ANTHROZOÖS VOLUME 20, ISSUE 2 PP 167–180

Do Animal-Assisted Activities Effectively Treat Depression? A Meta-Analysis

Megan A. Souter and Michelle D. Miller
Department of Psychology, Northern Arizona University, USA

Address for correspondence: Dr Michelle Miller, Department of Psychology, Northern Arizona University, Flagstaff, AZ 86011-5106, USA. E-mail: Michelle.Miller@nau.edu

ABSTRACT We conducted a meta-analysis to determine the effectiveness of animal-assisted activities (AAA) and animal-assisted therapy (AAT) for reducing depressive symptoms in humans. To be included in the meta-analysis, studies had to demonstrate random assignment, include a comparison/control group, use AAA or AAT, use a self-report measure of depression, and report sufficient information to calculate effect sizes, a statistical standardization of the strength of a treatment effect. Five studies were identified for analysis. The aggregate effect size for these studies was of medium magnitude and statistically significant, indicating that AAA/AAT are associated with fewer depressive symptoms. This analysis revealed gaps in the research on AAA/AAT, which we attempted to identify in order to better understand the factors that make AAA and AAT effective at reducing depression.

Keywords: animal-assisted activities, animal-assisted therapy, depression, meta-analysis, pet therapy

Animal-assisted activities (AAA) and animal-assisted therapy (AAT) are becoming increasingly common as therapeutic interventions in health care facilities such as nursing homes and hospitals. Various terms have been used to describe AAA and AAT such as pet therapy, pet-facilitated therapy, pet-assisted therapy, animal-facilitated therapy, and animal visitation (Connor and Miller 2000). The Delta Society, a leading international, not-for-profit organization that provides training for AAA and AAT practice, categorizes these types of interventions under the preferred terms of animal-assisted activities (AAA) and animal-assisted therapy (AAT; "Standards of Practice" 1996). The following are formal definitions of AAA and AAT:

> AAA provide opportunities for motivational, educational, recreational, and/or therapeutic benefits to enhance quality of life. AAA are delivered in a variety of environments by specially trained professionals, paraprofessionals, and/or volunteers, in association with animals that meet specific criteria ("Standards of Practice" 1996).

> AAT is a goal-directed intervention in which an animal that meets specific criteria is an integral part of the treatment process. AAT is directed and/or delivered by a health/human service provider working within the scope of practice of his/her profession. AAT is designed to promote improvement in human physical, social, emotional, and/or cognitive functioning. AAT is provided in a variety of settings and may be group or individual in nature. This process is documented and evaluated.

Although by definition AAA and AAT are distinguishable, in actual practice the two are not clearly differentiated and at times may overlap.

In most situations, AAA and AAT are provided to individuals or groups with the aid of volunteers and/or health care practitioners (Barba 1995). The intention of these activities is to promote health and well-being among those individuals who are affected by a variety of issues, including depression, autism, emotional disorders, Alzheimer's, and physical disabilities. These types of human–animal interactions can be experienced in a variety of settings such as hospitals, nursing homes, hospices, rehabilitation facilities, oncology units, acute and critical care units, psychiatric facilities, psychotherapy, and prisons (Connor and Miller 2000).

A number of researchers have attempted to show that AAA and/or AAT positively affect a wide array of outcome variables such as blood pressure, heart rate, exercise level, stress, social interaction, anxiety, loneliness, and depression. However, these research studies have produced mixed results regarding the effectiveness of AAA and AAT, and the picture is further clouded by the fact that these studies vary along such important dimensions as setting, patient population, type of animal, duration of visits, and frequency of animal interactions. We conducted the present study to determine whether existing research supports the effectiveness of AAA/AAT in therapeutic settings.

In such settings, AAA and AAT may produce benefits related to health and quality of life or well-being. Some of the claimed physiological benefits of these human–animal interactions include decreased blood pressure, increased activity and mobility, decreased heart rate, improved recovery rate, better coping with illness, and general benefits from physical contact or touch (Delta Society n.d.; McCulloch 1983; Boldt and Dellman-Jenkins 1992). Potential psychological benefits include increased empathy, relaxation, improved self-esteem and acceptance, stress and anxiety reduction, reality orientation, nurturing skills, mental stimulation, decreased loneliness, increased positive affect, and opportunities to reminisce about past experiences with pets or life experiences in general (Delta Society n.d.; McCulloch 1983; Boldt and Dellman-Jenkins 1992; Barba 1995; Brasic 1998).

Does AAT/AAA actually deliver these potential benefits? Some studies seem to offer evidence to this effect, reporting improved social interaction in residents of institutional care facilities (Francis, Turner and Johnson 1985; Winkler et al. 1989; Fick 1992; Perelle and Granville 1993; Batson et al. 1998). Positive effects have also been reported for patient mood (Crowley-Robinson, Fenwick and Blackshaw 1996; Jessen, Cardiello and Baun 1996). Although such results are promising, much of the research lacks the methodological leverage to make clear inferences about AAA/AAT effects. One common limitation is the absence of random assignment (Norris 1983; Hagman 1997). Some studies also fail to include control groups (Winkler et al. 1989; Vaughan 1990; Perelle and Granville 1993). For those studies using observational methods, observer bias and blinding are also issues to consider (Jendro, Watson and Quigley 1984; Perelle and Granville 1993).

