Equine-Related Treatments For Mental Disorders Lack Empirical Support: A Systematic Review of Empirical Investigations

Michael D. Anestis,[1] Joye C. Anestis,[1] Laci L. Zawilinski,[1] Tiffany A. Hopkins,[1] and Scott O. Lilienfeld[2]

[1] *University of Southern Mississippi*
[2] *Emory University*

Context: Equine-related treatments (ERT) for mental disorders are becoming increasingly popular for a variety of diagnoses; however, they have been subjected only to limited systematic investigation. **Objective:** To examine the quality of and results from peer-reviewed research on ERT for mental disorders and related outcomes. **Method:** Peer-reviewed studies (k = 14) examining treatments for mental disorders or closely related outcomes were identified from databases and article reference sections. **Results:** All studies were compromised by a substantial number of threats to validity, calling into question the meaning and clinical significance of their findings. Additionally, studies failed to provide consistent evidence that ERT is superior to the mere passage of time in the treatment of any mental disorder. **Conclusion:** The current evidence base does not justify the marketing and utilization of ERT for mental disorders. Such services should not be offered to the public unless and until well-designed studies provide evidence that justify different conclusions. © 2014 Wiley Periodicals, Inc. J. Clin. Psychol. 70:1115–1132, 2014.

Keywords: equine assisted psychotherapy; therapeutic horseback riding; pseudoscience; evidence based psychotherapy

The establishment of empirical support for specific treatments for specific mental disorders requires several steps (see Chambless et al., 1998, for a summary). Some steps prioritize internal validity, which necessitates rigorously controlled evidence that the treatment outperforms control conditions or other active treatments, thereby demonstrating that the treatment itself (as opposed to nonspecific factors, such as placebo effects or regression to the mean) affects client outcomes. Other steps prioritize external validity, which necessitates evidence that the treatment produces positive effects in the real world, thereby demonstrating generalizability to treatment settings that differ from those seen in rigorously controlled intervention studies.

Furthermore, both steps ideally test for incremental efficacy above and beyond extant treatments to justify increases in cost (direct or indirect). This system provides a set of checks and balances to ensure that scientific integrity (e.g., use of a proper sample, faithful implementation of treatments) is consistent throughout and that validity threats common in one setting (e.g., diminished diagnostic comorbidity in controlled studies) are addressed in studies in different settings (e.g., greater comorbidity in a later effectiveness trial). In this sense, a treatment must earn its position in the marketplace through a stepwise process whereby multiple independent research teams establish that the intervention yields consistent and clinically meaningful improvements that are commensurate with the costs associated with the intervention.

One novel but widely administered set of interventions that has been subjected only to minimal empirical investigation is equine-related treatment (ERT), a category that includes both equine-assisted psychotherapy (EAP) and therapeutic horseback riding (THR). EAP is an experiential therapy that utilizes equines to facilitate personal exploration (e.g., awareness of thoughts, emotions, and behaviors) and provide a context within which the therapist and client can interpret the nature and meaning of the client's difficulties (e.g., Klontz, Bivens, Leinart,

Please address correspondence to: Michael D. Anestis, Department of Psychology, University of Southern Mississippi, 118 College Drive, Box #5025, Hattiesburg, MS 39406. E-mail: michael.anestis@usm.edu

 © 2014 Wiley Periodicals, Inc.
Published online in Wiley Online Library (wileyonlinelibrary.com/journal/jclp). DOI: 10.1002/jclp.22113

& Klontz, 2007). The equines are intended to serve as prompts for specific responses and as metaphors through which to understand internal experiences (e.g., Bachi, Terkel, & Telchman, 2012; Klontz et al., 2007).

Furthermore, the equine is theorized to provide feedback regarding subtle changes in mood, with some authors positing that the equine serves as a "large biofeedback machine" (Professional Association of Therapeutic Horsemanship [PATH] International, 2013a). The client engages in a range of equine-related activities (e.g., selecting an equine, grooming, feeding) and the client and therapist determine the specific targets for treatment on a case-by-case basis. EAP differs from THR, in which clients gain horsemanship skills under the supervision of sidewalkers who monitor the client and equine interaction and optimize safety. In THR, the experience of riding and directing the equine through voice commands is thought to confer therapeutic benefits (e.g., Jenkins & DiGennaro Reed, 2013). The primary goals of THR involve amelioration of physical and communication difficulties (e.g., Bertoti, 1988; Snider, Korner-Bitensky, Kammann, Warner, & Saleh, 2007).

