A Meta-Analysis of the Effectiveness of Occupational Therapy for Older Persons

Mike Carlson, Shan-Pin Fanchiang, Ruth Zemke, Florence Clark

Key Words: aged • outcome study

Given the current health care debate, it is imperative to document the usefulness of various health services for older persons, a rapidly growing population at increased risk for a wide variety of physical and functional impairments. A meta-analysis was conducted to examine the degree of effectiveness of occupational therapy for older persons. For a sample of 15 distinct tests of occupational therapy, a positive unweighted mean effect size of .51 (.54 when corrected for instrument unreliability) was obtained, along with a highly significant cumulative result for treatment success ($p < .001$). Beneficial treatment effects extended to activities of daily living—functional and psychosocial outcomes. The results for physical outcomes suggested a beneficial effect, although not every meta-analytic test yielded significant results. It was concluded that factors such as publication bias or poor study design are incapable of accounting for the positive meta-analytic result and that occupational therapy represents a worthwhile treatment option for older persons.

The growing needs of our nation's burgeoning population of older persons represent one of several trends that may contribute to an impending health care crisis unless major improvements in the health service delivery system are initiated (Pew Health Professions Commission, 1992). In relation to the currently shifting health care market, occupational therapy services for older persons are noteworthy because they conform to recent policy recommendations such as a focus on the treatment of chronic conditions, an emphasis on health promotion and life satisfaction, and availability within home and community-based settings (Clair, 1990; Crabtree, 1991; National Institute of Child Health and Human Development, 1993; Pew Health Professions Commission, 1992). Occupational therapists have used one or more techniques, such as life review, sensory stimulation, activity groups, rehabilitative procedures, home safety instruction, and assistance with adaptive equipment, to enhance the daily functioning and psychosocial well-being of older persons who live either in the community or in nursing homes (Hunt, 1988; Trace & Howell, 1991; Winston, 1981).

Given the current health care dialogue, it is imperative to evaluate the effectiveness of health services for older persons. In this context, policymakers can use such information to make determinations of how to allocate budgetary resources to best achieve national health care goals.

The present study was designed to address the ques-
tion of occupational therapy's effectiveness in treating older persons. To answer this research question, a meta-analysis of the existing outcome literature was performed to determine whether and how much older persons' psychosocial well-being, daily functioning, physical health, or any combination of these are enhanced by occupational therapy. In meta-analysis, individual study outcomes constitute the units of investigation and are quantified through transformation into a common scale termed effect size, which corresponds to the mean amount of therapeutic gain divided by the standard deviation of subjects' scores on a given outcome measure. By statistically pooling the effect sizes from a group of occupational therapy outcome studies, it is possible to arrive at a single overall indication of whether a positive treatment effect exists, as well as to estimate the magnitude of the average degree of therapeutic change. Informative examples of meta-analysis within the occupational therapy literature are provided by Ottenbacher (1982) and Ottenbacher and Petersen (1985).

It should be noted that the present meta-analysis is primarily "Glassian" in its approach (Glass, McGaw, & Smith, 1981; Hunter & Schmidt, 1990; Smith & Glass, 1977) in that it synthesizes the results of diverse studies to address the broad conceptual question of occupational therapy's overall effectiveness for older persons. Thus, the current investigation is not intended to examine the results of any particular type of occupational therapy, a research question that could only be studied if a much larger literature existed on the topic.

Method

Literature Search

The following procedure was used to locate a sample of studies on the effectiveness of occupational therapy in the treatment of older persons. First, a manual search of the following journals was done: American Journal of Occupational Therapy (1979-1994), British Journal of Occupational Therapy (1979-1993), Canadian Journal of Occupational Therapy (1979-1993), Journal of Gerontology (1984-1993), Occupational Therapy in Health Care (1980-1993), Occupational Therapy Journal of Research (1980-1993), and Physical Therapy and Occupational Therapy in Geriatrics (1980-1993). Second, a computerized search of MEDLINE (1979-1994) was conducted with the following key words: occupational therapy plus geriatric, gerontic, gerontology, gerontological, elderly, elders, older, and aged. For each potentially qualifying article, the reference section was examined for possible citations of additional relevant studies. This backward search process was continued until no new citations were found, ensuring that the acquired set of articles was reasonably exhaustive. Fifty-two research publications were gathered in the initial literature search stage.