Some of the research indicates that AAA and AAT have little impact. For example, one study demonstrated no significant differences between a pet therapy group and an exercise control group on an observational scale assessing self-care functioning, disoriented behavior, depressed or anxious mood, irritable behavior, or withdrawn behavior (Zisselman et al. 1996). Another study indicated no beneficial effects of AAA for psychological or functional variables; however, participants showed significantly more purposeful behavior (i.e., "physical movements/gestures and/or verbal expressions exhibited with the intent of having needs or wants met," p. 419) while in the AAA session (Jendro, Watson and Quigley 1984). In addition, other studies report mixed findings in regard to the

effectiveness of AAA and AAT (Harris, Rinehart and Gerstman 1993; Batson et al. 1998). The variability in the designs and types of programs being implemented makes it difficult to determine which programs or which components of programs could be the most effective. Thus, a narrative review cannot clearly elucidate whether or not AAA or AAT is useful.

Meta-analysis, on the other hand, offers a more systematic way to summarize and evaluate the diverse literature on the effectiveness of AAA and AAT. Meta-analysis is a procedure used to summarize the quantitative results of empirical research studies (Lipsey and Wilson 2001). The process of executing a meta-analysis involves gathering comparable empirical research on the topic of interest and coding and analyzing study characteristics and statistical findings that are reported in individual studies (Lipsey and Wilson 2001). Statistical procedures are then performed to quantify results across studies in order to summarize the research finding or question of interest. In selecting studies for our meta-analysis, we chose to focus on those that used depression or depressed mood as an outcome variable. Our reasoning was as follows: Depression is a serious condition that affects many people (approximately 19 million adults in the US alone; National Institute of Mental Health [NIMH] n.d.) and has disabling effects (National Institute of Health [NIH] 1992).

Depression can also intensify the effects of co-existing physical illnesses, an important consideration given that it often co-exists with other serious diseases such as heart disease, stroke, cancer, and diabetes (NIMH n.d.). A recent NIMH study (Pratt et al. 1996) suggested that an individual with depression has an increased risk for having a heart attack in the future. According to the Global Burden of Disease study in 1990, depression is predicted to be the second cause of disease burden in the nation, second to ischemic heart disease for 2020 (Murray and Lopez 1996). Depression is a particularly pressing concern among older adults, especially institutionalized older adults, with prevalence rates of major and minor depression among older adults in nursing homes reaching 15% to 25% (NIH 1992).

Methods

Literature Search

The studies for this meta-analysis were located by searching the following databases: PsychINFO, ERIC, Social Services, Sociological Abstracts, Academic Search Premier, NIMH, Health and Wellness Resource Center, PubMed, Agricola, Melvyl Catalog, Cinahl, Health Source-nursing/academic edition, Kluwer, Medline, Project Muse, Science Direct, Cochrane Database of Systematic Reviews, and Dissertation Abstracts International. Key terms used in searches included pets, pet therapy, companion animal, pets and hospitals, pets and human health, pet-facilitated therapy, pet-facilitated therapy and depression, pet therapy and depression, bonding-human-pet and aged, bonding-human-pet and nursing homes, bonding-human-pet and psychotherapy, pets and elderly, animals and therapeutic use, human-pet and bonding, human–animal relationships, pets-social aspects, animal-assisted activities, animal-assisted therapy, and animal therapy. The journal *Society and Animals* was also searched for articles related to AAA and AAT.

Websites on AAA and AAT and websites of universities with programs focusing on human–animal relationship research were reviewed for possible research articles, including www.deltasociety.org, www.dog-play.com/therapy.html, www.censhare.umn.edu, www.therapyanimals.org, www.vetmed.ucdavis.edu/CCAB/paws.htm, www.latham.org, www.tufts.edu/vet/cfa/index.html, and www.vet.upenn.edu/cias. Articles were restricted to those published in English; however, there were no restrictions for year of publication. If abstracts of the studies were available, they were read, and the studies were retrieved if they appeared to be related to the research question of focus. The reference lists of the collected papers were reviewed, and articles that appeared to meet the meta-analysis inclusion criteria were retrieved and examined.

Because this meta-analysis focused on AAA and/or AAT, studies on pet ownership were excluded. Articles that appeared to concentrate on depression and AAA and/or AAT were searched for and examined. After excluding articles on pet ownership, the literature search led to 165 articles

that were reviewed to determine if they would be included in the meta-analysis (not all of the 165 articles examined depression). Of the 165 articles, 105 were summaries/reviews, theoretical papers, editorials, methodological papers, and anecdotal papers, which did not include data that could be analyzed. Of the 60 remaining studies, approximately half did not include a measure of depression and only five met the selection criteria set for this meta-analysis (Brickel 1984; Struckus 1989; McVarish 1994; Wall 1994; Panzer-Koplow 2000).

Selection Criteria

In order to be included in the meta-analysis, each study was required to meet several criteria. These criteria were random assignment, inclusion of a control group, exposure to some form of AAA or AAT, and a measure of depressive symptoms. Restrictions for the depression measures were that they had to be a self-report questionnaire, either filled out by the participant or verbally administered by the researcher or a volunteer. More importantly, all studies had to report results in sufficient detail for us to calculate effect sizes.

The five studies that met the selection criteria for this meta-analysis consisted of four dissertations and one conference paper (Brickel 1984; Struckus 1989; McVarish 1994; Wall 1994; Panzer-Koplow 2000). Two additional studies were retrieved that met all criteria except for providing sufficient data to calculate effect sizes. Attempts were made to contact the authors of these studies by calling departments of the schools where they conducted their research. The dissertation chairperson of one of the authors was also reached to inquire about contact information of the author. These attempts did not result in sufficient information to contact either author; therefore, these studies were not included in the meta-analysis.