A recent systematic narrative review by Selby and Smith-Osborne (2013) examined the effectiveness of complementary and adjunctive interventions involving equines. The authors did not restrict their review to the use of ERT in the treatment of mental illness; instead, they reported results from several trials examining such outcomes and referenced several others. The authors concluded that the overall results were promising and justified consideration of ERT as a potentially useful adjunctive approach in the treatment of mental illness (Selby & Smith-Osborne, 2013). The review did not speak directly to the utility of ERT as a standalone treatment; however, the treatment is often marketed as such (e.g., PATH International, 2013b) and the financial costs associated with treatment are often substantial (e.g., SouthSource, 2013; True Balance PLLC, 2013).

Furthermore, although Selby and Smith-Osborne (2013) did not provide any evidence that adjunctive ERT is iatrogenic (psychologically harmful), they also did not offer a clear theoretical rationale or empirical test of the processes of change associated within the treatment. As such, it remains unclear how ERT is theorized to provide an effect that extends beyond those of either common factors (e.g., placebo effects, regression to the mean) or other treatments (if ERTs are utilized as adjuncts).

The current review seeks to systematically review the empirical support for ERT in the treatment of mental disorders and associated outcomes (e.g., emotional difficulties). Our aim is to establish the degree to which research on these interventions meets conventional standards of quality and indicates efficacy and effectiveness in the treatment of mental disorders or related conditions. Given the increasing popularity, promotion, and accessibility of these treatments (e.g., PATH International, 2013b), it is paramount to determine the degree to which ERT has been reliably shown to produce clinically meaningful outcomes for individuals presenting with specific diagnoses.

Indeed, ERT has been promoted on widely popular "celebrity" mental health programs (e.g., *Dr. Phil, The Dr. Oz Show*), and some practitioners have built upon such coverage to make sweeping claims about the utility of ERT across a broad range of conditions. For example, one promotion claims that "research has proven it [ERT] to be an effective form of alternative therapy for people struggling with issues like depression, addiction, eating disorders, recovery from trauma or abuse, and more," noting that "in a recent study of Equine Assisted Psychotherapy (EAP), 82% of teens showed more improvement in just five sessions of EAP than they had in years of traditional therapy" (Corcoran, n.d.).

Such powerful claims levied by professionals and marketed by celebrities run the risk of driving up opportunity costs by directing individuals with diagnoses for which evidence-based treatments exist (a) away from treatments with a strong empirical foundation and (b) towards alternative approaches devoid of such support. Given the clinical and economic implications of such concerns, it is imperative that consumers be provided with coherent and comprehensive summaries of the evidence for treatments available in the marketplace.

Consistent with the approach used in evaluating other novel therapies (e.g., Marino & Lilienfeld, 1998, 2007), we used established validity criteria (i.e., Cook & Campbell, 1979; Kendall, Flannery-Schroeder, & Ford, 1999; Shadish, Cook, & Campbell, 2002) to systematically evaluate

the quality of the research underlying ERT. In doing so, we also sought to clarify two other critical points: the extent to which (a) studies actually assess the same form of treatment (e.g., EAP vs. THR) and (b) findings from individual studies provide support that the treatment is efficacious and effective in addressing any specific mental illnesses, as either a standalone or adjunctive intervention.

Our review differs in several meaningful ways from the review by Selby and Smith-Osborne (2013). First, we aimed to provide a description of threats to the internal and external validity of each study in an effort to place findings in their appropriate context. Second, we included only peer-reviewed publications that present research findings. Third, we did not restrict our review to the use of ERTs as adjunctive interventions. In this sense, although our reviews approach similar topics, they do so in an entirely dissimilar manner.

Method

The current review examined peer-reviewed empirical investigations of the efficacy and effectiveness of ERT in the treatment of mental illness. We searched for papers relevant to this topic in a variety of ways. First, we entered the search term "equine" and the search terms "therapy," "psychotherapy," and "therapeutic" into the PsycINFO database. We then examined the reference lists of identified studies to uncover others that would meet our criteria. Finally, we obtained references through personal communications with ERT practitioners and other colleagues.