Inclusion Criteria for Studies

To qualify for entry into the meta-analysis, a study had to meet each of the following four conditions. First, the study had to be published in a book or journal between 1979 and 1994. The decision to exclude earlier studies was based on their potential nongeneralizability to current practice outcomes (e.g., because of subject cohort differences or use of outdated treatment approaches).

Second, the publication had to report a test of an occupational therapy intervention applied to four or more subjects, with a mean age of 60 years or more. In applying this criterion, it was necessary for the words occupational therapy or occupational therapist to occur in conjunction with the treatment. Interventions administered, developed, or overseen by occupational therapists were included, as well as interventions administered by occupational therapy assistants. However, interventions similar in content to occupational therapy, but not conceptualized as occupational therapy by the study's author(s), were rejected (e.g., Carter, Howard, & O'Neil, 1983; Haight, 1992; Powell, Felce, Jenkins, & Lunt, 1979). Further, studies that evaluated occupational therapy used in combination with other treatments, such as physical therapy, were excluded (e.g., Davidoff, Keren, Ring, & Solzi, 1991; Tangeman, Banaitis, & Williams, 1990) because the effects that are specifically associated with occupational therapy are unclear in such instances.

Third, the study had to provide a sufficiently detailed quantitative evaluation of the effectiveness of occupational therapy to permit one or more effect size calculations. Such statistical information included means and standard deviations, reported percentages of positive outcomes, test statistic results (e.g., z test, t test, and F test), presentation of raw data for which the hand calculation of an effect size estimate was possible, or reported means along with sufficient data to estimate a within-condition standard deviation.

Fourth, one or more of the study's effectiveness outcome variables had to correspond to a construct of potential interest to policymakers. Such measures reflected subjects' physical health, daily functioning, psychological well-being, or level of social involvement. Studies using outcome measures such as verbal interaction patterns (Fick, 1993) or the number of repetitions engaged in during reaching exercises (Riccio, Nelson, & Bush, 1990) were rejected on the basis of this criterion.

It should be noted that if an effect size estimate was
calculable, experimental design adequacy was not used as a precondition for entry into the study. This decision, which is consistent with the recommendations of Glass et al. (1981) and Hunter and Schmidt (1990), led to the inclusion in the meta-analysis of several single-group pretest-posttest studies, a type of design Campbell and Stanley (1963) described as a poor example of how to conduct research. Several justifications underscored the decision to allow such designs in the present study. First, these designs were excluded when the outcome variable would be expected on theoretical grounds to improve over time, even in the absence of treatment, because in such cases, spontaneous recovery could produce a spurious positive result. For example, a study of therapy intended to reduce unilateral visual neglect in patients with cerebral vascular accident was eliminated for this reason (Cermak, Trombly, Hauser, & Tiernan, 1991). Second, an analysis of control group outcomes contained within the sampled set of studies revealed that untreated older persons exhibited a small mean pretest-to-posttest decline, on the average, across a wide set of dependent variables ($M_{\text{effect size}} = -.07, n = 33$ assessments; using a single mean per study, $M_{\text{effect size}} = -.04, n = 7$). From this result, it is reasonable to assume that treated subjects would have remained, on the average, approximately the same or declined slightly had they remained untreated. Therefore, single-group pretest-posttest designs are not typically associated with systematic, upwardly biased effect sizes in the gerontic and geriatric literature because of the confounds, such as testing, maturation, statistical regression, or reactivity, identified by Campbell and Stanley (1963). Third, only effect sizes corresponding to occupational therapy versus control group comparisons were included when it was also possible to calculate effect sizes solely on the basis of pretest-posttest change in therapy condition. Finally, the meta-analytic design permitted the opportunity to empirically assess whether studies with weak designs were associated with a different outcome than other designs. Given the absence of a meaningful difference, the inclusion of single-group pretest-posttest designs increased the sample size and thereby promoted a more reliable evaluation of occupational therapy’s effectiveness.

Fourteen of the initial 52 publications gathered met the inclusion criteria for the analysis. Because one study contained two occupational therapy conditions (Bennett & Maas, 1988), the total meta-analytic sample size was 15. (The wide discrepancy in therapeutic outcome for the two conditions examined by Bennett and Maas [see Results section] suggests that the assumption of statistical independence was not violated by the decision to include two observations from the same study.)