Variable-Depressive Symptom Instruments

The studies included in the meta-analysis used the following validated self-report measures of depressive symptoms: Zung Self-Rating Depression Scale (Zung 1965; Zung 1974), the Beck Depression Inventory (BDI; Beck et al. 1961), the Beck Depression Inventory, II (BDI-II; Beck, Steer and Brown 1996), the Geriatric Depression Scale (GDS; Brink et al. 1982; Yesavage et al. 1983), and the NIMH Mood Scales-Elderly Depressed factor (MS-E; Raskin and Crook 1988).

Reported posttest means and standard deviations for the chosen depression measures were used to calculate an effect size statistic for each of the five studies. The posttest means and standard deviations included those from both treatment and control groups. The sample size of each group was also used in the calculation. Due to variations in the operationalization of depression across studies, the standardized mean difference between the treatment and control groups of each study was the effect size statistic chosen for this meta-analysis.

Study Characteristics

In order to qualitatively describe our studies, we noted the following characteristics of each one: (a) type of publication; (b) publication year; (c) role of experimenter in treatment; (d) discipline of research; (e) sample source (e.g., nursing home, hospital, psychiatric hospital, hospital-based nursing home care unit, VA medical hospital, day care center, rehabilitation unit, homebound); (f) mean age of sample at beginning of intervention; (g) predominant sex of sample (i.e., proportion of females in sample); (h) predominant ethnicity; (i) special characteristics of participants (e.g., psychiatric, dementia, diagnostics); (j) type of assignment to conditions; (k) whether or not equivalence of groups was tested at pretest; (l) pretest differences, if tested; (m) total sample size at beginning of study; (n) treatment group sample size; (o) control groups sample sizes; (p) type of treatment (e.g., animal-assisted therapy, animal-assisted activity, mascot or resident animal, live-in animals in individual rooms, animal-assisted activity and current events discussion); (q) type of animal interaction for treatment groups (e.g., individual or group); (r) duration (in minutes) of each visit for treatment group and control groups; (s) number of total visits for treatment group and control groups; (t) duration of treatment in weeks (i.e., length of time between pre and posttesting) for

treatment group and control groups; (u) intensity of treatment (e.g., petting, grooming, walking); (v) type of intervention for control groups (e.g., person visit, person and stuffed animal visit, person and photo of animals, recreational activity, no activity or treatment, conventional therapy, exercise, current events discussion); (w) delivery of intervention for treatment group and control groups (e.g., facility staff, researcher, volunteers or owners of pets, veterinarian or veterinarian technician, not reported); (x) number of volunteers per participant for animal-assisted activities; (y) type of animal; (z) survey design for variable of interest; (aa) dependent variable (e.g., depression, anxiety, stress, loneliness, affect, mood) and instrument used to measure the dependent variable; (bb) other dependent variables; and (cc) additional types of treatment groups. Given the small number of studies in the analysis, we did not examine how each of these characteristics influenced the outcome of treatment or whether any influence was statistically significant.

Effect Size Calculations

We used the formulas for the standardized mean difference described by Lipsey and Wilson (2001) to calculate effect sizes (also known as Cohen's *d*). The standardized mean difference involves calculating the pooled standard deviation, the biased effect size, the corrected effect size, the standard error of the effect size, and the inverse variance weight (see formulas 1–5 in Table 1). This effect size statistic was used because it allowed us to compare the effects of AAA and/or AAT on depression among control and treatment groups to determine if AAA and AAT are effective as therapeutic interventions. The standardized mean difference effect size formulas were most appropriate to assess this because they are used to compare the means of two groups on a dependent variable when the dependent variable is continuous and not operationalized in the same manner across research studies (Lipsey and Wilson 2001).

The effect size formula includes subtracting group means for the numerator and estimating the pooled standard deviation for the denominator (Lipsey and Wilson 2001). When comparing treatment and control groups, a positive effect size is an indication that the intervention was effective. For situations when higher scores on a dependent measure indicate progress, it is appropriate to subtract the control group mean from the treatment group mean when calculating the numerator of the effect size formula. Because lower scores rather than higher scores indicated progress (e.g., less depressed) on all of the dependent measures examined in the present meta-analysis, the order of subtraction for the numerator was reversed (i.e., the treatment group mean was subtracted from the control group mean) (Lipsey and Wilson 2001). By making this change in the subtraction order for the numerator, a positive effect size remained as an indicator that the intervention was effective. All studies reported the required information in order to calculate the effect size; however, for one study (Panzer-Koplow 2000) the pretest standard deviations were used as an estimate of the pooled standard deviation because posttest standard deviations were not reported.

After the effect size was calculated, it was used to compute a corrected or unbiased effect size estimate (Lipsey and Wilson 2001). A corrected effect size is necessary to calculate because the original effect size value tends to be upwardly biased when estimated on small sample sizes (Hedges 1981). In addition to calculating the corrected effect size estimate, the standard error of the effect size and the inverse variance weight are calculated (Lipsey and Wilson 2001). These statistics are computed in order to obtain a weight for each effect size. Because the various effect sizes are based on a range of sample sizes, weights are used to represent the precision of each effect size (Lipsey and Wilson 2001). These weights are derived from the standard error of the effect size by calculating the inverse of the squared standard error value. This value is termed the inverse variance weight.

The analysis of the meta-analytic data involved using the corrected effect sizes, standard errors of the effect sizes, and inverse variance weights of the effect sizes to analyze the effect size mean and distribution (see formulas 6–10 in Table 1). The effect size mean was found by weighting each

Table 1. Equations used in the meta-analysis.