To be included in our review, studies were required to meet a number of criteria. First, they needed to involve the use of an intervention in which an equine plays a pivotal role either as a standalone or adjunctive treatment of a specific mental illness, a variety of mental illnesses, or conditions closely related to mental illness (e.g., "at risk adolescents"). The latter group, though broad, was included due to concerns that limiting our review to specific diagnoses would prohibit examination of relevant research that did not rely on standard diagnostic categories.

Additionally, to be included in our review, the studies needed to employ an experimental protocol (between-groups or single-case design) and report how efficacy and effectiveness was assessed. The search and subsequent application of inclusionary criteria yielded 14 studies (see Table 1). Ten of these studies used child participants, whereas four used adult participants. Eight studies reported outcomes of THR and related programs, and six utilized EAP or related approaches.

Results

A summary of the threats to validity within each study meeting inclusion criteria can be found in Table 1. Threats for which zero or one study met criteria were not included in the table. Each threat is included within one of five categories: (a) threats to statistical conclusion validity (the degree to which appropriate statistical approaches were used and conclusions matched data); (b) threats to internal validity (the degree to which the study exhibited precision and specificity in measurement); (c) threats to construct validity (the degree to which the study assessed the constructs of interest and provided results that adequately tested hypotheses); (d) threats to external validity (the degree to which the findings generalize to other populations); and (e) matters of design (flaws in the design of the study). Matters of design, in some ways, touch upon issues related to several of the other categories; however, we opted to consider them separately in an effort to highlight their importance in clinical research. Examination of Table 1 reveals that each study was found to be in violation of numerous threats to validity (modal number of threats = 11, median = 13.5, range = 9–21).

Additionally, three violations occurred across most studies. These violations are the result of an almost uniform lack of appropriate experimental controls, proper experimental procedures necessary to test treatment outcome, and independent, unbiased raters. Any one of these violations would typically disqualify studies in the psychotherapy outcome literature from serious scholarly consideration. Due to the frequency of these violations, they are each addressed in isolation below. Additionally, a summary of study characteristics (e.g., description of sample, presence and absence of treatment manual) can be found in Table 2. Following the discussion of

Table 1
Threats to Validity in Each of the Studies That Met Inclusion Criteria for the Review

	Bass	Bizub	Bachi	Burgon	Corring	Cuypers	Ewing	Gabriels	Jenkins	Kaiser	Klontz	Lutter	Schultz	Trotter
Threats to statistical conclusion validity														
Low statistical power[a,b]	X	X	X	X	X	X		X	X	X	X		X	X
Fishing and error rate problem[a,b]	X								X					X
Unreliability of measures[a,b]						X	X						X	X
Unreliability of treatment implementation[a,b]			X	X	X		X					X		
Heterogeneity of units (respondents)[a,b]	X	X	X	X								X		
Threats to internal validity														
Ambiguous temporal precedence[a,b]		X		X	X									
Selection[a,b]	X		X					X			X	X		X
History[a,b]		X		X		X						X	X	X
Maturation[a,b]		X		X				X		X	X	X	X	X
Attrition[a,b]				X			X	X			X	X	X	X
Testing[a]						X	X	X						
Additive and interactive effects of threats to internal validity[a,b]		X	X	X	X	X	X	X			X	X	X	
Threats to construct validity														
"Matters of procedure"[c]	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Inadequate explication of constructs[a,b]		X	X	X		X	X			X	X	X	X	X
Construct confounding[a,b]		X	X		X							X	X	X
Mono-operation bias[a,b,c]			X				X			X	X	X	X	
Mono-method bias[a,b,c]	X	X	X							X	X	X		
Reactive self-report changes[a]			X				X							
Reactivity to the experimental situation[a]	X	X	X	X	X					X	X			X
Experimenter expectancies[a,b]	X	X	X	X		X	X		X	X	X	X	X	X
Novelty and disruption effects[a]	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Resentful demoralization[a,b]			X											X

(*Continued*)

Table 1
Continued

	Bass	Bizub	Bachi	Burgon	Corring	Cuypers	Ewing	Gabriels	Jenkins	Kaiser	Klontz	Lutter	Schultz	Trotter
Threats to external validity														
Interaction of the causal relationship with units[a,b]		X	X	X								X		
Interaction of the causal relationship over treatment variations[a]						X						X		
Interaction of the causal relationship with outcomes[a,c]		X		X										
Interactions of the causal relationship with settings[a]			X	X			X					X		
Matters of design[c]														
Posttreatment, follow-up, and intratreatment evaluations[c]	X							X	X	X		X	X	X
Comparing alternative treatments[c]					X	X	X	X	X	X	X	X	X	
Clinical significance[c]	X				X	X			X	X		X	X	X
Mediators and moderators[c]									X	X	X	X		

[a]Shadish, Cook, & Campbell (2002).
[b]Cook & Campbell (1979).
[c]Kendall, Flannery-Schroeder, & Ford (1999).

