A Systematic Review of Effectiveness of Complementary and Adjunct Therapies and Interventions Involving Equines

Alison Selby
Child and Family Guidance Centers, Plano, Texas

Alexa Smith-Osborne
University of Texas at Arlington

Objective: This systematic review examines the empirical literature in an emerging body of evidence for the effectiveness of biopsychosocial interventions involving equines across populations with chronic illness or health challenges. Method: Selected quantitative studies published in peer-reviewed journals were reviewed for inclusion; the gray literature and white papers were also explored. Population, Intervention, Comparison, and Outcome (PICO) criteria and Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) were applied to all studies. Fourteen full reports meeting a priori inclusion criteria were extracted from 103 studies accessed through 16 electronic databases and a hand search. Data were synthesized in relation to three research questions informing evidence-based practice. Results: No randomized clinical trials were located. Two studies provided a moderate level of evidence for effectiveness. Nine studies demonstrated statistically significant positive effects. Three studies did not find significant psychosocial effects for the target group, although one found significant positive effects for the comparison group. Conclusion: In the aggregate, the evidence is promising in support of the effectiveness of complementary and adjunct interventions employing equines in the treatment of health challenges. Future studies are needed that utilize rigorous and creative designs, especially longitudinal studies and comparisons with established effective treatments.

Keywords: chronic illness, complementary and adjunct therapies, animal-assisted interventions, therapeutic horseback riding, equine-facilitated psychotherapy

Therapeutic interventions involving horses for people with various disabilities, illnesses, and psychosocial challenges are not recent phenomena: history is rich with accounts of the curative effects associated with interaction with equines. Bustad and Hines’s (1984) review cited historical examples of the use of riding therapy, from treating various illnesses in the ancient world as early as 5 B.C. to rehabilitating injured British World War I veterans. However, significant scholarly attention was not directed toward animal-assisted therapeutic interventions in general until the 1960s (Levinson, 1962), and only relatively recently has there been an interest in this adjunct therapy.

In the past 10 years, building on an accumulating body of anecdotal and other evidence, interest in this adjunct therapy has increased. Concomitantly, the development of safety standards and their maintenance through formal accreditation procedures have minimized the risks inherent in this type of activity (Chitnavis, Gibbons, Hirigoyen, Lloyd-Parry, & Simpson, 1996).

Equines as nonpredator animals provide unique opportunities in the therapeutic process when compared to predator companion animals. These therapeutic values have been associated with the human–equine bond and the multiplicity of potential roles equines can play in such a relationship, potentially supporting therapeutic gains that span and blend exercise, recreational, competence-building, and other psychosocial domains. Some of the attributes that horses bring to the therapeutic environment are cooperation, patience, willingness, receptiveness, and, after millennia of domestication, an orientation toward people (Ewing, MacDonald, Taylor, & Bowers, 2007; Hayden, 2005; Karol, 2007; Vidrine, Owen-Smith, & Faulkner, 2002).

Therapeutic horsemanship involves any of a number of active physiotherapeutic exercises on and around the horse. To date, research on equine-assisted activities (EAAs) and therapies has concentrated on the physical benefits of these types of alternative treatments. In particular, most of the research attention has focused on the effects of hippotherapy (utilizing equine movement in physical, occupational, or speech therapy treatment), primarily for the rehabilitative effects on physical limitations associated with physical mobility challenges (Ungermann & Gras, 2011; Zadnikar & Kastrin, 2011).

Other therapies that have received less research attention include therapeutic horseback riding (THR), therapeutic carriage-driving (TCD), equine-facilitated experiential learning (EFEL), interactive vaulting, and equine-facilitated psychotherapy (EFP). THR, TCD, and vaulting use group or individual riding or driving experiences targeted to biopsychosocial effects related to the health condition or disability (Professional Association of Therapeutic Horsemanship International, n.d.). EFEL and EFP models use mounted and unmounted activities in a team facilitator or individual psychotherapy approach based on the philosophy that people do not change unless they are uncomfortable (Kersten &
Thomas, 2004; Kohanov, 2001). Like horses, people often remain in painful situations and “run or hide from pressure” (Russell-Martin, 2006, p. 31). In both EFEL and EFP models, exercises involving the horse generate experiences in the “here and now” (Frame, 2006) that can be observed by both the facilitator or therapist and the client, and they provide opportunities to experiment with cognitive, affective, behavioral, and sensory modulation and regulation. These five EAAs are being used as complementary and adjunct therapies (CATs) to intervene in a broader range of health conditions and disabilities, and in support of a broader range of therapies, than hippotherapy.

Until recently, most of the supporting evidence for the beneficial mobility rehabilitative effects of activities involving horses consisted of single-subject designs that lacked comparison groups, studies that lacked standardized measures, and results that were frequently reported outside peer-reviewed publications (Silkwood-Sherer & Warnbiert, 2007; Snider, Korner-Bitensky, Kammann, Warner, & Saleh, 2007). However, more rigorous evidence has begun to accumulate, as illustrated by two systematic reviews (Snider et al., 2007; Sterba, 2007), which demonstrated clinically significant benefits of hippotherapy for children with cerebral palsy. Recent research (Shurtleff, Standeven, & Engberg, 2009) has showed large effect sizes for dynamic trunk and head stability and functional reach improvement in children with spastic diplegia cerebral palsy, through 12 weeks postintervention.

Because of the unique nature of this treatment approach, health personnel with widely differing philosophies and from diverse disciplines and theoretical backgrounds find that the incorporation of horses into clinical practice can have beneficial effects. It should be noted, however, that although interventions involving horses may be therapeutic, the mere presence of equines in the therapy session does not fit the clinical definition of therapy (Fredrickson, 1992; see also Taylor, 2001). In the therapeutic setting, horses are engaged as change agents to facilitate the process of enhanced biopsychosocial development, growth, and education. Clarification of the elements of a therapeutic relationship, as opposed to the benefits of recreational activities, is critical if this approach is to achieve credible status in the professional community (Beck & Katcher, 2003).

