quality Kinematic analysis: step length; velocity; ankle angle at swing; heelstrike

Abbreviations: ?, not clear; NA, not applicable; AROM, active range of motion.
* Range given in parentheses.

underlying population parameters and partly due to the sampling error of the estimator about the parameter value, the observed effect size variance is decomposed into both parts. If 100% of the observed variance is explained by sampling error, this indicates that the data are homogeneous. This program uses the methods described by Hedges and Olkin.[7]

For meta-analyses that were not statistically significant, power analyses were calculated for three estimates of clinically important differences (selected by JDM) between the treatment and control groups. Using the standard deviations available for each study, the alternative hypotheses as specified by the clinically important differences were used to estimate the effect size from each study. The alternative hypothesis values were summarized with the arithmetic mean and this was divided by the same standard error that was computed from the software to estimate a Z coefficient. Probabilities for these were computed to be a power estimate. Detailed probabilities were estimated from each Z value with an HP48GX that has the normal probability as one of its software features.

We hypothesized that differences in effect sizes across studies may be caused by differences in the samples, treatments, or experimental designs. These included blinded versus nonblinded assessment of outcomes, whether there was inclusion of conventional therapy in the treatment group, whether a placebo was given to the control group, and whether treatment monitoring was used. If heterogeneity was identified in the analyses, sensitivity analyses based on these factors were planned.

RESULTS

Seventy-nine studies were identified as relevant and 12[9-20] met the selection criteria. The interobserver agreement was .67. The selected studies were all identified by the database searches. No additional (published or unpublished) studies were located by writing to first authors. Four trials[15,18-20] were subsequently excluded because outcome data were not available in the published report and could not be obtained from the authors. The characteristics of the remaining studies are summarized in table 1. Gait speed was reported in 6 studies,[12-14,16,17,20] ankle range of motion was reported in 4 studies,[9-11,17] gait quality was reported in 3 studies,[9,11,16] ankle angle and stride length during gait were assessed in 3 studies,[13,14,16] and ankle strength was measured in two studies.[9,11]

Study Quality

The results of the methodologic assessment are summarized in table 2. Interobserver agreement on each criterion varied from 0 to 1. In many trials, insufficient data were reported to determine group comparability. The outcome measures were assessed blindly in 7 of 8 studies. Six of 8 studies had at least 95% follow-up of subjects and analyzed the withdrawals in their original group. None of the studies randomized therapists to groups. Treatment was monitored in fewer than half of the studies.

Meta-Analyses

Ankle muscle strength. Two studies[9,11] assessed ankle muscle strength. Basmajian et al[9] measured ankle dorsiflexion muscle strength using a spring dynamometer in a semi-reclining position, while Burnside and coworkers[11] used a scale of 0 to 5 with the patient sitting. Both effect sizes were positive, and Burnside[11] reported a statistically significant result. Basmajian[9] did not report statistical significance; however, we calculated a t test from Basmajian's data, and it was statistically significant ($p < .05$).

The mean (combined) effect size was statistically significant $p = .0006$ with a mean of 1.17 (95% CI, .50-1.85), Z test. In natural units this mean effect size represents a difference of 2.5kg muscle force in favor of biofeedback. The individual effect sizes and combined effect sizes are illustrated in figure 1. Homogeneity of the effect sizes is evident. The test for homogeneity was not significant ($p = .43$, $Q = .62$) and the amount of variance explained by sampling error was 100%. It was noted that the results of the fixed effects weighted integration method gave the same mean effect size and was statistically significant ($p < .05$).