Table 2
Study Sample and Administration Characteristics

	Treatment length	Sample	Control group	Random assignment	Diagnostic assessment	Manual used	Fidelity check	Follow-Up collected
Bass	12 1-hour sessions, weekly	34 children (female N = 5) diagnosed with autism spectrum disorder	Yes (waitlist control)	No	Method not specified	No	No	No
Bizub	10 2-hour sessions, weekly	5 adults (age 26–46) of varying diagnoses	No	No	Method not specified	No	No	No
Bachi	50 minute sessions, weekly for 7 months	29 adolescents of varying diagnoses	Yes (no treat-ment)	No	Method not specified	No	No	Yes (12-month)
Burgon	Varied by participant	7 foster children with varying diagnoses (e.g., ADHD, autism)	No	No	Method not specified	No	No	No
Corring	10 1.5 hour group sessions, weekly; Adjunctive to Assertive Community Treatment	6 outpatients (age 36–59) diagnosed with schizophrenia or schizoaffective disorder	No	No	SCID-I (First & Gibbon, 2004)	No	No	No
Cuypers	16 sessions, twice weekly	5 boys (age 10–11) diagnosed with ADHD	No	No	Method not specified	No	No	Yes (8 weeks)
Ewing	2 hour sessions, twice weekly for 9 weeks	28 adolescents (age 10–13) of varying diagnoses	Yes (waitlist control)	No	Method not specified	Yes	No	No
Gabriels	10 1-hour sessions, weekly	42 (female N = 6) individuals (age 6–16) diagnosed with autism spectrum disorder	Yes (waitlist control)	No	Method not specified	No	No	No
Jenkins	9 1-hour sessions, weekly	7 children (age 6–14) diagnosed with autism spectrum disorder	Yes (waitlist control)	No	Method not specified	No	No	No
Kaiser	5-day camp	16 children (age 7–17) with no mental or physical disability	No	No	Method not specified	No	No	No
Klontz	4.5-day residential treatment program	66 individuals (age 23–70) of unknown diagnostic profile; results only provided for 31 with complete data	No	No	Method not specified	Yes	No	Yes (6-month)

(*Continued*)

Table 2
Continued

	Treatment length	Sample	Control group	Random assignment	Diagnostic assessment	Manual used	Fidelity check	Follow-Up collected
Lutter	Varied. Some equine elements across diverse treatments	72 women (age 18–64) diagnosed with eating disorder(s)	No	No	Method not specified	No	No	No
Schultz	0-116 sessions	63 individuals (age 4–16) diagnosed with a range of mental illnesses	No	No	Method not specified	Yes	No	No
Trotter	12 2-hour group sessions, weekly	164 "at risk" children (3rd to 8th grade)	Yes (school-based counseling program)	No	Method not specified	Yes	No	No

Note. control group = presence/absence of a control group in the study; random assignment = presence/absence of random assignment of participants to condition; diagnostic assessment = method utilized to develop diagnostic decisions; manual used = presence/absence of treatment manual; fidelity check = presence/absence of treatment fidelity checks; follow-up collected = presence/absence of posttreatment data collection.

several specific validity threats as they applied across studies, descriptions of the findings from each study are provided, thereby allowing for a succinct summary of the evidence base for the efficacy and/or effectiveness of ERT in the treatment of mental disorders and related conditions.

Novelty Effects

Novelty effects refer to the elation and energy that ensue from a new and exciting experience (Shadish et al., 2002). Equine approaches are especially vulnerable to novelty effects because interacting with a horse is probably an unusual and exciting experience for most individuals. The potential importance of this effect on equine therapy research is highlighted by Trotter, Chandler, Goodwin-Bond, and Casey (2008), in which children chose between ERT and a school-based therapy, thereby violating the key principle of random assignment. As would be expected, substantially more children opted for the equine program. Combine the novelty of the animal itself with the experience of journeying to a farm, ranch, or camp, spending time outdoors, and at times participating in additional adventure therapy activities (e.g., Trotter et al., 2008), and it is impossible to parse the effect of the intervention from other factors.