**Coding Procedures**

The first author coded all study design variables. The design variables included year of publication, type of research design underlying the effect size statistic, treatment setting (nursing home vs. independent living), mean age of subjects, treatment and control group sample sizes, type of occupational therapy (general activity, life review, other), duration of treatment, and type of outcome variable (activities of daily living [ADL]-functional, physical, or psychosocial).

**Effect Size Calculation**

A separate effect size estimate was generated for each relevant outcome measure contained within each study. Then, separately within each study, the mean of all corresponding effect size estimates was calculated and used as the primary unit of analysis. Additionally, within each study, a single effect size estimate specific to each represented outcome variable domain (i.e., ADL—functional, physical, or psychosocial) was obtained by calculating the mean of all effect sizes falling within the category in question.

Both posttherapy and follow-up results were included, as well as results obtained during the course of the intervention when neither posttherapy nor follow-up outcomes were reported. In cases where raw outcome data were presented along with individual subject ages, only subjects ages 60 years or more were included in the effect size calculations. In instances where more than one control group was present, the control group that was experimentally treated least similarly to the occupational therapy group was used as the basis for effect size comparison. In one study, the condition for the group that the authors labeled as the control group actually consisted of an alternate occupational therapy intervention (Bennett & Maas, 1988). As noted earlier, for the purpose of meta-analysis, this group was treated as a second occupational therapy observation within the study. Two splitting studies (Langlois, Pederson, & MacKinnon, 1991; McPherson, Kreimeyer, Alderks, & Gallagher, 1982) included two or more occupational therapy conditions that each had three or fewer subjects. Because in both studies no meaningful differences in outcome between the occupational therapy conditions were observed, for meta-analytic purposes, the conditions were combined within each study to form respective single observations (with corresponding final $n$ of 6 for the Langlois et al. study and 4 for the McPherson et al. study).
For cases in which outcome measure means and standard deviations were reported for both the occupational therapy and the control group, effect size estimates were calculated by subtracting the control group mean from the therapy group mean and dividing by the pooled estimate of standard deviation (Hunter, Schmidt, & Jackson, 1982). Because many of the sample sizes were small, this pooled estimate was used in lieu of the control group standard deviation to minimize sampling error. When both pretest and posttest standard deviations were reported for occupational therapy and control groups, the pooled average of the four standard deviations (obtained by taking the square root of one fourth the sum of the variances) was used in the effect size formula denominator. In the instance of single-group pretest–posttest designs, the signed degree of mean change was divided by the pooled average of the two raw score (i.e., pretest and posttest) standard deviations.

The formulas found in Glass et al. (1981) were used to calculate effect size estimates on the basis of F, z, t, paired t, or r statistics. In generating estimates for paired t values (or from pretest–posttest mean gain scores and standard deviations), a test–retest correlation of .80 was assumed to estimate the raw score effect size through the formulas contained in Glass et al. For dichotomous outcome measures, probit transformation procedures were used (Glass et al.). Because small samples tend to produce upwardly biased effect size estimates, all resulting values were adjusted for sample size bias in accordance with Hedges’ correction procedure (Hedges & Olkin, 1985).

The effect size calculation procedures were performed independently by the first two authors, and the mean of the two resulting values was used as the final effect size estimate for each observation in question. Reliability for the variable effect size was evaluated with the Spearman–Brown formula (Rosenthal, 1991).

To help interpret the final mean effect size, a correction for attenuation due to measurement unreliability was applied through Hunter et al.’s (1982) procedure of dividing effect size by the square root of the reliability coefficient corresponding to the dependent rating scale in question. To perform this adjustment, reliability estimates were obtained either through direct reporting by the study author(s) or on the basis of referenced instrument development studies concerning the dependent measure; the square root of the mean (using Fisher’s r to z transformation procedure [Hays, 1981]) of all reliability values obtained in this manner was used as a single overall correction factor to be applied to the grand effect size mean.

**Statistical Analysis**

Both unweighted and sample-size–weighted (Cooper, 1989; Hunter & Schmidt, 1990) means and .90 confidence intervals were calculated for the set of effect sizes. Sample size weighting was based on the number of subjects in the treatment condition because a number of studies lacked a control group.

The Stauffer procedure for combining z values (Rosenthal, 1991) was followed to determine the overall probability of the cumulative result for occupational therapy and the results within key subdivisions of the data set. When possible, z values were assigned on the basis of probability values reported in the initial studies. In cases where exact p values were not available, z values were derived, through transformation, from effect size values in conjunction with sample size (Rosenthal). Within studies that included more than one dependent measure, the average z value for the set of relevant outcome variables was used for cumulation purposes.