Gait quality. Three studies[9,11,16] assessed gait quality. A

Table 2: Validity Criteria

Criteria	Basmajian,[9] 1975	Binder,[10] 1981	Burnside,[11] 1982	John,[12] 1986	Mulder,[17] 1986	Cozean,[14] 1987	Colborne,[13] 1993	Intiso,[16] 1994
Sample*	Y	Y	Y	Y	Y	N	Y	N
Group comparability								
Duration of stroke	N	N	N	N	?	?	N	N
Age	Y	N	Y	N	?	N	Y	N
Receptive aphasia	N/A	N/A	N	N/A	N/A	N/A	N/A	N/A
Sensory loss	N	N/A	N	?	?	?	Y	?
Outcomes at baseline	?	N	N	?	?	?	?	N/A
Intervention								
Equal time	Y	Y	N	Y	Y	Y	Y	?
Therapists randomized	N	N	N	N	N	N	N	?
Treatment monitored	Y	Y	N	N	Y	N	N	?
Treatment standardized	Y	Y	N	N	Y	Y	Y	?
Contamination/cointervention avoided	Y	Y	N	Y	Y	N	Y	?
Outcomes								
Blinded assessment	Y	Y	Y	Y	N	Y	Y	Y
Withdrawals analyzed in original group	N/A	N/A	N/A	N/A	N/A	N	N/A	N

Abbreviations: Y, criterion met; N, criterion not met; N/A, not applicable; ?, not clear.
* >95% assessed for outcome measures.

scale developed by Basmajian et al[9] was used. This 6-point ordinal scale measures the degree of dorsiflexion control during gait (from complete foot drop to normal heel-toe gait pattern). All three studies demonstrated positive changes, and one[16] of these reported a statistically significant change for the experimental group. Basmajian[9] did not report on statistical significance; however, from the data presented, the *t* test was statistically significant ($p < .05$). For calculation of individual effect sizes, Burnside and associates[11] did not report postscores and an effect size could not be calculated from their statistical testing (Fisher's exact probability test). Therefore the postscore pooled standard deviation was imputed from Intiso et al.[16]

The mean effect size was not statistically significant ($p = .08$) with a mean of .48 (95% CI, −.06-1.01), *Z* test. In natural units this mean effect size represents approximately 0.5 of a point on the Basmajian gait scale in favor of biofeedback. Individual effect sizes and the combined effect size are shown in figure 1. Homogeneity of the effect sizes is evident and the test for homogeneity was not significant ($p = .99$, $Q = .007$). The amount of variance explained by sampling error was 100%, which further substantiates homogeneity. Power analyses are given in table 3.

Ankle range of motion. Ankle range of motion was assessed in four studies.[9-11,17] In two studies,[10,17] the effect sizes were negative and in two studies,[9,11] the effect sizes were positive (fig 1). For calculation of individual effect sizes, Mulder and associates[17] did not report standard deviations of the results. Although they stated they performed an ANOVA, descriptive and inferential statistics were not given. Therefore, the pooled standard deviation from Burnside[11] was imputed. The mean effect size was not statistically significant ($p = .76$) with a mean of .07 (95% CI, −.42-0.57), *Z* test. In natural units, this mean effect size represents a difference of 0.7° between biofeedback treatment and control treatment. The test for homogeneity was not significant ($p = .43$, $Q = 2.7$) and the amount of variance explained by sampling error was 100%. Power analyses are given in table 3.

Ankle angle during gait. This was assessed in three studies.[13,14,16] All had positive effect sizes (fig 1). The mean effect size was not statistically significant ($p = .14$) with a mean of .51 (95% CI, −.17-1.21), *Z* test. In natural units, this mean effect size corresponds to a difference of 5.7° in favor of biofeedback. The test for homogeneity was not significant ($p = .43$, $Q = 1.70$) and the amount of variance explained by sampling error was 100%. Power analyses are given in table 3.

Stride length. Stride length was assessed in three studies,[13,14,16] two of which had findings favoring conventional therapy. The mean effect size was not statistically significant ($p = .80$) with a mean of .09 (95% CI, −.56-.73), *Z* test. In natural units, this mean effect size corresponds to 2.0cm. Individual effect sizes, as well as the combined mean effect size, are shown in figure 1. The test for homogeneity was not significant ($p = .79$, $Q = .47$) and the amount of variance explained by sampling error was 100%. Power analyses are given in table 3.

Gait speed. Gait speed was assessed in 6 studies.[10,12-14,17] In 2 studies,[12,17] the effects sizes were negative and in 4[10,13,14,16] they were positive (fig 1). For the calculation of individual effect sizes, Binder et al[10] did not provide posttest standard deviations. Therefore, the standard deviation from Mulder[17] was imputed. An effect size could not be calculated from the statistical test in Binder[10] because a Mann-Whitney *U* test was used. The mean effect size was not statistically significant ($p = .19$) with a mean of .31 (95% CI, −.16-.78), *Z* test. In natural units, this mean effect size corresponds to a difference of 8.4m over 2 minutes in favor of biofeedback. The test for homogeneity was not significant ($p = .90$, $Q = 1.58$) and the amount of variance explained by sampling error was 100%. It was not possible to calculate power for this meta-analysis because the data from each study could not be converted to the same unit of measurement.