For a condition to control for the novelty effects inherent in ERT, participants in the control condition would need to interact with another large and familiar social animal (e.g., cow), ideally holding all other factors equal (e.g., setting). It should be noted that this would be a particularly stringent test, as it assumes that therapeutic gains are specific to horses rather than generalizing to other large animals. An additional possibility would be to replicate all aspects of the treatment except for the horse (e.g., bringing participants to the ranch, meeting for the same amount of time and number of sessions, conducting all aspects of the treatment except for the horse-specific components). Although 6 of the 14 studies reviewed used a control group (see Table 2), none controlled for the presumed novelty of the treatment.

Experimenter Expectancies

When an experimenter is not blinded to the participant's condition, experimenter expectancies become plausible concerns (Cook & Campbell, 1979; Shadish et al., 2002). Experimenters can convey their expectations about desirable responses to the participant, thereby influencing participant response. Such expectations can affect client responses on self-report and other treatment outcome measures. This confound was deemed to be a problem in 12 of the 14 studies.

An additional concern regarding the role of experimenter bias focuses on the subjective ratings of treatment outcome completed by therapists in many of these studies. When subjective ratings of treatment outcome are necessary, employing unbiased raters who are independent of the treatment team and the investigation and who are blinded to the study hypotheses is paramount. In at least five of the examined studies, ratings of outcome or treatment benefit were conducted by the therapists themselves or the participant's treatment team, all stakeholders in the effectiveness of the experimental condition.

Matters of Design

Kendall et al. (1999) noted the importance of procedural issues in the design of treatment outcome research. We focus on two here: manual-based treatments and fidelity or integrity checks. To truly replicate the implementation of a particular treatment, the treatment must be manualized or at least described explicitly. Of the 14 studies examined, six did not report the use of any manual or guidelines (see Table 2). Four studies described general guidelines for treatment but did not provide specific steps or manuals for these guidelines, and many times noted that they were modified for individual participants (Bass, Duchowny, & Llabre, 2009; Bizub, Joy, & Davidson, 2003; Gabriels et al., 2012; Jenkins, & DiGennaro Reed, 2013). A brief comparison of these guidelines makes clear that they differed greatly across studies of the "same" intervention. Four studies reported using a treatment manual, albeit a different manual or protocol was used in each one (Ewing, MacDonald, Tayloer, & Bowers, 2007 ["Horse Power"]; Klontz, et al., 2007;

[Longville, 2000]; Schultz, Remick-Barlow, & Robbins, 2007 [EAGALA guidelines]; Trotter et al., 2008 [Trotter, 2006]).

Merely because a therapy manual or treatment guidelines exist and are used, however, does not guarantee that the treatment provided is the treatment described or that the treatment is provided competently. As such, integrity or fidelity checks by expert and impartial judges are necessary to ascertain and maintain fidelity to the treatment. Although some of the studies mentioned following guidelines and manuals, none of the 14 studies reported having any evaluation of treatment fidelity in place. This combination of lack of therapy structure and lack of a system of checks to determine the faithful implementation of the treatment raises significant questions about the independent variable under study.

It should be noted that some groups (e.g., EAGALA) require that therapists who provide ERT be trained and certified in its administration. This training does not eliminate concerns regarding treatment fidelity–therapist drift is possible once training has been completed, particularly in the absence of a treatment manual–but it does indicate that efforts have been made to increase treatment adherence.

Therapeutic Horseback Riding

For each study described below, see Tables 1 and 2 for a description of threats to validity and methodology.

Bizub et al. (2003). Bizub et al. (2003) evaluated the efficacy of THR for adults with severe psychiatric illness. Transcripts from semistructured interviews were analyzed for thematic content and reported qualitatively. The authors concluded that participants (a) were influenced by expanded "reach" into the physical and interpersonal world, (b) acquired a sense of agency, (c) developed increased coping skills and insight, and (d) experienced lower levels of psychiatric symptoms.