Interventions employing equines have been applied across age groups and cultures (Dell, Chalmers, Dell, Sauve, & MacKinnon, 2008). It has been applied in work with families (Kersten, 1997; Thomas, 2002) and groups (Trotter, Chandler, Goodwin-Bond, & Casey, 2008; Vidrine et al., 2002), and is not gender-specific. This treatment approach has been used to address terminal illness (Haylock & Cantril, 2006), comorbid chronic developmental disabilities and health conditions (Ewing et al., 2007; Greenland, 2001; Iannone, 2003), behavioral and attentional issues (Bassil & Antoon, 1996; Zanin, 1997; Bowers, 2001; Gamache, 2004; Tetreault, 2006; Beckman-Devik & Ansin, 2008), substance abuse and addiction disorders (Dell et al., 2008), eating disorders (Christian, 2005; Colclasure, 2004; Cornelius, 2002; Cumella, 2003; Lutter & Smith-Osborne, 2011), depression (Bray, 2002; Frame, 2006), anxiety (Moreau & McDaniel, 2000), relationship problems (Russell-Martin, 2006), and posttraumatic stress disorder (Myers, n.d.; Yorke, 1997, 2010; Yorke, Adams, & Coady, 2008). It has been applied to populations ranging from at-risk youths (Chandler, 2005; Cole, 2005; Hayden, 2005; Kaiser, Smith, Heleski, & Spence, 2006; Sapir, 2007; MacDonald, 2004) to brain-injured and aging veterans (Myers, n.d.; Smith-Osborne & Selby, 2009) and to health care settings such as medical domiciliary care (Smith-Osborne & Selby, 2009), residential treatment facilities (All, Loving, & Crane, 1999; Lutter & Smith-Osborne, 2011), hospices and hospitals (Boysen, 1985; Myers, n.d.), and prisons (Cushing & Williams, 1995).

Typically, EAA CATs are brief and experiential in nature (Klontz, Bivens, Leinart, & Klontz, 2007). Theoretical and clinical orientations that have incorporated equines include empowerment theory and the health belief model (Smith-Osborne & Selby, 2009), the palliative care model (Boysen, 1985), dual representation theory (Yorke, 1997, 2010; Yorke et al., 2008), cognitive remediation (Myers, n.d.; Zanin, 1997), recreation therapy (Myers, n.d.; Smith-Osborne & Selby, 2009), cognitive behavioral therapy (Eggiman, 2006; Frame, 2006), humanistic and transpersonal models (Lutter & Smith-Osborne, 2011; Myers, n.d.; Tranutt, 2003), and psychodynamic therapy (Karol, 2007). An intervention that employs equines in the treatment plan, by its very nature, may be appealing and relevant to client populations not otherwise amenable or accessible to traditional hospital or office-based forms of treatment.

The quality of the therapeutic relationship has been described as the most significant factor in successful intervention outcomes (Lambert & Bergin, 1994); horses may facilitate the establishment of this relationship. For example, the exercises employed in EFP are problem-solving tasks that occur in novel surroundings, and that enable not only the client, but also the therapist, to view situations in a different light than may be possible in traditional psychotherapeutic settings; it is based primarily on observations rather than solely on what is said. Sometimes hidden emotions emerge in this setting while the client is actively engaged in solving the problems at hand (Christian, 2005). Each activity is followed by processing time in which the client is encouraged to make connections between what occurred in the session and his or her internal state (Eggiman, 2006).

Research Aims

Despite a call for research dating back at least to the 1970s (Mayberry, 1978), prior reviews have identified a gap in research on the effects of horsemanship as an adjunct and complementary therapy for chronic health conditions and disabilities (other than the mobility effects of hippotherapy for disabilities, primarily cerebral palsy) that are considered methodologically sound, and examine associated psychosocial improvements (Smith-Osborne & Selby, 2010). Because reviews are now in place for hippotherapy for the mobility effects (Snider et al., 2007; Sterba, 2007), the purpose of this systematic review was to synthesize the evidence for the more generalized biopsychosocial effects resulting from involving equines in therapeutic processes, considered as an alternative and complementary intervention for chronic health conditions and disabilities (Bausell & Berman, 2002; Druss & Rosenheck, 1999). Our aim was to determine whether the research supports recommendations for EAAs to be integrated into the course of biopsychosocial interventions.
Specifically, the following questions were examined:

1. Does involving equines in the intervention process result in health benefits beyond increased mobility for participants?

2. What are the quality and strength of the research that has examined the above question?

**Method**

Methodology for this systematic review follows the current analytical methods and standards established by the PRISMA Group for systematic reviews and meta-analyses (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) and delineated in the items of the AMSTAR tool (Oxman, Schünemann, & Fretheim, 2006). Study quality was measured using Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) criteria developed by the GRADE Working Group in association with the World Health Organization (GRADE Working Group, 2004; Oxman et al., 2006). Research synthesis is treated as a separate research method, distinct from qualitative, quantitative, and mixed methods used in single studies, because the sample does not consist of persons, groups, organizations, or communities, but studies themselves (Cooper & Hedges, 1994).

**Literature Search and Retrieval Process**

Approval for the study was obtained from the institutional review board of The University of Texas at Arlington. Studies used in this review were obtained from electronic searches of the following databases from 2000 through 2009 (inclusion dates were chosen to reflect the timeframe characterizing the reviews in the hippotherapy domain, reflecting our intent to synthesize nonhippotherapy EAAs from 2000 to 2009): Academic Search Complete, Alt HealthWatch, CINAHL Plus with Full Text, EBSCO Animals, E-Journals, ERIC, Health Source: Nursing/Academic Edition, MasterFILE Premier, MEDLINE, Professional Development Collection, PsycARTICLES, Psychology and Behavioral Sciences Collection, PsycINFO, PubMed, Social Work Abstracts, and TOPICSearch. Keywords entered were: equine and psychosocial, therapeutic riding, therapeutic horsemanship, horse and therapy, horse and psychotherapy, equine and psychotherapy, equine and therapy, equine facilitated therapy, and equine assisted therapy. In addition, a hand search of the journal Anthrozoös, the leading professional journal for publication of evidence-based practice of equine-facilitated interventions, was performed for years between 2000 and 2004 not available electronically.