	Statistic	Equation
1.	Pooled standard deviation	$s_p = \sqrt{\frac{(n_{G1}-1)\,s_{G1}^2 + (n_{G2}-1)\,s_{G2}^2}{(n_{G1}-1)+(n_{G2}-1)}}$
2.	Biased effect size	$ES_{sm} = \frac{\overline{X}_{G1}-\overline{X}_{G2}}{s_p}$
3.	Corrected effect size	$ES'_{sm} = \left[1 - \frac{3}{4N-9}\right] ES_{sm}$
4.	Standard error of the effect size	$SE_{sm} = \sqrt{\frac{n_{G1}+n_{G2}}{n_{G1}n_{G2}} + \frac{(ES'_{sm})^2}{2(n_{G1}+n_{G2})}}$
5.	Inverse variance weight	$w_{sm} = \frac{1}{SE_{sm}^2} = \frac{2n_{G1}n_{G2}(n_{G1}+n_{G2})}{2(n_{G1}+n_{G2})^2 + n_{G1}n_{G2}(ES'_{sm})^2}$
6.	Weighted mean effect size	$\overline{ES} = \frac{\Sigma(w_i ES_i)}{\Sigma w_i}$
7.	Standard error of the mean effect size	$SE_{\overline{ES}} = \sqrt{\frac{1}{\Sigma w_i}}$
8.	Confidence intervals around the mean effect size	$\overline{ES}_L = \overline{ES} - z_{(1-\alpha)}(SE_{\overline{ES}})$ $\overline{ES}_U = \overline{ES} + z_{(1-\alpha)}(SE_{\overline{ES}})$
9.	z-test	$z = \frac{\lvert\overline{ES}\rvert}{SE_{\overline{ES}}}$
10.	Test for homogeneity	$Q = (\Sigma w_i ES_i^2) - \frac{(\Sigma w_i ES_i)^2}{\Sigma w_i}$
11.	Method of moments estimate for random variance component	$v_\theta = \frac{Q-(k-1)}{\Sigma w_i - (\Sigma w_i^2 / \Sigma w_i)}$
12.	Percentage of total variation across studies due to heterogeneity	$I^2 = 100 \times \frac{(Q-\mathrm{df})}{Q}$
13.	Fail-safe *N*	$k_O = k\left[\frac{\overline{ES}_k}{\overline{ES}_c} - 1\right]$

Table 2. Study descriptions and findings.

Study	Characteristic	Description
Brickel (1984)	Participants	Total of 15 participants from a nursing home unit in a hospital
	Conditions/Groups	Conventional therapy (*n* = 5); Pet-facilitated psychotherapy (*n* = 5); No-treatment control group (*n* = 5), no activity
	Animal-Assisted Activity Session	8 sessions over a 4-week period; Each session approximately 67.5 minutes; Dog
	Self-Report Measure Examined in Meta-Analysis	Zung Self-Rating Depression Scale
McVarish (1994)	Participants	Total of 74 participants from two psychiatric hospitals
	Conditions/Groups	Pet-facilitated therapy group (*n* = 24); Animal photograph group (*n* = 26); Control group (*n* = 24), no activity
	Animal-Assisted Activity Session	One 40-minute group visit; Dogs and cats
	Self-Report Measure Examined in Meta-Analysis	Beck Depression Inventory
Panzer-Koplow (2000)	Participants	Total of 35 participants from a nursing home
	Conditions/Groups	Animal-assisted therapy group (*n* = 15); Control group (*n* = 19), no activity
	Animal-Assisted Activity Session	One individual visit per week, lasting 15 minutes, over a ten-week period; Dog
	Self-Report Measure Examined in Meta-Analysis	Beck Depression Inventory II
Struckus (1989)	Participants	Total of 50 participants from a nursing home
	Conditions/Groups	Animal visitation group (*n* = 25); Comparison group (*n* = 25), alternative recreational activity (i.e., sing-alongs, reading aloud)
	Animal-Assisted Activity Session	24 individual visits over a 12-week period, with each visit lasting approximately 20 minutes; Dog
	Self-Report Measure Examined in Meta-Analysis	Geriatric Depression Scale II
Wall (1994)	Participants	Total of 80 participants from nursing home
	Conditions/Groups	Dog with visitor (*n* = 20); Visitor with a stuffed animal (*n* = 20); Visitor alone (*n* = 20); No treatment control condition (*n* = 20), no activity
	Animal-Assisted Activity Session	3 individual visits over a two and a half week period, with each visit lasting approximately 8 minutes; Dog
	Self-Report Measure Examined in Meta-Analysis	NIMH Mood Scales - Elderly (Depressed Factor)

effect size by its inverse variance weight, summing these values, and dividing by the sum of weights (Lipsey and Wilson 2001). The standard error of the mean was also calculated. Confidence intervals for the mean effect size were found using the standard error of the mean effect size and a critical *z*-value. To test the significance of the mean effect size, a *z*-test was computed. The effect size distribution was also tested for homogeneity using the *Q* statistic.

Results

Study Characteristics

All five chosen studies were conducted in the United States, and published between 1984 and 2000 (see Table 2 for a brief description of each study). The studies took place in various institutional settings including a hospital-based nursing home care unit, a psychiatric hospital, and nursing homes. The mean age of individuals participating in these studies ranged from 47 to 85. The majority of these studies included females (50% to 95%), with the exception of one study, which included only males (Brickel 1984). The predominant ethnicity of participants in these studies was Caucasian (i.e., greater than 60% were Caucasian).