Because the sample was composed of published studies only, it was important to determine whether publication bias represented a viable explanation for any observed cumulatively significant result (i.e., a bias in favor of occupational therapy’s effectiveness; unpublished studies, which are generally less likely to exhibit a significant outcome, were not represented in the sample). To assess this possibility, a fail-safe n statistic was computed (Cooper, 1989; Rosenthal, 1991). The fail-safe n corresponds to the number of unpublished studies with an average effect size of zero that would need to exist to render an overall meta-analytic outcome nonsignificant.

To facilitate an interpretation of the magnitude of the average effect of occupational therapy, Rosenthal’s (1991) binomial effect size display procedure was followed with respect to the mean reliability-corrected unweighted effect size value. The binomial effect size display technique transforms an effect size value into an equivalent change in the expected success rate of treated subjects relative to controls.

Because diverse subjects, settings, and treatment approaches were included in the meta-analysis, a test of potential moderators of therapeutic efficacy was conducted. To achieve this, zero-order Pearson product–moment correlations (or point–biserial correlations for dichotomous variables) with effect size were performed for the variables’ year of publication, quality of experimental design (randomized control group study [1] vs. single-group pretest–posttest design [−1] vs. other [0]), residency of subjects (community [1] vs. nursing home [0]), mean age of subjects, treatment group sample size, and duration of treatment.

In general, significance tests were conducted at the .05 alpha level and were one tailed in accordance with the
hypothesis of a positive effect for therapy. However, two-tailed tests were performed to assess each of the potential moderator variables noted above. The Statistical Analysis System computing package (SAS Institute Inc., 1988) was used to conduct all data analyses.

Results
The correlation between the first two authors' effect size estimates was .98. On the basis of the Spearman-Brown formula (Rosenthal, 1991), the reliability coefficient for the average of the two values was .99. The mean of the differences in absolute value between the raters' effect size estimates was .05.

Table 1 presents the study characteristics and effect size estimate for each meta-analytic observation. Descriptive summary statistics for each variable are presented at the bottom of the table.

A wide variety of occupational therapy approaches were represented in the sample, including activity groups, life review therapies, sensory techniques, and multistrategy programs. Randomized occupational therapy versus control group pretest-posttest studies (n = 6) were the most prominent type of research design. Slightly more than half of the studies examined the effectiveness of therapy for inpatients.

For both unweighted and sample-size-weighted effect sizes, the mean and .90 confidence interval are presented in Table 2. Also included in the table is the combined significance test result and associated fail-safe n. These outcomes are provided for the entire set of observations, as well as for selected breakdowns on key study characteristics. For both the unweighted and weighted analyses, the overall mean effect size significantly exceeded zero (both p values < .05, as evidenced by the lower confidence interval boundaries, which exceeded zero). Additionally, a highly significant cumulative z value (p < .001) was obtained. Thus, the null hypothesis of no positive effect for occupational therapy was rejected. The fail-safe n corresponding to the cumulative results revealed that 110 nonsampled studies averaging no effect would need to exist to overturn the significance of the obtained cumulative outcome.

Reliability coefficient estimates were available for nine of the outcome instruments included in the therapeutic assessments. The average reported reliability among this set of measures was .89. The average effect size after correcting for unreliability was .54. On the basis of Rosenthal's (1991) binomial effect size display technique, this mean value corresponds to an approximate change in the outcome success rate of 37% for untreated older persons to 63% for those treated. Across individual studies, this figure varied widely (from a -2% to an 82% success rate differential associated with treatment).

Table 2 also reveals that a significant overall mean effect size for therapy was still obtained in both unweighted and weighted analyses when the cases were limited to randomized control group designs. The cumulative p value for this subset of the data was less than .001. Further, the beneficial effects of treatment extended to both outcome areas of ADL-daily functioning and psychosocial well-being (cumulative p values < .001). For physical outcomes, both the unweighted mean effect size and the combined significance test results were significant (p < .05). However, the weighted mean effect size was not reliably greater than zero for the physical outcome domain.