DISCUSSION

Research overviews have been criticized for their potential biases. Publication bias may occur when studies having unfavorable trends or nonsignificant results are not published. Although the potential for this bias always exists and is difficult to measure, we addressed this issue by attempting to search for unpublished studies by contacting authors of papers that met our selection criteria. A weakness is that we only searched for English language studies. The effect of limiting studies to the English language has not been examined.[21] Personal bias may occur during study selection and validity evaluation. The agreement between investigators for the selection criteria was moderate to high but the agreement for the validity criteria was generally poor. This was overcome by reaching a consensus on each of the validity criteria to reduce errors caused by oversight and ensure that each study was carefully reviewed. A potential weakness of this study is that the evaluators were not blinded to authors, institutions, or results when evaluating the validity of the studies.

In calculating effect sizes, for two outcomes, ankle muscle strength and ankle range of motion, it was necessary to use the standard deviation of the change scores in the denominator. This is not ideal because the magnitude of the standard deviation of change scores is dependent on the correlation between the pretest and posttest measures.[6] Since we were combining very similar outcomes, it is expected that the correlation between pretest and posttest measures would be approximately equal for each study within each outcome (hence, we would not be combining apples and oranges). This method has been used in another meta-analysis of poststroke rehabilitation examining the effectiveness of functional electrostimulation.[22]