Current methodology guidelines posit that the exclusion of the gray literature puts not only meta-analyses, but also systematic reviews at higher risk for publication bias and therefore overestimation of effects (Wilson, 2009). The assumption that peer-reviewed published studies constitute the high-quality evidence in a field has been brought into question (Torgerson, 2006). Thus, the inclusion of only peer-reviewed publications in systematic review syntheses, as well as meta-analyses, has been discouraged by the PRISMA group (Moher et al., 2009), and other scholars concerned about the quality of this research synthesis methodology (Littell, 2008; Moher, Tetzlaff, Tricco, Sampson, & Altman, 2007). Now, the inclusion of only peer-reviewed publications is considered a source of bias, not quality improvement, in this type of research methodology, and systematic review procedures (which are the starting point and foundation of any meta-analysis) are recommended to be consistent with those of meta-analysis. Therefore, the gray literature was also searched for relevant papers, with no limits on publication date, using Dissertation Abstracts International and ProQuest Dissertations and Theses, as were white papers pertaining to the broad subject topic. The accumulated articles were scanned for references through March 2009. Finally, requests for relevant papers were made to professionals involved in the practice of equine-facilitated interventions.

**Inclusion criteria and study quality assessment.** Eligibility of studies was assessed using Population, Intervention, Comparison, and Outcome (PICO) criteria, as recommended by Gambrill (2006). The answers to PICO questions define the population under study, the specified intervention, the comparison or lack thereof, and desired outcomes, all with an element of time; for example, How old are the participants? How long is the treatment protocol? When is the outcome measured?

In keeping with the research aims to address a gap in the EAA literature, inclusion criteria for this review included studies that quantified the biopsychosocial benefits of nonhippotherapy EAs across age groups. Review criteria excluded articles not in English, except those that could be translated using computer translation software. One was found (Boysen, 1985), but it did not meet inclusion criteria for this review. Although the European Union has a long tradition of hippotherapy, EAAs that target broader biopsychosocial benefits of these CATs have emerged synchronously with those in the United States (Broesma, 2009). In addition, the European environmental context (e.g., policies, national health care systems, local resources, provider regulation patterns) for such programs is quite different, suggesting the need for a separate follow-up comparison review using enhanced translator resources.

Articles pertaining to the mobility effects of hippotherapy were excluded from this review, because recent, readily available reviews have already been done of hippotherapy in this area (Snider et al., 2007; Sterba, 2007). Reports that could not be read in full were excluded. Papers that were published in professional membership journals were excluded because of the possibility of lack of objective peer review resulting in publication bias. Anecdotal reports and case studies were excluded in an effort to isolate the quantitative research that has been done to date. Qualitative studies were excluded in order to apply current review standards fully and to permit quantitative synthesis if the level of the located studies were appropriate. The two researchers, working independently, reviewed the abstracts of papers retrieved through the search processes, determined whether inclusion criteria were sufficiently met to indicate retrieval of the full article, and compared their results. Differences were discussed and recorded until consensus was reached. Full-text articles were retrieved for those remaining abstracts, and the same independent review process followed with those. From the initial search, 103 articles were retrieved. Exclusion of commentaries and articles not available in English yielded 40 articles for consideration for inclusion in this systematic review. Five studies could not be retrieved in full for review. Fourteen studies that met inclusion criteria were identified from the literature. See Figure 1 for a summary of the retrieval process.
Studies that met inclusion criteria were observational, quasi-experimental, or descriptive pre–post designs, except for one experimental study of efficacy. Studies were evaluated further for quality using GRADE criteria developed by the GRADE Working Group in association with the World Health Organization (GRADE Working Group, 2004; Oxman et al., 2006), adapted for use in this review. Quality GRADES may be found in Table 1.

Data Extraction and Synthesis

The first abstractor entered all relevant information into the prepared evidence tables. The second reviewer subsequently checked each abstraction for accuracy and completeness against the original articles. Abstractors reconciled all disagreements concerning the information reported in the evidence tables. Finally, evidence-based and -informed practice levels of evidence were considered during the data synthesis phase of this review, as recommended by Guyatt and Rennie (2002) and Gambrill (2006). Only in the past few years has methodologically rigorous evidence begun to emerge in the field of equine-facilitated and -assisted psychotherapy, and to date no randomized clinical trials, considered the criterion standard in research methodology, have been conducted. Some studies using nonclinical samples met inclusion criteria due to the overlap of intervention types and domains in
Results of Articles Reviewed: GRADE Criteria