The sample size for treatment groups ranged from 5 to 25. For control groups, the sample size also ranged from 5 to 25. For the studies that had more than one comparison group, the comparison group whose intervention included no treatment or activity was chosen to be compared with the treatment group for the meta-analysis. However, the comparison group for one of the studies in the meta-analysis engaged in a regularly scheduled recreational activity (Struckus 1989). The type of treatment for four of the five studies involved AAA or pet visitations. One study used AAT as the intervention.

All animal interactions involved individual visits with the exception of one study using group visits (McVarish 1994). The duration in minutes of each visit for treatment groups among the five studies ranged on average from 8 minutes to 67.5 minutes. The number of total visits for treatment groups ranged from one visit to 24 visits. The duration of treatment in weeks (length of time between pretest and posttest) ranged from 0.71 weeks to 12 weeks. The type of animal participating in the intervention was a dog for all studies. The study facilitating group visits also included a cat as part of the animal intervention (McVarish 1994). Four out of the five studies reported a significant reduction in depression for the animal intervention groups from pre to posttest. The Panzer-Koplow (2000) study found that the control group experienced a reduction in depressive symptoms from pre to posttest; however, this difference was not significant.

Mean Effect Sizes

An alpha level of 0.05 was used for all statistical tests. Number of participants, means, and standard deviations for both treatment and comparison/control groups of each study were used in the calculations (see Table 3). For each study, using the standardized mean difference formulas, a pooled standard deviation, a biased effect size, a corrected effect size, a standard error, a variance, and an inverse variance weight were calculated (see Table 4). The calculated effect sizes were all statistically independent. Figure 1 displays the distribution of effect sizes in a forest plot generated with Stata 8.2 (Stata Corp., College Station, TX). Data were analyzed using a random effects model.[1] The random variance component was 0.30. The random effects weighted mean effect size was 0.61, with a standard error of 0.30.

The 95% confidence interval for the mean effect size was 0.03 to 1.19. A z-test showed that the mean effect size for the sample of studies was statistically significant, $z = 2.05$, $p \leq 0.05$. This finding supported the hypothesis that AAA and AAT are effective at alleviating depression. According to Cohen (1988), a standardized mean difference effect size of 0.50 is considered a medium effect size, and 0.80 is considered large. The mean effect size found in this meta-analysis falls between these two means and is closer to a medium mean effect size.

Homogeneity

To test the homogeneity of the distribution of effect sizes, a Q statistic was computed and was found to be significant, $Q(4) = 13.61$, $p \leq 0.05$, therefore rejecting the null hypothesis of homogeneity and determining that the distribution of effect sizes was heterogeneous. In addition to the Q statistic, another statistical approach was used to quantify the effect of heterogeneity. I^2 was calculated to determine the percentage of total variation across studies that was due to heterogeneity rather

Table 3. Mean depression scores and standard deviations for meta-analysis studies.

	Control/Comparison			Treatment		
Study	***M***	***SD***	***n***	***M***	***SD***	***n***
Brickel (1984)	60.60	5.90	5	56.00	6.10	5
McVarish (1994)	38.40	5.71	24	24.34	12.79	24
Panzer-Koplow (2000)	7.05	7.36	19	9.33	6.57	15
Struckus (1989)	13.50	5.80	25	9.40	4.30	25
Wall (1994)	1.70	0.85	20	1.35	0.54	20

Table 4. Standardized mean difference calculations for meta-analysis studies.

Study	s_p	ES_{sm}	ES'_{sm}	SE_{sm}	v_{sm}	w_{sm}
Brickel (1984)	6.00	0.77	0.69	0.65	0.42	2.36
McVarish (1994)	9.90	1.42	1.40	0.32	0.10	9.65
Panzer-Koplow (2000)	7.03	-0.32	-0.32	0.35	0.12	8.28
Struckus (1989)	5.11	0.80	0.79	0.29	0.09	11.59
Wall (1994)	0.71	0.49	0.48	0.32	0.10	9.72

Note. s_p = pooled standard deviation of the standardized mean difference effect size; ES_{sm} = biased standardized mean difference effect size; ES'_{sm} = corrected or unbiased standardized mean difference effect size; SE_{sm} = standard error of the standardized mean difference effect size; v_{sm} = variance of the standardized mean difference effect size; w_{sm} = inverse variance weight of the standardized mean difference effect size. According to Cohen (1988), an $ES'_{sm} \leq 0.20$ is small, an $ES'_{sm} = 0.50$ is medium, and an $ES'_{sm} \geq 0.80$ is large.

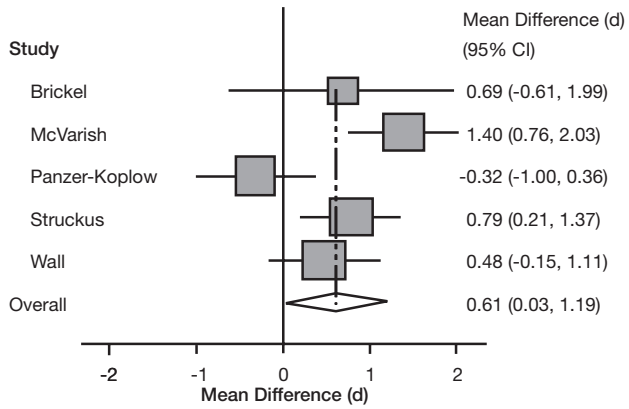

Figure 1. Forest plot of corrected standardized mean difference effect sizes for depressive symptoms following animal-assisted activities/animal-assisted therapy. Positive difference scores reflect lower depressive symptoms in the therapy group. Box areas are proportional to the sample size of each study.