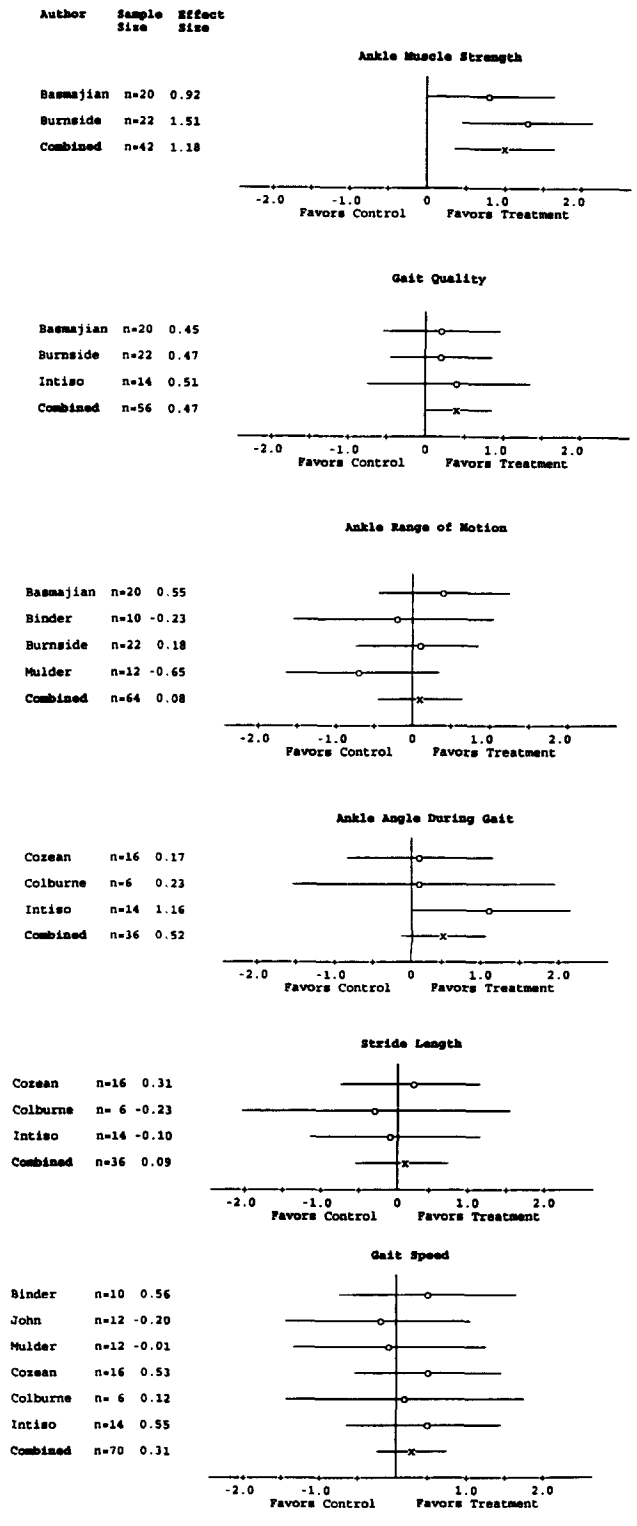

Fig 1. Individual and combined effect sizes for the lower extremity functional measures studied.

Table 3: Power Analyses

Outcome	Difference Between Treatment and Control Groups	Power
Gait quality	.25 point	.16
	.50 point	.41
	1.0 point	.91
Ankle range of motion	5°	.99
	10°	1.00
	15°	1.00
Ankle angle during gait	5°	.54
	10°	.95
	15°	1.00
Stride length	.03m	.69
	.06m	.84
	.09m	.93

For each outcome, three differences were selected based on possible clinical importance.

In this review, a random effects model was chosen because the studies examined somewhat different samples and treatment applications. This model incorporates the concept that each study provides a sample estimate drawn from a population of effects. It is more conservative than the fixed effects method of pooling study results. We found that in all meta-analyses, the pooled mean effect size was identical for the random and fixed effects analyses; however, the confidence intervals were slightly wider for the random effects analyses. There were no substantial differences in significance levels between the fixed effects and random effects methods.

Mohr and Olkin[21] comment that too much homogeneity of studies may stifle generalizations to a larger population, but on the other hand too much study heterogeneity will weaken the results. In all meta-analyses the variance explained by sampling error was 100% (ie, variance was not attributed to systematic error) and this supports the finding that none of the tests for homogeneity were significant. Although the tests for homogeneity would have low power to detect heterogeneity given the small sample sizes, the visual analyses in figure 1 support the conclusion of homogeneity. Because of the finding of homogeneity and because factors for sensitivity analyses were confounded between the studies, the a priori hypotheses were not tested.

Ankle muscle strength. Both effect sizes were positive (fig 1). The treatments and length of follow-up were similar for both Basmajian[9] and Burnside.[11] The study of Burnside[11] had the largest effect size and although this study may have had bias resulting from methodologic problems as outlined in table 2, it did provide placebo biofeedback to the control group. Both studies had statistically significant effects, with Basmajian[9] having reasonably strong methodology. This lends support to the conclusion that EMG biofeedback is effective for improving ankle dorsiflexion muscle strength. We are unaware of any literature that indicates whether the size of the effect, 2.5kg, represents a clinically important difference. It may be important for overcoming foot drop during the swing phase of gait. Because of the small sample size, generalization can only be made to populations that are similar to the samples of these two studies, ie, patients who are more than 4 months poststroke with residual foot dorsiflexion paresis, a minimum passive dorsiflexion range to neutral position, and ability to ambulate with or without aids.

Burnside[11] studied hospital outpatients. The patients had non–weight-bearing treatment that then was progressed to gait training with biofeedback. Biofeedback focused on tibialis anterior or peroneus longus, and training was extended to the gastrocnemius if spasticity was present. This was done for 12 sessions over 6 weeks. Basmajian[9] emphasized training for dorsiflexion. Training was done with the patient sitting with the knee at various angles and there were 15 sessions over 5 weeks.

Gait quality. The point estimates of the three effect sizes are very similar (fig 1). As was the case, it would be expected that the Intiso study[16] would have the largest effect size since it provided a substantially larger number of treatment sessions. Intiso[16] did not provide a detailed description of the biofeedback treatment. In the first phase of treatment, the subjects learned to contract tibialis anterior, and in the second phase, training occurred during gait. Both acute and chronic patients participated daily for 40 sessions over 8 weeks. All patients had mild spasticity. All studies incorporated training of ankle dorsiflexion.