<table>
<thead>
<tr>
<th>Study</th>
<th>Study design</th>
<th>Quality of evidence</th>
<th>Directnessa</th>
<th>Higher if strong associationb</th>
<th>GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowers &amp; MacDonald (2001)</td>
<td>Pilot</td>
<td>Low</td>
<td>b = −2</td>
<td>−</td>
<td>Very low</td>
</tr>
<tr>
<td>Kaiser, Spence, Lavergne, &amp; Vanden Bosch (2004)</td>
<td>Pilot</td>
<td>Low</td>
<td>a = −1</td>
<td>d = +1</td>
<td>Low</td>
</tr>
<tr>
<td>Graham (2007)</td>
<td>Pilot</td>
<td>Low</td>
<td>a = −1</td>
<td>d = +1</td>
<td>Moderate</td>
</tr>
<tr>
<td>Schultz, Remick-Barlow, &amp; Robbins (2007)</td>
<td>Pilot</td>
<td>Low</td>
<td>a = −1</td>
<td>d = +1</td>
<td>Low</td>
</tr>
<tr>
<td>Tetreault (2006)</td>
<td>Pre-/posttest comparison</td>
<td>Low</td>
<td>b = −2</td>
<td>−</td>
<td>Very low</td>
</tr>
<tr>
<td>Shultz (2005)</td>
<td>Pre-/posttest comparison</td>
<td>Low</td>
<td>a = −1</td>
<td>−</td>
<td>Very low</td>
</tr>
<tr>
<td>Ewing, MacDonald, Taylor, &amp; Bowers (2007)</td>
<td>Pre-/posttest comparison</td>
<td>Low</td>
<td>a = −1</td>
<td>d = +1</td>
<td>Low</td>
</tr>
<tr>
<td>Kaiser, Smith, Heleski, &amp; Spence (2006)</td>
<td>Pre-/posttest comparison</td>
<td>Low</td>
<td>b = −2</td>
<td>d = +1</td>
<td>Low</td>
</tr>
<tr>
<td>Russell-Martin (2006)</td>
<td>Pre-/posttest comparison</td>
<td>Low</td>
<td>d = −1</td>
<td>d = +1</td>
<td>Low</td>
</tr>
<tr>
<td>Klontz, Bivens, Leinart, &amp; Klontz (2007)</td>
<td>Pre-/posttest with follow</td>
<td>Low</td>
<td>c = −1</td>
<td>d = +1</td>
<td>Low</td>
</tr>
<tr>
<td>Shambo, Seely, &amp; Vonderfecht (2008)</td>
<td>Pre-/posttest with follow</td>
<td>Low</td>
<td>b = −2</td>
<td>d = +1</td>
<td>Very low</td>
</tr>
<tr>
<td>Greenwald (2001)</td>
<td>Program evaluation</td>
<td>Low</td>
<td>a = −1</td>
<td>d = +1</td>
<td>Low</td>
</tr>
<tr>
<td>Iannone (2003)</td>
<td>Program evaluation</td>
<td>Low</td>
<td>a = −1</td>
<td>d = +1</td>
<td>Low</td>
</tr>
<tr>
<td>Trotter, Chandler, Goodwin-Bond, &amp; Casey (2008)</td>
<td>Efficacy pre-/posttest</td>
<td>Low</td>
<td>b = −2</td>
<td>d = +1</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

Note. GRADE = Grading of Recommendations, Assessment, Development, and Evaluation.

A = 1 (some uncertainty); b = −2 (major uncertainty); c = −1 (sparse data); d = −1 (high probability of reporting bias).

a = +1 (strong, no plausible confounders, consistent and direct evidence; p < .05; >2 observational studies); b = +2 (very strong, no major threats to validity and direct evidence; p < .02); c = +1 (evidence of a dose gradient); d = +1 (all possible confounds would have reduced effect).

Biopsychosocial Effects of Equine-Assisted Interventions

Pilot studies. Of the four pilot studies investigating THR and EFP, two (Bowers & MacDonald, 2001; Graham, 2007) included measurements of levels of depression among other effects, and the remaining two measured anger, quality of life, and perceived self-confidence (Kaiser et al., 2004), and global assessment of functioning (Schultz et al., 2007). One THR study (Bowers & MacDonald, 2001) noted low statistical power and found no statistically significant effects on any measures, except depression, for which they reported a statistically significant decrease, strengthened by consistent qualitative findings (participant self-report). The other (Kaiser et al., 2004) reported that total scores on anger decreased significantly on all subscales except frustration, but noted no significant differences in quality of life or perceived self-confidence; they did, however, note a positive trend in the Global Self Worth subscale.

Of the two EFP pilot studies, Graham’s (2007) study of depression resulting from catastrophic loss among adults showed a statistically significant increase (p = .001) in perceived wellness, although physical wellness did not appear to increase. Graham also reported a statistically significant decrease in levels of depression (p = .001); Graham’s hypothesis that posttest scores of depression and wellness would be inversely correlated was supported, adding strength to the study, but the hypothesis that physiological measures would decrease in an optimal direction was not supported statistically. All participants rated the sessions positively in self-report. The other EFP pilot (Schultz et al., 2007) studied children and teens with a range of symptoms, and reported a statistically significant increase in the percentage of improvement in global functioning in relation to the number of sessions of EFP (p =
### Table 2

**Summary of Pilot Studies**

<table>
<thead>
<tr>
<th>Study design</th>
<th>Sample</th>
<th>Hypotheses</th>
<th>Intervention/comparison</th>
<th>Measures</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowers &amp; MacDonald (2001): Observational pre-/posttest</td>
<td>$N = 10$. At-risk adolescents, ages 14–18 years. Boys, girls, various ethnicities, abused; psychotherapy.</td>
<td>Sense of self-worth, empathy, internal LOC will ↑; depression, loneliness ↓.</td>
<td>Weekly for 7 weeks for 90 min natural horsemanship approach: round pen and riding. a Description of treatment/comparison groups unclear.</td>
<td>Self-Perception Profile for Adolescents, Empathy Questionnaire, Locus of Control Scale, Children's Depression Inventory, Children's Loneliness, Pre-/posttest: a. Children's Inventory of Anger; b. Pediatric Quality of Life Inventory, Version 4.0; c. Self-Perception Profile for Children.</td>
<td>Depression sig ↓</td>
</tr>
<tr>
<td>Schultz, Remick-Barlow, &amp; Robbins (2007): Observational pre-/posttest</td>
<td>$N = 63$, boys = 37, girls = 26, ages 4–16 years, multi-ethnic, various behavioral conditions.</td>
<td>GAF–Children ↑.</td>
<td>EAGALA model: 6–116 sessions of EAP by licensed social worker and equine specialist.</td>
<td>GAF–Children, at 3 months. Intervals until treatment complete.</td>
<td>GAF–Children ↑; greatest ↑ in youngest, history of abuse, female.</td>
</tr>
</tbody>
</table>

**Note.** LOC = locus of control; ↑ = increase; ↓ = decrease; H1 = Hypothesis 1; H2 = Hypothesis 2; H3 = Hypothesis 3; H4 = Hypothesis 4; GAF–Children = Global Assessment of Functioning–Children Scale; NARHA = North American Riding for the Handicapped Association; EAGALA = Equine Assisted Growth and Learning Association; EAP = equine-assisted psychotherapy; sig = significant.