than chance and was found to be 70.61% (Higgins et al. 2003). Due to the Panzer-Koplow study having a corrected or unbiased standardized mean difference effect size of –0.32, which was in the opposite direction of the other studies, statistics were recalculated excluding the data from the Panzer-Koplow (2000) study. The random variance component was 0.05. The random effects weighted mean effect size was 0.87 with a standard error of 0.21. The 95% confidence interval for

the mean effect size was 0.45 to 1.29. A *z*-test showed that the mean effect size for the sample of studies was statistically significant, $z = 4.05$, $p \leq 0.05$. A *Q* statistic was recalculated excluding the Panzer-Koplow data to examine the heterogeneity of the five studies compared with the four studies. A *Q* statistic was computed for the four studies and was not significant, $Q(3) = 4.29$, $p > 0.05$, therefore failing to reject the null hypothesis of homogeneity. Figure 2 displays the distribution of effect sizes in a forest plot generated with Stata 8.2 (Stata Corp., College Station, TX). Heterogeneity was no longer significant when excluding the Panzer-Koplow study.

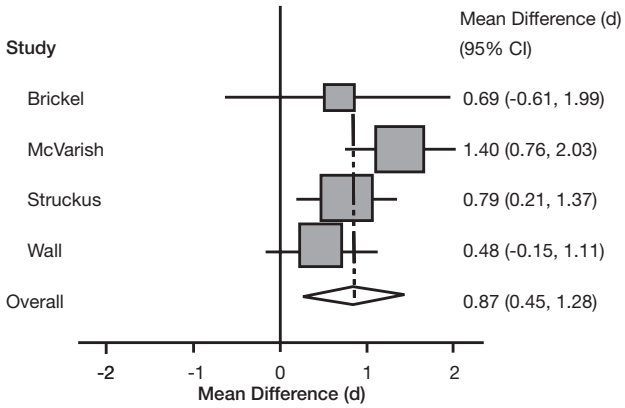

Figure 2. Forest plot of corrected standardized mean difference effect sizes for depressive symptoms following animal-assisted activities/animal-assisted therapy with the Panzer-Koplow (2000) study excluded. Positive difference scores reflect lower depressive symptoms in the therapy group. Box areas are proportional to the sample size of each study.

Fail-Safe N Calculations

To determine the influence of unpublished studies on this meta-analysis (a.k.a., the file drawer problem), the fail-safe *N* was calculated. The fail-safe *N*, a statistic developed by Rosenthal and later adapted to the standardized mean difference effect size by Orwin, estimates the number of unpublished studies with nonsignificant findings that would decrease the cumulated effect of the studies to nonsignificant (Rosenthal 1979; Orwin 1983; Lipsey and Wilson 2001). Criterion effect size levels were set at 0.50, a medium standardized mean difference effect size according to Cohen, and 0.20, a small effect size according to Cohen (1988). For the criterion effect size level of 0.50, one unpublished study with a zero effect size would be needed. For the criterion level of 0.20, the number of unpublished studies with a zero effect would be ten.

Discussion

The results of our meta-analysis support the effectiveness of AAA, and in one case, AAT, as treatments for depression. Although the number of chosen studies was small, taken together, the results of these studies indicate that exposure to AAA/AAT produces significant improvement in depression, as measured with a range of well-accepted instruments. Furthermore, these improvements are unlikely to have resulted from artifacts in study design, as we included only studies that met certain design standards, most notably the inclusion of a control group. By combining the results across a number of studies, our meta-analysis makes a new contribution to what is known about the therapeutic effectiveness of AAA/AAT.

The present study also establishes what we can expect in terms of how much improvement patients will experience as a result of AAA/AAT. By standardizing across the results of a number of studies, we can place the likely effect size of AAA/AAT in the medium range, which Cohen (1988)

characterizes as a difference between groups that is not large enough to be grossly obvious, but is large enough to be noticeable to the casual observer. In other words, although patients undergoing AAA/AAT are unlikely to experience a dramatic decrease in depression, they will likely experience a noticeable degree of relief. By considering effect sizes, our study provides evidence to endorse not only the statistical significance of animal therapies, but their practical significance, as well.

Although the distribution of the five effect sizes was found to be heterogeneous, it is likely that the heterogeneity was due to extraneous variables in the Panzer-Koplow (2000) study. In the Panzer-Koplow study, participants in the control group showed a greater reduction in depressive symptoms than the animal-assisted therapy group. There are possible explanations for this finding. Participants reported that the double negative wording on the BDI-II was confusing. Toward the end of the intervention, a caged bird and dog were brought to the facility. It is possible that these extraneous variables could have affected the outcomes in this study, especially since the bird and dog were introduced to the facility a few weeks before the post-tests were administered. These issues are unique to the Panzer-Koplow study. When excluding the Panzer-Koplow study and recalculating the heterogeneity, the studies appear more homogenous.

A further contribution of our meta-analysis is to highlight the design shortcomings associated with studies of AAA/AAT. Although we considered a large number of studies for inclusion in the analysis, only a very small proportion met our (admittedly high) standards of research design. Many of the studies we considered failed to randomly assign participants to control and treatment groups. Some also designed pretest/posttest studies in which participants were exposed to AAA before the pretest, thereby obscuring whether or not the animal intervention was the factor leading to a reduction in depression.