Correlations computed to examine the relationships between potential moderator variables and effect size were all nonsignificant. Among the six variables tested, in all cases the resulting r² value was less than .18 (all two-tailed p values > .10). Within the sample, less than 1% of the effect size variability was associated with the quality of study design variable (r² = .003).

Discussion
The central finding of this investigation is that occupational therapy services for older persons produce positive results across a wide range of treatment contexts. This outcome strongly suggests that occupational therapy can contribute to the quality of life of our nation's growing senior population and that such services can be expected to make a meaningful contribution when included as a health care option for older persons. Because the positive result was still obtained when the analysis was limited to randomized control group studies, and because study design quality was unrelated to outcomes, it cannot be argued that the inclusion of weaker designs was responsible for the significant overall finding.

As noted in the introductory section of this article, this analysis was not intended to examine the effectiveness of any particular type of treatment but rather to provide a wider evaluation that encompasses a variety of treatment populations of older persons, settings, and therapeutic approaches. Therefore, with respect to health care policy, this study's results suggest that at the level of across-the-board decision making, occupational therapy is, in general, beneficial when used to assist older persons.

Publication Bias
The obtained overall fail-safe n value revealed that 110 nonsampled studies averaging an effect size of zero would need to exist to render the cumulative positive outcome
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year Published</th>
<th>Design</th>
<th>Underlying Effect Size Statistic</th>
<th>Description of Subjects</th>
<th>Mean Subject Age (years)</th>
<th>Treatment Group n</th>
<th>Control Group n</th>
<th>Type of Occupational Therapy</th>
<th>Duration of Treatment (weeks)</th>
<th>Outcome Variables</th>
<th>Mean Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnetz, Eyre, &amp; Theorell</td>
<td>1982;1983</td>
<td>Nonequivalent control group</td>
<td>Residence of senior apartment building in Sweden</td>
<td>78.2</td>
<td>30</td>
<td>30</td>
<td>Interest and activity groups</td>
<td>26</td>
<td>Social activation, depression, psychosomatic complaints</td>
<td>+.34</td>
<td></td>
</tr>
<tr>
<td>Bennett &amp; Maas</td>
<td>1988</td>
<td>Single-group pretest-posttest</td>
<td>British, Irish, and Australian residents of nursing homes or hostels</td>
<td>82.2</td>
<td>13</td>
<td>—</td>
<td>Music-based life review</td>
<td>6</td>
<td>Life satisfaction, ego integrity</td>
<td>+.87</td>
<td></td>
</tr>
<tr>
<td>Bennett &amp; Maas</td>
<td>1988</td>
<td>Single-group pretest-posttest</td>
<td>British, Irish, and Australian residents of nursing homes or hostels</td>
<td>81.2</td>
<td>13</td>
<td>—</td>
<td>Verbal life review</td>
<td>6</td>
<td>Life satisfaction, ego integrity</td>
<td>−.02</td>
<td></td>
</tr>
<tr>
<td>Burton</td>
<td>1980</td>
<td>Multiple baseline and intervention (ABAB)</td>
<td>British psychogeriatric ward residents (nearly all with dementia)</td>
<td>82.0</td>
<td>10</td>
<td>—</td>
<td>Activity and social reinforcement of positive behaviors</td>
<td>1.7</td>
<td>Reception to people and things in environment, absence of inappropriate or repetitive behavior</td>
<td>+.15</td>
<td></td>
</tr>
<tr>
<td>Gauthier, Dalziel, &amp; Gauthier</td>
<td>1987</td>
<td>Randomized experimental versus control group with pretesting and posttesting</td>
<td>Canadians with Parkinson’s disease who resided in the community</td>
<td>63.1</td>
<td>30</td>
<td>29</td>
<td>Rehabilitative therapy, including mobility, dexterity, ADL, and education components</td>
<td>5</td>
<td>Functional autonomy, finger dexterity, reduction in physical and motor symptoms (posttest, 6-month and 1-year follow-up)</td>
<td>+.30</td>
<td></td>
</tr>
<tr>
<td>Jongbloed &amp; Morgan</td>
<td>1991</td>
<td>Randomized experimental versus control group with pretesting and posttesting</td>
<td>Canadian patients with CVA previously discharged from a rehabilitation hospital</td>
<td>68.9</td>
<td>20</td>
<td>20</td>
<td>Assistance with participation in leisure activities</td>
<td>5</td>
<td>Satisfaction with activity involvement (posttreatment, 13-week follow-up)</td>
<td>+.13</td>
<td></td>
</tr>
<tr>
<td>Kiernat</td>
<td>1979</td>
<td>Correlational</td>
<td>American nursing home residents with confusion</td>
<td>—</td>
<td>19</td>
<td>—</td>
<td>Life review</td>
<td>10</td>
<td>Social and adaptive behavior</td>
<td>+.50</td>
<td></td>
</tr>
<tr>
<td>Kirchman, Reichenbach, &amp; Giambalvo</td>
<td>1982</td>
<td>Single-group pretest-posttest</td>
<td>American well older persons attending a medical center nutrition site</td>
<td>69.7</td>
<td>31</td>
<td>—</td>
<td>Avocational and recreational activities, physical exercise, and educational sessions</td>
<td>26–35</td>
<td>ADL, household activities, economic resources, life satisfaction, general affect, perceptions of health</td>
<td>+.12</td>
<td></td>
</tr>
<tr>
<td>Langlois, Pedersen, &amp; MacKinnon</td>
<td>1991</td>
<td>Single-group pretest-posttest</td>
<td>Canadians living in institutions or in the community and who have spastic hemiplegia due to stroke</td>
<td>70.8</td>
<td>6</td>
<td>—</td>
<td>Hand splinting</td>
<td>4</td>
<td>Reduction in spasticity (posttest)</td>
<td>+.49</td>
<td></td>
</tr>
<tr>
<td>McPherson, Kreimeyer, Addlerks, &amp; Gallagher</td>
<td>1982</td>
<td>Single-group pretest-posttest</td>
<td>American older persons with medical conditions producing hypertonus</td>
<td>—</td>
<td>4</td>
<td>—</td>
<td>Dorsal or volar hand splinting</td>
<td>5</td>
<td>Reduction in hypertonus</td>
<td>+.77</td>
<td></td>
</tr>
</tbody>
</table>
nonsignificant. Given the current area of investigation, the existence of 110 such studies does not seem possible. Because of the paucity of outcome investigations on the topic, nonsignificant findings may be more publishable than is the case for most other types of research literature. For example, a number of the studies included in the sample failed to detect a significant effect (Jongbloed & Morgan, 1991; Robichaud, Hebert, & Desrosiers, 1994; Stevens-Ratchford, 1993). Because studies with nonsignificant findings are often deemed publishable in this area, there was a diminished possibility of capitalizing on publication bias that favored positive outcomes. It should also be noted that many of the nonqualifying studies (e.g., studies that would have otherwise qualified but lacked adequate statistical reporting or were published before the sampling time frame) were associated with a positive effect of therapy (e.g., Porszt-Miron, Florian, & Burton, 1988; Turbow, 1975; Turton & Fraser, 1990). If such outcomes are only sometimes present in the set of existing unpublished studies, then the number of nonsampled articles exhibiting an overall null effect would need to be much greater than 110 to discount the cumulative positive effect observed for occupational therapy. Given these considerations, we conclude that publication bias cannot explain the global positive result.