a Confounder/limitation. b Gray literature study.
<table>
<thead>
<tr>
<th>Study design</th>
<th>Sample</th>
<th>Hypotheses</th>
<th>Intervention/comparison</th>
<th>Measures</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>aTetreault (2006): Observational one-group pre-/posttest</td>
<td>N = 10, ages 10–12 years; boys = 4, girls = 6, diagnosed with emotional disturbance. All taking medications.</td>
<td>H1: Self-identification, behavior management; H2: Use skills to interact with others.</td>
<td>Five sessions at 2 hr × 5 weeks. EAGALA model.</td>
<td>Self-created survey</td>
<td>Claims statistical significance.</td>
</tr>
<tr>
<td>aShultz (2005): Observational pre-/posttest comparison</td>
<td>N = 29; ages 12–18 years, in residential care or outpatient treatment; ethnic diversity. Treatment group n = 15, boys = 8, girls = 7, comparison group n = 14, boys = 9, girls = 5.</td>
<td>H1: psychosocial functioning †</td>
<td>Group or individual EAP conducted by therapy teams (therapists and horse specialists). EAGALA model.</td>
<td>Youth Outcome Questionnaire Self-Report/Parent Report/Caregiver Report</td>
<td>Reports clinical significance.</td>
</tr>
<tr>
<td>Ewing, MacDonald, Taylor, &amp; Bowers (2007): Observational quantitative and qualitative pre-/posttest comparison</td>
<td>N = 28; youths from alternative school with SED, ages 10–13 years. Boys, girls, varying ethnicity, low SES. Teacher-referred. 4–5 youths per 9 week session, groups selected for similar age and IQ. Wait list comparison.</td>
<td>H1: sense of self-worth, self-esteem ↑; H2: interpersonal empathy; H3: ↓ internal LOC; H4: depression ↓, loneliness ↓.</td>
<td>Two hours, twice weekly for 9 weeks = 36 hr + classroom. Some riding. Participation in all aspects of care and handling (except turn-out in pen). Supervised by teachers and *sometimes school psychologist, with volunteers.</td>
<td>Self-Perception Profile for Children; Empathy Questionnaire; Nowicki-Strickland Internal-External Control Scale for Children; Children’s Depression Inventory; Children’s Loneliness Questionnaire.</td>
<td>Qualitative observational data from teacher, instructor, and volunteers did support positive effects.</td>
</tr>
<tr>
<td>aRussell-Martin (2006): Observational pre-/posttest comparison</td>
<td>N = 20 dyads, ages 21–45 years. Treatment group: n = 10 dyads receiving EFT. Comparison group: n = 10 dyads receiving SFT; random assignment to groups; groups similar in descriptive measure.</td>
<td>H1: relational adjustment compared to SFT ↑.</td>
<td>All received six 1-hr treatment sessions over 6 weeks. EAGALA model.</td>
<td>DAS (self-report) both groups after Sessions 1, 3, and 6, experience, as well as demographic data.</td>
<td>EFT group ↑ relational adjustment.</td>
</tr>
</tbody>
</table>

Note. SED = severe emotional disturbance; SES = socioeconomic status; NA = Not Available; EFT = equine facilitated therapy; SFT = solution-focused therapy; H1 = Hypothesis 1; H2 = Hypothesis 2; H3 = Hypothesis 3; LOC = locus of control; H4 = Hypothesis 4; ↑ = increase; ↓ = decrease; EAGALA = Equine Assisted Growth and Learning Association; EAP = equine assisted psychotherapy; DAS = Dyadic Adjustment Scale.

a Gray literature study.
Table 4  
**Summary of Pre- and Posttest Studies With Follow-Up**

<table>
<thead>
<tr>
<th>Study design</th>
<th>Sample</th>
<th>Hypotheses</th>
<th>Intervention/comparison</th>
<th>Measures</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klontz, Bivens, Leinart, &amp; Klontz (2007): Observational one-group pre-/posttest with follow-up</td>
<td>N = 31 adults, ages 23–70 years, 90% White. Men = 9, women = 22.</td>
<td>H1: general symptom severity ↓. H2: enhanced psychological well-being.</td>
<td>Equine-assisted experiential therapy (based on psychodrama). Manualized. Two male and 3 female psychotherapists led, assisted by equine specialist. Adherence closely monitored.</td>
<td>BSI; BSI Summary scale; GSI; Personal Orientation Inventory; pre-/posttest, 6-month follow-up.</td>
<td>H1: ↓ GSI scores from pre- to posttest; not from posttest to follow-up; H2: ↑ from pre-to posttest; not from posttest to follow-up.</td>
</tr>
<tr>
<td>Shambo, Seely, &amp; Vonderfecht (2008): Observational one-group pre-/posttest with follow-up</td>
<td>N = 6 female with complex PTSD with previous “refusal of or failure to benefit from traditional group therapy”.</td>
<td>H1: depression, dissociative symptoms, anxiety, overall treatment effectiveness ↓.</td>
<td>Ten 2-hr sessions, once a week for 10 weeks. EFP psychoeducation + psychoeducation + group process and support. Half session in classroom, half with horses. Epona model: nonriding, insight-oriented.</td>
<td>Hamilton Depression Scale, pre-/posttest 4-month follow-up, BAI, mid- to posttest 4-month follow-up, DES, pre-to mid- to posttest 4-month follow-up, OQ, pre-to mid- to posttest 6-month follow-up.</td>
<td>H1: ↓ in depression, pre-/posttest, post follow-up. No pre BAP, but mid- to posttest follow-up showed trend. DES pretest to follow-up, OQ: pre-to posttest and, pretest to follow-up.</td>
</tr>
</tbody>
</table>

**Note.** PTSD = posttraumatic stress disorder; EFP = equine-facilitated psychotherapy; H1 = Hypothesis 1; H2 = Hypothesis 2; BSI = Brief Symptom Inventory; GSI = Global Severity Index; BAI = Beck Anxiety Inventory; DES = Dissociative Experience Scale; OQ = Outcome Questionnaire.  
* Confounder/limitation.