Our systematic survey of the literature revealed several crucial gaps. Although we chose to focus on depression as an outcome measure, we believe that it would be equally important to research the effect of AAA/AAT on physiological measures (e.g., heart rate, blood pressure), yet few studies incorporate such measures. Blood pressure would be particularly important to measure because it is a health risk marker (e.g., high blood pressure is considered a major risk factor for cardiovascular disease; NIH n.d.). Of the 165 studies we reviewed, only nine incorporated physiological measures. Of these nine studies, five did not measure depression, one was a case history, two did not use random assignment, and one study's author could not be contacted to gain further statistical information. A number of pet ownership versus non-pet ownership studies assess physiological measures; however, this is much less common among studies examining AAA/AAT.

Another gap in the literature is the lack of research specifically addressing the degree to which the positive effects of AAA/AAT are attributable to contact with the human being facilitating the animal visit. Some studies do include human contact-only groups to compare with animal interaction and control groups; however, in the present meta-analysis only one study included a "person visit only" group to be compared with the AAA group (Wall 1994), precluding us from investigating this factor in a systematic way.

In most of the studies included in the meta-analysis, either the investigator or the volunteer animal handler facilitated the intervention and collected the data, with the exception of one study (Struckus 1989) in which the intervention and data collection was facilitated by individuals naïve to the research component of the study. Therefore, experimenter bias is a possible flaw of the studies and could have inflated the effect sizes. However, because the data used in the meta-analysis were based on self-report measures, the likelihood of experimenter bias is somewhat lessened. A further limitation of our study was that the therapies included only dogs, with the sole exception of one combined dog/cat study. It is therefore impossible to know how much of the observed improvement resulted from animal interaction in general, versus interaction with a dog and/or interaction with the human facilitating treatment.

It would also be valuable for research to assess the long-term effects of AAA/AAT. This would shed light on whether the beneficial effects continue after exposure to AAA/AAT and on the duration of the effects. Lastly, research comparing individual versus group interactions is lacking, despite

the fact that the effectiveness of individual versus group animal visits is an important practical question for therapists who are planning animal interaction programs.

Technically, our studies comprised a heterogeneous array of treatments, but shared the important common thread of animal interaction. Only one study in the meta-analysis used AAT, precluding us from directly comparing the effectiveness of AAT versus the more general category AAA. For purposes of the meta-analysis, the benefits of including another study on animal interaction and depression outweighed the disadvantage of having two separate approaches represented in the same analysis. Further research is needed to contrast AAA versus AAT, but in the absence of such research, we believe it is appropriate to say that our results reflect the effectiveness of AAA/AAT in general, especially given that, as noted in the introduction, the two treatments are often not clearly distinguished in actual practice. Thus we have extended our results to both types (AAA/AAT).

Conclusions

The results of our meta-analysis offer some empirical support for the therapeutic effectiveness of dog-assisted AAA/AAT for treating depression. Five empirical studies show that AAA/AAT has positive effects on depression that are both statistically significant and large enough to be of practical significance. However, our survey of the literature suggests that only a small proportion of the existing research on AAA/AAT meets even minimal standards of research design, and few focus on important physiological measures such as blood pressure. Well-designed studies specifically focusing on physiological measures such as blood pressure would be a useful addition to the existing body of knowledge on AAA/AAT.

Acknowledgements

This article is based on a master's thesis written by M.S. We thank Steven Barger and Heidi Wayment for their helpful guidance on the design and analyses.

Notes

1. Analyses using a fixed effects model produced similar findings and did not affect conclusions. The authors would like to thank an anonymous reviewer for raising the issue.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Barba, B. E. 1995. The positive influence of animals: Animal-assisted therapy in acute care. *Clinical Nurse Specialist* 9(4): 199–202.

Batson, K., McCabe, B., Baun, M. M. and Wilson, C. 1998. The effect of a therapy dog on socialization and physiological indicators of stress in persons diagnosed with Alzheimer's disease. In *Companion Animals in Human Health*, 203–215, ed. C. C. Wilson and D. C. Turner. Thousand Oaks, CA: Sage Publications.

Beck, A. T., Steer, R. A. and Brown, G. K. 1996. *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. and Erbaugh, J. 1961. An inventory for measuring depression. *Archives of General Psychiatry* 4: 53–63.

Boldt, M. A. and Dellmann-Jenkins, M. 1992. The impact of companion animals in later life and considerations for practice. *The Journal of Applied Gerontology* 11(2): 228–239.

Brasic, J. R. 1998. Pets and health. *Psychological Reports* 83: 1011–1024.

*Brickel, C. M. 1984. Depression in the nursing home: A pilot study using pet-facilitated psychotherapy. In *The Pet Connection: Its Influence on Our Health and Quality of Life*, 407–415, ed. R. K. Anderson, B. L. Hart, and L. A. Hart. Minneapolis, MN: Center to Study Human–Animal Relationships and Environments, University of Minnesota.

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P. H., Adey, M. and Rose, T. L. 1982. Screening tests for geriatric depression. *Clinical Gerontologist* 1(1): 37–43.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences.* 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Connor, K. and Miller, J. 2000. Animal-assisted therapy: An in-depth look. *Dimensions of Critical Care Nursing* 19(3): 20–26.

Crowley-Robinson, P., Fenwich, D. C. and Blackshaw, J. K. 1996. A long-term study of elderly people in nursing homes with visiting and resident dogs. *Applied Animal Behaviour Science* 47: 137–148.

Delta Society. n.d. About animal-assisted activities and animal-assisted therapy. <http://www.deltasociety.org/aboutaaat.htm> Accessed April 7, 2003.