### Effect Size Magnitude

The present average effect size (.51 uncorrected for unre-
Table 2
Effect Size Summary Data and Cumulative Z Test Results for Combined Data and by Key Breakdowns

<table>
<thead>
<tr>
<th>Sample</th>
<th>Unweighted Mean</th>
<th>Unweighted Combined Significance Test</th>
<th>Weighted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>CI for M</td>
</tr>
<tr>
<td>All observations</td>
<td>15</td>
<td>.51</td>
<td>.27-.75</td>
</tr>
<tr>
<td>Randomized control group designs</td>
<td>6</td>
<td>.40</td>
<td>.14-.66</td>
</tr>
<tr>
<td>ADL/daily function outcomes</td>
<td>6</td>
<td>.67</td>
<td>.08-1.26</td>
</tr>
<tr>
<td>Physical health outcomes</td>
<td>6</td>
<td>.32</td>
<td>.09-.55</td>
</tr>
<tr>
<td>Psychosocial well-being outcomes</td>
<td>10</td>
<td>.37</td>
<td>.20-.54</td>
</tr>
</tbody>
</table>

Note. ADL = activities of daily living. CI = confidence interval. ES = effect size.

liability, .54 corrected) is somewhat smaller than the corresponding values obtained in two other meta-analyses of occupational therapy's effectiveness: .58 for the treatment of depression (Devereaux & Carlson, 1992) and .79 for sensory integrative procedures (Ottenbacher, 1982). However, the current mean is not significantly smaller than either of these previous values. Across these three meta-analyses of the effectiveness of occupational therapy, the mean corrected effect size (assuming an average reliability of .9 for outcome measures contained in the depression and vestibular stimulation analyses) is .66. According to Rosenthal’s (1991) binomial effect size display, this figure corresponds to a near doubling (34.5% to 65.5%) of the rate of successful outcomes among treated patients relative to those not receiving therapy.