Table 5  
**Summary of Program Evaluations**

<table>
<thead>
<tr>
<th>Study design</th>
<th>Sample</th>
<th>Hypotheses</th>
<th>Intervention/comparison</th>
<th>Measures</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Iannone (2003): Observational program evaluation/impact assessment</em></td>
<td>N = 19, boys = 13, girls = 6, ages 12–16 years, ethnically and diagnostically diverse in residential treatment, with SED and limited/no success with previous treatment. Selected by school treatment team. Comparison: n = 8 boys = 4, treatment = 4, ages 12–15 years; wait listed.</td>
<td>H1: Self-perceptions ↑ as result of time in program. P1: self-esteem ↑ T1. ↑ T2; P2: ↑ internal LOC; H2: Psychiatric symptoms and behavior problems ↓. H3: Vocational skills ↑.</td>
<td>Two days week for 2 hr. Three phases for 6 weeks. ↑ depending on successful completion of each P1. Basics of horse care and barn management plus basic riding skills. 2. Peer mentorship of incoming P1. 3. Work-study continued training. Participants responsible for all aspects of horse care during time at site.</td>
<td>Rosenberg Self-Esteem Scale; Novicki-Strickland Locus of Control Scale; Symptom Checklist–90; Behavior Rating Index for Children; knowledge test; proficiency skill tests; chart review; brief survey by prospective employers plus number who went to work.</td>
<td>P1, P2: ↑ self-esteem. P2: ↑ shift toward external LOC. P1: strongly supported. P6, P7: strongly supported. * Data gaps and reduced sample size.</td>
</tr>
</tbody>
</table>

**Note.** SED = severe emotional disturbance; H1 = Hypothesis 1; H2 = Hypothesis 2; H3 = Hypothesis 3; T1 = Time 1; T2 = Time 2; LOC = locus of control; P1 = Phase 1; P2 = Phase 2; P6 = Phase 6; P7 = Phase 7.  
* Gray literature study.  
* Confounder/limitation.
<table>
<thead>
<tr>
<th>Study design</th>
<th>Sample</th>
<th>Hypotheses</th>
<th>Intervention/comparison</th>
<th>Measures</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trotter, Chandler, Goodwin-Bond, &amp; Casey (2008): Observational pre-/post experimental comparison (efficacy)</td>
<td>( N = 164 ), boys = 102, girls = 62. Majority White, at-risk third to eighth graders. Treatment ( n = 126 ); comparison ( n = 38 ).</td>
<td>( H_1 ): self-awareness; healthy relationships ↑; ( H_2 ): problem behaviors ↓.</td>
<td>EAC: 12 weeks for 2 hr each, 6-8 group participation. Traditional talk, group processing, EAGALA model plus complementary adventure-based treatment activities. Provided by an experienced and certified mental health counselor plus equine specialist. Manualized. Comparison: (Exemplary Substance Abuse Prevention Award: SAMHSA, 1999): indoor, school-based group counseling, 6-8 participants, 12 weeks for 1 hour. Provided by trained school counselor. Manualized.</td>
<td>BASC SRS and PRS pre- and posttest. + PSF (Chandler, 2005; not standardized) to track progress between sessions. Comparison group data not begun until second semester → data for only first semester.</td>
<td>( \downarrow ) negative behaviors and ( \uparrow ) positive behaviors, EAC SRS improved; comparison: ( \downarrow ) negative behavior and ( \uparrow ) positive behaviors, BASC SRS.</td>
</tr>
</tbody>
</table>

**Note.** EAC = equine assisted counseling; \( H_1 \) = Hypothesis 1; \( H_2 \) = Hypothesis 2; EAGALA = Equine Assisted Growth and Learning Association; SAMHSA = Substance Abuse and Mental Health Services Administration; BASC SRS = Behavior Assessment System for Children Self-Report Scale; BASC PRS = Behavior Assessment System for Children Parent Self-Report; PSF = Psychosocial Session Form; Animal Assisted Therapy.

\( a \) Confounder/limitation.
In examining the results of EFP for women with complex posttraumatic stress disorder who had a history of refusal of, or failure to benefit from, traditional group therapy, Shambo et al. (2008) found statistically significant decreases in levels of depression ($p = .03$) from pre- to posttest, and from posttest to follow-up at 4 months ($p = .02$). Pre- and posttest scores on the Dissociative Experience Scale were not statistically significant from pre- to posttest, but demonstrated statistical significance from pretest to follow-up ($p = .023$). Pre- and posttest scores on the Outcome Questionnaire were statistically significant ($p = .017$), as well as from pretest to follow-up ($p = .011$).

**Program evaluations.** Table 5 contains studies that evaluated existing programs. Greenwald (2001) measured the effects of recreational model EAAs in a residential school for children and adolescents with severe emotional disturbances and multiple mental health difficulties, and predicted that participation in the program would be correlated with increases in levels of frustration tolerance and self-esteem, and decreases in levels of anxiety and depression. Results indicated that levels of self-esteem and frustration tolerance were not affected. Outcome measures indicated a correlation between bonding with horses and decreased anxiety and depression. However, it is pertinent that participants with higher levels of depression and anxiety were less likely to be involved in the horsemanship program. The final hypothesis that behavior and overall conduct would show positive effects was not supported.

Iannone (2003) investigated a vocational preparation EAA program for institutionalized adolescents with emotional difficulties, who had limited or no success with previous interventions, and predicted that positive changes in levels of self-esteem, internal LOC, psychiatric symptoms, and behavioral problems would occur in relation to time in the program compared with a wait-listed group. In addition, scores on a vocational skills test were reported as a measure of work-readiness. Statistically significant increases in self-esteem levels were reported during Phase 1 of the three-phase program, with no other significant changes. During Phase 2, levels of self-esteem were reported to continue to increase, and an increasing shift toward external LOC was noted. No other significant changes were reported for Phase 2. Overall, positive effects on levels of self-esteem were the only statistically significant results obtained, with no measurable positive results on psychiatric symptoms and behavior problems, and a change in the opposite direction than was predicted for LOC. Scores on the vocational skills test did improve, as predicted, and were in the passing range as defined by the author.