Burgon (2011). Burgon (2011) utilized qualitative methods to examine the effects of a therapeutic horsemanship program on risk and resilience factors among at risk youth. Participants (female $n = 5$; male $n = 2$) were foster children with a variety of diagnoses, perceived disadvantages, or both. Dosage and length of the program varied widely. The author disclosed prior relationships with several participants in the study and reported collaborating and consulting with participants regarding the research design and methods, thereby raising concerns about experimenter expectancies and demand characteristics. The author concluded that four main themes emerged, including improved confidence and self-esteem, an increased sense of mastery and self-efficacy, development of empathy, and initiation of positive future opportunities.

Corring, Lundberg, and Rudnick (2013). Corring et al. (2013) examined whether THR can benefit patients diagnosed with schizophrenia or schizoaffective disorder as an adjunct to the care provided by Assertive Community Treatment teams. Participants included six outpatients (*female* $n = 2$; aged 36–59 years). Results of a comparative thematic analysis revealed four themes related to patient outcomes: having fun, bonding relationship with horse, increased confidence and self-esteem, and relationship gains. Although the findings seem encouraging, the relevance of these outcomes to the treatment of mental disorders is questionable.

Cuypers, de Ridder, and Strandheim (2011). The researchers investigated the effects of THR on children with attention deficit hyperactivity disorder (ADHD). Five 10–11-year-old boys who were diagnosed with ADHD and responded positively to and were stabilized on medication for more than one year were selected to participate. Results suggested improvement in various domains of behavior, quality of life, and motor performance from pre- to postintervention; however, the effects were inconsistent across measures from posttreatment to follow-up. Type I errors were a serious threat to validity, as 165 statistical comparisons were conducted, with no correction for family-wise error.

Kaiser, Spence, Lavergne, and Vanden Bosch (2004). In this study, 16 children aged 7–17 years took part in a 5-day THR camp. Any individual exhibiting mental or physical disability was excluded, resulting in exclusion of 20% of the initial sample ($n = 20$). Participants took part in THR activities, classroom activities centered on equine-related topics, and occasional recreational trail rides. Participants provided pre- and posttreatment data on measures of anger, quality of life, and perceived self-competence. Significant pre-post differences were noted on total anger and three anger subscales, but no significant differences were noted in any of the other 12 analyses conducted on quality of life and perceived self-confidence outcomes.

Bass et al. (2009). This study utilized a waitlist controlled trial to determine if THR improves social functioning in children diagnosed with autism spectrum disorders. The total sample comprised 34 children (female $n = 5$). Although treatment was not manualized, all sessions included the same list of activities spread out over the same time intervals. The authors conducted pre- and posttreatment measures of social functioning, which were administered to parents in both groups. THR was associated with significant and clinically meaningful improvement on overall Sensory Profile and Social Responsiveness Scale scores, as well as the subscales related to sensory seeking, attention and distractibility, sensory sensitivity, sedentary, and social motivation.

Gabriels et al. (2012). Gabriels and colleagues (2012) conducted a pilot study investigating the effects of THR on self-regulation behaviors, adaptive skills, and motor skills in children and adolescents diagnosed with an autism spectrum disorder ($N = 42$; aged 6–16 years). Sixteen of the 42 participants were assigned to a waitlist control condition prior to the intervention. Within-group results indicated that the THR group displayed significant improvements from baseline to posttreatment on self-regulation, adaptive-expressive language skills, motor skills, and verbal praxis and motor planning skills. In addition, the THR condition outperformed the waitlist condition on indices of irritability, lethargy, stereotypic behavior, and hyperactivity.

Jenkins and DiGennaro Reed (2013). In this study, seven children diagnosed with an autism spectrum disorder (aged 6–14 years) were divided into THR ($n = 4$) and waitlist ($n = 3$) groups. No information was provided regarding how group assignment was established or whether individuals in each group differed from one another on any treatment-relevant variables. Participants were observed at their homes and during an after-school program located at the same center as the THR activities. No information was provided as to whether children on the waitlist took part in the after-school program during the study, an important point to consider given the potential for "resentful demoralization," which is caused by exclusion from a novel experience in which other participants openly participated (Cook & Campbell, 1979).

The authors assessed behaviors indicative of nine outcomes: happiness, unhappiness, spontaneous initiations of linguistic communication, contextually appropriate responses to initiations by others, off-task behavior, compliance, problem behavior, commands to direct the horse, and posture. Additionally, the authors collected pre- and posttreatment data from the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), with both teachers and parents providing data. Results indicated that THR was not associated with significant improvements on any outcome other than posture. Additionally, no between-groups differences between the THR and waitlist groups were reported for any variables.