It should be stressed, however, that it is potentially problematic to compare or combine effect sizes across different research areas (Cooper, 1989; Hunter & Schmidt, 1990). For example, between-literature variations in subjects' potential for therapeutic change, the types of outcome domains, or the selected meta-analytic techniques and decisions may render the corresponding effect size estimates nonequivalent. In this regard, the average effect size in the present study may reflect a low estimate (relative to most other analyses) due to a series of meta-analytic decisions that were intended to produce conservative effect size values and thereby strengthen the case for therapeutic effectiveness given a significant finding. For example, in cases where intermediate numerical estimates were required to calculate effect sizes, values were chosen so as to err on the side of understating the magnitude of therapeutic change. Additionally, in one study, pretest differences in favor of the control group were ignored to produce a conservatively low series of effect sizes (Reichenbach & Kirchman, 1991).

The unweighted mean reliability-corrected effect size of .54 in the current study also needs to be viewed in light of its wide confidence limits. It would be premature to conclude that, in general, occupational therapy for older persons is associated with an average effect size in the vicinity of .5. Within the bounds of statistical reasoning, one cannot reject the possibility that the corresponding unweighted population mean is as small as .27 or as large as .75, values that, although positive, are associated with very different substantive meanings with regard to the magnitude of treatment gains.

**Outcome Domains**

The tendency in this study for occupational therapy's beneficial effects to radiate to a range of outcome areas (i.e., ADL–daily functioning, psychosocial well-being, possibly physical health) substantiates the profession's longstanding belief in the ability of prescribed occupation to enhance multiple areas of human functioning. For each of these three outcome domains, the most sensitive test (the combined significance test) produced a positive result (except in the case of physical health outcomes) along with a fail-safe n statistic that seems large enough to rule out publication bias as a viable explanation for the findings.

With respect to physical health outcomes, the nonsignificant result for the weighted analysis shows that the larger effect sizes observed within this domain were associated with studies that used small sample sizes. However, the finding of a positive (though nonsignificant) mean weighted effect size was, in principle, consistent with the unweighted analysis. It is possible that the nonsignificant result stemming from the weighted analysis reflects the small (n = 6) meta-analytic sample size in conjunction with an effect that is relatively hard to detect. For example, physical outcomes may be more strongly linked to extratherapeutic factors among older persons, such as general age-related health declines that are independent...
of treatment, and their evaluation may therefore be associated with increased variability that may partially mask the effects due to treatment.

Although moderators of effect size were not reliably documented (possibly because of the small sample size), the wide variation in study results (effect size range = 2.09) suggests that the outcomes of differing occupational therapy services for older persons may not be uniform. Because an overall positive effect has been documented, further research needs to address the question of which therapeutic approaches work best with which types of older patients to achieve selected types of therapeutic outcomes.

Conclusion

The results of this study have practical importance, particularly if shared with administrators, policymakers, and others who influence older persons' access to health care. The findings allow one to be confident in concluding that occupational therapy, in a global sense, is effective in achieving crucial outcomes in older persons and to argue that in contrast to other interventions for which comparable data are not available, there is strong justification for the need to incorporate occupational therapy as a part of overall health care plans. Furthermore, occupational therapists may wish to inform consumers of the positive findings as a way of providing them with a source of hope. The treatment of illness and disability in the later years through occupational therapy is likely to result in significant positive outcomes.

Acknowledgments

This investigation was supported by National Institute on Aging research grant # ROI AG1810-01, The Effectiveness of Two OT Treatments for the Elderly (F. Clark, Principal Investigator), the American Occupational Therapy Foundation (AOTF) Center at the University of Southern California for the Study of Occupation and Its Relation to Adaptation (F. Clark, Principal Investigator); and a grant from the RGK Foundation and Lumes, Inc., that was arranged by AOTF.

References


February 1996, Volume 50, Number 2