Fick, K. M. 1992. The influence of an animal on social interactions of nursing home residents in a group setting. *The American Journal of Occupational Therapy* 47: 529–534.

Francis, G., Turner, J. T. and Johnson, S. B. 1985. Domestic animal visitation as therapy with adult home residents. *International Journal of Nursing Studies* 22(3): 201–206.

Hagman, C. S. 1997. The effect of animal-assisted therapy on levels of measured anxiety and depression in residents of assisted living and health care communities. Unpublished master's thesis. The University of Tennessee, Knoxville.

Harris, M. D., Rinehart, J. M. and Gerstman, J. 1993. Animal-assisted therapy for the homebound elderly. *Holistic Nursing Practice* 8(1): 27–37.

Hedges, L. V. 1981. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6(2): 107–128.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J. and Altman, D. G. 2003. Measuring inconsistency in meta-analyses. *British Medical Journal* 327: 557–560.

Jendro, C., Watson, C. and Quigley, J. 1984. The effects of pets on the chronically-ill elderly. In *The Pet Connection: Its Influence on Our Health and Quality of Life,* 416–422, ed. R. K. Anderson, B. L. Hart, and L. A. Hart. Minneapolis, MN: Center to Study Human–Animal Relationships and Environments, University of Minnesota.

Jessen, J., Cardiello, F. and Baun, M. 1996. Avian companionship in alleviation of depression, loneliness, and low morale of older adults in skilled rehabilitation units. *Psychological Reports* 78: 339–348.

Lipsey, M. W. and Wilson, D. B. 2001. *Practical Meta-Analysis.* Thousand Oaks, CA: Sage.

McCulloch, M. J. 1983. Animal-facilitated therapy: Overview and future direction. In *New Perspectives on Our Lives With Companion Animals,* 410–426, ed. A. H. Katcher and A. M. Beck. Philadelphia, PA: University of Pennsylvania Press.

*McVarish, C. 1994. The effects of pet facilitated therapy on depressed institutionalized inpatients. Doctoral dissertation. California School of Professional Psychology, Alameda.

Murray, C. J. L. and Lopez, A. D. eds. 1996. *The Global Burden of Disease and Injury Series, Volume 1: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020.* Cambridge, MA: Harvard University Press.

National Institutes of Health: National Heart, Lung, and Blood Institute. n.d. Risk factors and heart disease. <http://nhlbisupport.com/chdl/chdexp1.htm> Accessed April 7, 2004.

National Institutes of Health. 1992. Diagnosis and treatment of depression in late life: NIH consensus development panel on depression in late life. *Journal of the American Medical Association* 268(8): 1018–1024.

National Institute of Mental Health. n.d. Depression research at the National Institute of Mental Health: Fact Sheet. <http://www.nimh.nih.gov/publicat/depresfact.cfm> Accessed February 11, 2004.

Norris, J. M. 1983. the effects of pet dogs and their owners on geriatric day care center participants' affect, social interaction, physical functioning, and activity level. Unpublished master's thesis. Loyola College, Baltimore.

Orwin, R. G. 1983. A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics* 8(2): 157–159.

*Panzer-Koplow, S. 2000. Effects of animal-assisted therapy on depression and morale among nursing home residents. Doctoral dissertation. Rutgers, The State University of New Jersey.

Perelle, I. B. and Granville, D. A. 1993. Assessment of the effectiveness of a pet facilitated therapy program in a nursing home setting. *Society & Animals* 1(1): 91–100.

Pratt, L. A., Ford, D. E., Crum, R. M., Armenian, H. K., Gallo, J. J. and Eaton, W. W. 1996. Coronary heart disease/myocardial infarction: Depression, psychotropic medication, and risk of myocardial infarction: Prospective data from the Baltimore ECA follow-up. *Circulation* 94(12): 3123–3129.

Raskin, A. and Crook, T. 1988. Mood scales-elderly (MS-E). *Psychopharmacology Bulletin* 24(4): 727–729.

Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. *Psychological Bulletin* 86(3): 638–641.

Standards of Practice for Animal-assisted Activities and Animal-assisted Therapy. 1996. Renton, WA: Delta Society.

*Struckus, J. E. 1989. The use of pet-facilitated therapy in the treatment of depression in the elderly: A behavioral conceptualization of treatment effects. Doctoral dissertation. University of Massachusetts.

Vaughan, B. K. 1990. Pet and patient: Enhancing nursing home life. Unpublished master's thesis. California State University, Long Beach.

*Wall, M. J. 1994. The effects of companion animal visitation on mood state and level of speech activity of nursing home residents. Doctoral dissertation. California School of Professional Psychology, San Diego.

Winkler, A., Fairnie, H., Gericevich, F. and Long, M. 1989. The impact of a resident dog on an institution for the elderly: Effects on perceptions and social interactions. *The Gerontologist* 29: 216–223.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M. and Leirer, V. O. 1983. Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research* 17(1): 37–49.

Zisselman, M. H., Rovner, B. W., Shmuely, Y. and Ferrie, P. 1996. A pet therapy intervention with geriatric psychiatry inpatients. *The American Journal of Occupational Therapy* 50(1): 47–51.

Zung, W. W. K. 1965. A self-rating depression scale. *Archives of General Psychiatry* 12: 63–70.

Zung, W. W. K. 1974. The measurement of affects: Depression and anxiety. In *Modern Problems of Pharmacopsychiatry: Vol. 7. Psychological Measurements in Psychopharmacology,* 170–188, ed. P. Pichot. Basel: S. Karger.