On the CBCL, which assesses internalizing, externalizing, and total childhood problems, the authors reported that, pretreatment, only 1 of 12 CBCL scores in the treatment group was in the clinical range, whereas six of nine CBCL scores in the waitlist group were in the clinical range. Posttreatment, 3 of 12 CBCL scores in the treatment group were in the clinical range and 1 of 9 CBCL scores in the waitlist group was in the clinical range. On the teacher rating form, 2 of 12 pretreatment scores were in the clinical range for the treatment group and 2 of 9 were in the clinical range in the waitlist group. Posttreatment, 4 of 12 scores were in the clinical range in the treatment group and 3 of 9 were in the clinical range in the waitlist group. In conjunction with the other results already noted, there appears to be no compelling evidence that THR is effective for children with an autism spectrum disorder.

Equine-Assisted, Equine-Facilitated Psychotherapy and Counseling

For each study described in the following section, see Tables 1 and 2 for a description of threats to validity and methodology.

Lutter and Smith-Osborne (2011). The authors examined the correlation between (a) the level of metabolic activity expended in ERT when this treatment was used as an adjunct intervention for eating disorders and (b) eating disorder and mood symptoms. The authors hypothesized that participants who expended more energy in ERT would improve more than those who expended less. The sample comprised 72 women (aged 18–64 years) whose treatment involved one or more of the following: unspecified group psychotherapy, unspecified individual psychotherapy, unspecified family psychotherapy, equestrian programs, canine therapy, experiential challenge course, and recreation therapy.

Results indicated that a greater level of physical activity in equine-related activities predicted greater change in both eating disorder and depressive symptoms, above and beyond the effects of eating disorder diagnosis and length of stay. Although promising, this analysis did not control for baseline levels of psychopathology. As such, given the absence of a control group, it is unclear to what extent greater physiological activity reflected more intensive treatment for more severe individuals (and a greater likelihood of regression to the mean). Furthermore, given the lack of control for type and amount of other forms of treatment given, there is no way to determine to what extent the physiological activity involved in equine-related activities influenced the outcomes. These results provide modest reassurance that the equine-related activities were not iatrogenic, but do not bear directly on the effectiveness of equine-related activities in the treatment of eating disorders.

Klontz et al. (2007). In this study, 66 individuals (aged 23–70 years) who entered a 4.5-day residential treatment program were given the opportunity to participate in EAP. Only 31 individuals remained in treatment for all three time points (pretreatment, posttreatment, 6-month follow-up), and the authors provided information only on those 31 participants (no information was provided as to whether significant differences existed between groups on any relevant variables).

The authors reported that overall symptom severity and self-actualization improved significantly from pre- to posttreatment and remained stable at 6-month follow-up. The lack of a comparison group, however, precludes any determination regarding the impact of treatment above and beyond common factors and regression to the mean. In addition, the lack of intent-to-treat analyses leaves open the possibility that dropouts were treatment nonresponders or more severe clients, thereby spuriously inflating positive outcomes.

Schultz et al. (2007). Schultz et al. (2007) presented a cross-sectional pilot study examining the impact of EAP on a broad variety of childhood behavioral and mental health problems. The sample ($N = 63$; aged 4–16 years) was reported to meet criteria for a broad array of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition diagnoses. The treatment team determined Global Assessment of Functioning (GAF) scores at pretreatment and at 3-month intervals until the conclusion of treatment.

The authors reported a significant improvement on GAF scores. In fact, they noted that all children included in the analyses had a GAF increase (improvement), and that there was a significant correlation between GAF scores and number of sessions attended ($r = 0.73$). Nevertheless, these results are difficult to interpret given several methodological problems. First, the use of and reliance on the GAF as the sole measure of treatment outcome compromises the construct validity of the study. In this case, mono-operation bias (the use of only one operationalization of a construct) is compounded by the use of an unreliable measure (e.g., Garb, 1998). These problems are further exacerbated by an absence of inter-rater reliability data or, at a minimum, ratings conducted by an independent reviewer (i.e., not a member of the treatment team).

Additionally, sampling issues raise questions about the internal validity of the study. The authors noted that they excluded children who received less than six sessions ($n = 14$), meaning that nearly 25% of the sample was removed prior to analysis. They did not mention the rationale for excluding these children. This decision could very well have influenced the findings, given the large correlation between GAF scores and number of treatment sessions (range = 0–116).