**Experimental study.** In a comparison study of the efficacy of EFP with added “adventure” experiential components compared to an award-winning school-based group counseling program (see Table 6), Trotter et al. (2008) demonstrated statistically significant decreases in negative behaviors and increases in positive behaviors on five subscales of the Behavior Assessment System for Children (BASC) Self-Report Scale, and 12 subscales of the BASC Parent Self-Report Scale, and one item on the BASC Parent Self-Report. Both interventions were shown to be efficacious, but the experimental equine-assisted counseling group demonstrated more positive gains and was judged to be superior to the control school-based intervention.

**Study Quality**

The studies included in this review used a broad range of equine-facilitated and -assisted therapeutic techniques and measurement instruments to address a wide variety of biopsychosocial outcomes in vastly different clinical populations, making comparison difficult. Two studies used the Rosenberg Self-Esteem Scale (Greenwald, 2001; Iannone, 2003). Three studies used the Nowicki-Strickland Locus of Control Scale (Bowers & MacDonald, 2001; Ewing et al., 2007; Iannone, 2003). Three studies used the Self-Perception Profile for Children (Ewing et al., 2007; Kaiser et al., 2006; Kaiser et al., 2004) and one used the Self-Perception Profile for Adolescents (Bowers & MacDonald, 2001). The Children’s Depression Inventory and the Children’s Loneliness Questionnaire were both used in two studies (Bowers & MacDonald, 2001; Ewing et al., 2007). Ten of the studies reported reliability and validity for the instruments used (Bowers & MacDonald, 2001; Ewing et al., 2007; Graham, 2007; Greenwald, 2001; Iannone, 2003; Kaiser et al., 2006; Kaiser et al., 2004; Russell-Martin, 2006; Schultz et al., 2007; Trotter et al., 2008). The majority of the studies in this review were conducted with small convenience samples, ranging from $N = 10$ (Shambo et al., 2008) to the largest of $N = 63$ (Schultz et al., 2007). However, two studies employed random-sample comparison design (Graham, 2007; Trotter et al., 2008). Three of the studies employed nonequivalent comparison groups (Iannone, 2003; Shultz, 2005; Trotter et al., 2008). Attrition was not addressed in several studies (Kaiser et al., 2006; Kaiser et al., 2004; Shultz, 2005; Tetreault, 2006; Trotter et al., 2008). The results of the application of GRADE criteria to the studies included for review have been reported in the Method section in Table 1. The application of the GRADE criteria to the aggregate of selected studies yielded a mostly low to very low level of quality, with a few exceptions: Graham’s (2007) study achieved a moderate result according to GRADE criteria, as did Trotter et al.’s (2008) study, the highest level that can be achieved for observational studies using GRADE criteria.

Three of the 12 low quality studies (Ewing et al., 2007; Greenwald, 2001; Shultz, 2005) found no significant positive effects, while the remaining nine found diverse positive effects of various EAAs across conditions and populations (Bowers & MacDonald, 2001; Iannone, 2003; Kaiser et al., 2004; Kaiser et al., 2006; Klontz et al., 2007; Russell-Martin, 2006; Shambo et al., 2008; Schultz et al., 2007; Tetreault, 2006). Shultz (2005) reported clinically significant, although not statistically significant, results in a study of adolescents in treatment, but there were significant limitations and confounders. Likewise, Tetreault’s (2006) study of 10 children with emotional disturbances was affected by significant limitations and confounding factors. It is not surprising that Bowers and MacDonald (2001) were unable to demonstrate statistically significant effects on targeted outcomes when it is considered that a major confounder to the study outcome occurred as a result of the sudden, unexpected, and imminent threat of the closure of the facility where the study was being conducted, an example of the type of difficulty that may be encountered when endeavoring to conduct in vivo studies with human participants. It is worth noting that another study conducted by MacDonald...
(2004), not reviewed here because of the lack of full report access, showed promising quantitative results and statistically significant effects on targeted behaviors with a sample size of 126 in a multiple program evaluation. Two studies achieved statistically significant results in the targeted areas (Iannone, 2003; Shambo et al., 2008), but with methodological limitations.

Two studies (Greenwald, 2001; Iannone, 2003) using a small group recreational therapy EAA model twice weekly for length of stay and a vocational rehabilitation EAA model twice weekly for 18 weeks, respectively, also reported a possible harmful significant outcome: a shift from internal to external LOC among institution-alized teens with a history of previous treatment failure for severe and persistent mental illness. This finding, pending replication, suggests a possible contraindication for these EAA models in this population. The positive effects in these low quality studies included increased motor coordination in one study, increased work readiness among teens in one study, mixed findings on self-esteem effects across studies, relational improvement in adult couples and children in some studies, decreased dissociative and depressive symptoms in adults with posttraumatic stress disorder in one study, increased adaptive functioning in children and teens in some studies, and decreased negative affect among children, teens, and adults across studies.

Thus, the two moderate quality studies (Graham, 2007; Trotter et al., 2008) constituted the best evidence practice for use of EAAs as CATs. These studies were supportive of: (a) brief (5-week) small-group manualized EFP provided by trained facilitators to adults with mild-to-moderate catastrophic loss-related depression for outcomes of increased perceived wellness and decreased depressive symptoms; and (b) standard (12-week) small-group manualized adventure-based EFP provided by a certified THR instructor and mental health specialist to school-identified at-risk children for outcomes of decreased externalizing behaviors, increased prosocial behaviors, and improvement on the BASC.