Of the three studies, Basmajian[9] and Burnside[11] performed follow-up measures after their studies. Muscle strength improvement in the biofeedback group was maintained at 6-week follow-up for Burnside.[11] Range of motion and gait quality were maintained in the biofeedback group, whereas the control group relapsed. These follow-up findings are consistent in supporting (EMG) biofeedback as a treatment. Basmajian[9] did follow-up examinations at 4 to 16 weeks after the trial; however, a complete description of these results was not given.

The power of this meta-analysis to detect a difference of 0.5 point was low (.41). This meta-analysis suggests a clinically important effect and further studies with adequate power are needed.

Ankle range of motion. For this outcome, two effect sizes were positive and two were negative (fig 1). There were no consistent differences in samples, treatments or methodologies between the positive and negative studies. Although the study of Burnside[11] is methodologically weaker, it does not have the largest effect size. It is possible that EMG biofeedback does not have an effect on active range of motion. This is supported by the meta-analysis of Glanz[5] in which the effect sizes were not statistically significant for lower extremity or upper extremity range of motion. As indicated by the power analyses, there was adequate power to detect a clinically important difference in our meta-analysis.

Ankle angle during gait. All effect sizes in this meta-analysis were positive (fig 1). The study of Intiso[16] showed the largest effect size. The most striking feature of this study is the large number of treatment sessions provided (40). In this meta-analysis, the power to detect a clinically important difference of 5° was poor (.54). Because of the small sample sizes and small number of studies, further research using ankle angle as an outcome is needed.

Stride length. The point estimates of effect sizes for this outcome are close to 0 (fig 1). Stride length may be primarily determined by movement of the hip, which was not targeted in any of the interventions. The power of this meta-analysis was not adequate, however, to detect a difference of 3cm between the treatment and control groups.

Gait speed. For this outcome, there were four positive effect sizes and two negative effect sizes (fig 1). In both negative studies, biofeedback was not combined with conventional therapy. It is noted that methodologically, the strongest study, Binder,[10] had the largest positive effect size. Despite having the largest number of studies (six), this meta-analysis was not statistically significant. Since walking velocity is an important functional outcome, further larger rigorous studies are needed.

In general, the quality (validity criteria) of the included studies was moderate. All were randomized controlled trials. Only one of the studies (Mulder[17]) did not report blinding of the outcome evaluator. Despite the lack of blinding, the outcomes for gait speed and ankle range of motion had effect sizes which supported traditional therapy. Although, for all studies, the measures at baseline were not comparable between the treatment and control groups (table 2), the use of change scores in the analysis was a strategy to compensate for this. None of the studies randomized the therapists to the treatment and control groups and three studies reported monitoring of the treatments. This may be a source of bias. Future studies should include treatment monitoring to ensure that equal encouragement is given to each study group.

How do our results compare to previous meta-analyses? Schleenbaker and Mainous[3] combined the effect sizes of functional measures and found the mean effect size for lower extremity function was .89. It was not stated whether this was statistically significant. In our study, the effect sizes varied from .07 to 1.17.

Glanz[4] found a pooled effect size of 1.50 (not statistically significant) for lower extremity range of motion measures. This effect size is large compared to ours, which was .07 for ankle range of motion. Glanz[4] included one study that did not have any intervention for the control group, which may have increased the magnitude of the pooled effect size for this meta-analysis.

Further research on EMG biofeedback for lower extremity training after stroke is required. Rigorous methodology and adequate sample sizes are needed. Other areas for research concern optimal frequencies and durations of treatment. The literature about different types of feedback compared to EMG feedback is also an area for review and investigation.

CONCLUSIONS

One meta-analysis provides support for the efficacy of EMG biofeedback for improving ankle dorsiflexor muscle strength. Generalization is limited to patients who are ambulatory, more than 3 months after stroke, with paresis of ankle dorsiflexion and range of motion of ankle dorsiflexion to neutral.

Ankle range of motion and stride length had small effect sizes and were not statistically significant. Although not statistically significant, the point estimates of gait quality, ankle angle during gait, and gait velocity showed moderate effect sizes in favor of biofeedback. The power analyses indicate that larger, more rigorous studies are needed. As more studies are done, more precise estimates of the effects will become possible.

Acknowledgment: The authors acknowledge the assistance of Charles H Goldsmith, PhD, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, with statistical analysis.

References

1. Moreland J, Thomson MA. Efficacy of electromyographic biofeedback compared with conventional physical therapy for upper extremity function in patients following stroke: a research overview and meta-analysis. Phys Ther 1994;74:534-47.
2. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988;138:697-703.
3. Schleenbaker RE, Mainous AG. Electromyographic biofeedback for neuromuscular reeducation in the hemiplegic stroke patient: a meta-analysis. Arch Phys Med Rehabil 1993;74:1301-4.
4. Glanz M, Klawansky S, Stason W, Berkey C, Shah N, Phan H, et al. Biofeedback therapy in poststroke rehabilitation: a meta-analysis of the randomized controlled trials. Arch Phys Med Rehabil 1995; 76:508-15.
5. Mandel AR, Nymark JR, Balmer SJ, Grinnell DM, O'Riain MD.

Electromyographic versus rhythmic positional biofeedback in computerized gait retraining with stroke patients. Arch Phys Med Rehabil 1990;71:649-54.
6. Lund T. Some metrical issues with meta-analysis of therapy effects. Scand J Psychol 1988;29:1-8.
7. Hedges LV, Olkin I. Statistical methods for meta-analysis. Orlando (FL): Academic Press; 1985.
8. Schwarzer R. Meta-analysis programs, Version 5.1. Berlin: Free University of Berlin; 1989.
9. Basmajian JV, Kukulka CG, Narayan MG, Takebe K. Biofeedback treatment of foot-drop after stroke compared with standard rehabilitation technique: effects on voluntary control and strength. Arch Phys Med Rehabil 1975;56:231-6.
10. Binder SA, Moll CB, Wolf SL. Evaluation of electromyographic biofeedback as an adjunct to therapeutic exercise in treating the lower extremities of hemiplegic patients. Phys Ther 1981;61:886-93.
11. Burnside IG, Tobias S, Bursill D. Electromyographic feedback in the remobilization of stroke patients: a controlled trial. Arch Phys Med Rehabil 1982;63:217-22.
12. John J. Failure of electrical myofeedback to augment the effects of physiotherapy in stroke. Int J Rehabil Res 1986;9:35-45.
13. Colborne GR, Olney SJ, Griffin MP. Feedback of ankle joint angle and soleus electromyography in the rehabilitation of hemiplegic gait. Arch Phys Med Rehabil 1993;74:1100-6.
14. Cozean CD, Pease WS, Hubbell SL. Biofeedback and functional electric stimulation in stroke rehabilitation. Arch Phys Med Rehabil 1988;69:401-5.
15. Hurd WW, Pegram V, Nepomuceno C. Comparison of actual and simulated EMG biofeedback in the treatment of hemiplegic patients. Am J Phys Med Rehabil 1980;59:73-82.
16. Intiso D, Santilli V, Grasso MG, Rossi R, Caruso I. Rehabilitation of walking with electromyographic biofeedback in foot-drop after stroke. Stroke 1994;25:1189-92.
17. Mulder T, Hulstijn W, van der Meer J. EMG feedback and the restoration of motor control. Am J Phys Med Rehabil 1986;65:173-87.
18. Shahani BT, Conners L, Mohr JP. Electromyographic audiovisual feedback training effect on the motor performance in patients with lesions of the central nervous system [abstract]. Arch Phys Med Rehabil 1977;58:519.
19. Shiavi RG, Champion SA, Freeman FR, Bugel HJ. Efficacy of myofeedback therapy in regaining control of lower extremity musculature following stroke. Am J Phys Med Rehabil 1979;58:185-94.
20. Takebe K, Basmajian JV. Gait analysis in stroke patients to assess treatments of foot-drop. Arch Phys Med Rehabil 1976;57:305-10.
21. Moher D, Olkin I. Meta-analysis of randomized controlled trials: a concern for standards. JAMA 1995;274:1962-4.
22. Glanz M, Klawansky S, Stason W, Berkey C, Chalmers TC. Functional electrostimulation in poststroke rehabilitation: a meta-analysis of the randomized controlled trials. Arch Phys Med Rehabil 1996;77:549-53.