### Discussion

Does the research support recommendations for EAAs to be integrated into the course of biopsychosocial interventions? Best practice dictates that the decision-making process and data synthesis be designed to reduce the gaps between clinical practice and the empirical evidence base (Gambrill, 2006), but, to date, there have been no systematic reviews undertaken to examine the biopsychosocial effects of EAAs other than hippotherapy on health challenges. This review examines these effects as reported in retrievable research available from around the time of those hip-challenges. The outcomes of decreased negative affect and increased adaptive functioning and perceived wellness, for example, have implications for use of EAAs with conditions known to respond to increased positive affect with decreased pain, decreased absenteeism, and decreased health care utilization (e.g., Gil et al., 2004; Lyubomirsky, King, & Diener, 2005; Raysamb, Tambs, Reichborn-Kjennerud, Neale, & Harris, 2003).

### Study Limitations and Future Directions

These results should be considered in light of several limitations. Five studies could not be retrieved or could not be retrieved in their entirety for full review. A listing of the excluded studies may be obtained from coauthor Selby. Publication bias is another common limitation, which was addressed by inclusion of the gray literature and published white papers. Because of the paucity of studies and the lack of randomized controlled trials to date, the current review is limited by the methodological weaknesses of the studies included. In addition, the heterogeneity of the included studies and instruments make comparison difficult.

Although research on the effectiveness of interventions involving horses is emerging, practice appears to be thriving (Taylor, 2001), a reflection of the common finding that participant and staff perceptions sometimes exceed statistical evidence. The results of this review demonstrate promise, which may inform policy on reimbursement by third-party payers.

The final aim of this systematic examination of the data is to highlight future research needs and recommendations. Although the literature is replete with qualitative studies, the standard for exploratory research in new areas, much work remains to be done in the quantitative realm if this treatment approach is to gain credibility. Evidence from the preliminary studies reviewed here is promising, but they underscore the need for more rigorous investigation, including longitudinal studies and clinical trials. Further high-quality research is needed on the biopsychosocial effects on consumers with additional chronic illnesses and health challenges, particularly with regard to an expanded range of outcomes such as pain management and reduced absenteeism (whether for school or work). Many of the identified studies addressed psychiatric disorders or physical wellness and other outcomes in nonclinical populations such as bereaved persons. Further investigation could also target different types of equine-involved interventions for head-to-head comparisons on varying outcomes for the same condition. For example, this investigation uncovered two systematic reviews that examined mobility effects of hippotherapy for populations with cerebral palsy (Snider et al., 2007; Sterba, 2007); additional studies could be done to investigate hippotherapy effects for the same condition in other domains such as pain management and affective regulation, compared with effects of THR in those domains. Currently, some equine-assisted and -facilitated therapeutic approaches use unmounted activities, while others integrate riding into the treatment plan. It would be helpful to investigate which aspects of these approaches are helpful for which client populations. Some studies investigate EFP, while others investigate a combination of THR and EFP or other EAAs. Dismantling studies...
that excavate the underlying effective mechanisms of action are indicated.

There is a dearth of information comparing this treatment approach to other psychosocial techniques that have been shown to be effective both statistically and clinically. Future studies should include the use of comparison groups receiving established interventions for chronic conditions, such as cognitive–behavioral therapy, exercise and relaxation strategies, support groups, family health education and psychoeducation, and systematic desensitization, examining EAAs as an adjunct or complementary therapy to the established treatments. These studies should seek to establish the dosages, diagnoses, and outcomes for which adjunctive use is effective.

In addition, it would be useful to explore the physiological responses that occur when clients interact with horses in a therapeutic setting. Moreover, few studies exist on the physiological and behavioral effects of this application on the horses that are integral to this approach—three studies were identified during the course of this review (Kaiser et al., 2006; Pyle, 2006; Suthers et al., 2006). Although research into the psychosocial effects of therapeutic techniques employing horses can be considered to be in its infancy, this initial review illustrates that there are a number of preliminary and pilot studies that demonstrate the promise of this approach. The next logical step indicated in the progression is for the design and implementation of high-quality studies with sufficiently large sample sizes based on statistical power analyses. Although it may be practically impossible in investigations of this nature to reach the level of rigor that is the criterion standard in medical research, and on which the principles of evidence-based practice are based, investigators should strive with all due diligence to increase the methodological quality of future studies. Because this review primarily identifies effectiveness studies of this treatment alternative, with only one study designed as a randomized experiment and one as a nonrandomized efficacy trial, it is difficult to differentiate whether any beneficial effects noted were caused by the intervention or other factors. Efficacy studies are necessary to inform evidence-based practice. Only one study was found that examined the efficacy of this approach (Trotter et al., 2008); despite the limitations of the study noted by the authors, the results were very promising, and more research into the efficacy of this treatment approach should be undertaken.

More systematic reviews of high-quality research studies of the psychosocial effects of interacting with equines are necessary. Reviews that are conducted transparently and thoroughly can inform practice and assist consumers in making enlightened and practical determinations in seeking effective and relevant treatment options. Creatively and carefully designed and conducted research and the subsequent promotion of interventions involving equines as a credible form of treatment will be critical to its further implementation and ultimate success.

References


*Iannone, V. N. (2003).*


(Appendix follows)
Appendix

Criteria for GRADE

Type of evidence:
High (randomized, controlled trial)
Low (observational study)
Very low (any other evidence)

Decrease grade if:
- Serious (-1) or very serious (-2) limitation to study quality
- Important inconsistency (-1)
- Some (-1) or major (-2) uncertainty about directness
- Imprecise or sparse data (-1)
- High probability of reporting bias (-1)

Increase grade if:
- Strong evidence of association-significant relative risk of > 2 \( (p < .5) \) based on consistent evidence
- From ≥ observational studies, with no plausible confounders (+1)

Very strong evidence of association-significant relative risk of > 2 \( (p < .2) \) based on direct evidence with no major threats to validity (+2)
- Evidence of dose-response gradient (+1)
- All plausible confounders would have reduced effect (+1)

Ratings:
- High
- Moderate
- Low
- Very low