Bachi et al. (2012). Bachi et al. (2012) hypothesized that EAP would result in improvements in self-control, self-image, and trust among at-risk adolescents ($N = 29$). The study compared the EAP treatment group with a no-treatment control group; placement was nonrandomized and based on case manager referral. Measures and interviews were completed pre- and postintervention; brief interviews were conducted 1-year postintervention. The authors reported no significant differences between groups on any outcome.

Ewing et al. (2007). Ewing et al. (2007) evaluated the effectiveness of ERT for adolescents with severe emotional disturbances. Participants included 28 adolescents aged 10–13 years who were diagnosed with a range of mental illnesses. The ERT, named Horse Power, comprised caring for the horse, riding the horse, and a process group focused on discussing the experience of Horse Power and individual goals for the day. Participants awaiting the intervention served as the waitlist control condition. Results revealed no statistically significant differences between pre- and posttreatment on any measure.

Trotter et al. (2008). Trotter and colleagues (2008) compared efficacy of equine-assisted counseling (EAC) for at-risk third- to eighth-graders with an in-class, school-based counseling program. Students identified by school counselors as being at risk for school failure, social failure, or both, were invited to participate. The students were allowed to choose between EAC and an in-class, school-based (SB) program. The EAC group ($n = 126$) contained a much larger number of participants than did the SB group ($n = 38$). Of the 205 participants, 41 dropped out, resulting in a final sample of 164.

The authors reported a large number of analyses: within-subjects and between-groups comparisons of two measures of behavioral problems comprising 32 total subscales, and within-subjects comparisons of another measure of behavioral problems comprising three subscales. Significant within-subjects differences were reported for the EAC group for 18 of the 32 subscales of the initial measure.

The authors reported that the SB group demonstrated significant pre-post changes on 5 of the 32 subscales of those same measures. The authors then reported results from seven analyses of covariance (ANCOVAs) demonstrating a "statistically significant greater decrease in five negative behaviors and statistically significant greater increase in two positive behaviors" (p. 273) relative to the SB treatment group. The authors did not explain how they selected these 7 of the possible 32 scales or whether they examined all of the other scales. The final set of results presented was from three repeated-measures ANOVAs, which showed significant improvement on the three scales of the third measure: an increase in overall behaviors (an unconventional outcome apparently representing a sum of all behaviors, positive or negative), an increase in positive behaviors, and a decrease in negative behaviors.

The authors provided no data on pretreatment scores and no comparisons of pretreatment scores between groups. It seems plausible that these groups significantly differed at pretreatment, raising questions about the impact of the intervention. The authors also noted that 41 participants dropped out, but did not report how many dropped out from each group or whether these individuals differed from those who remained in treatment. Instead, dropouts were excluded from the analyses.

Although novelty effects threaten the validity of all 14 studies in this review, the novelty of the EAC treatment may have played a particularly potent role, as EAC participants met at a ranch, whereas control participants met at their school. Because the control group was given the choice of joining the EAC group, they were aware of the novel treatment they declined. As such, resentful demoralization may have played a role. Despite these limitations, the first author claimed on his website that: "the results proved that horse therapy gets authentic results in increasing

positive behaviors in clients while decreasing undesirable responses" (Trotter, n.d.). Furthermore, EAGALA has referred to this study as a randomized control group study that demonstrates equal or superior results for ERT relative to "established interventions" (EAGALA, 2013a).

This study was the only one in our review in which ERT was compared with a form of treatment as usual (TAU). Although a strong argument could be made that the optimal comparison group for any treatment is an evidence-based alternative approach, there is certainly value in ascertaining how a novel treatment compares with what would otherwise be offered in a given environment. In this sense, TAU comparisons enhance external validity by considering the novel treatment within the context of the general marketplace and without the restraints of highly controlled clinical trials.

Summary of Results

When considered in aggregate, findings from studies on ERT provide inconsistent and less than compelling support for its efficacy in the treatment of any mental disorder. Studies that lack control groups often yielded improvement from pre- to posttreatment; however, such findings cannot demonstrate an effect of treatment. Studies with control groups neglected to use random assignment or to report on pretreatment between-group differences on any variable. Furthermore, several studies with control groups were unable to demonstrate favorable results for ERT. Given the lack of consistent follow-up data, the absence of treatment manualization or integrity checks, and several consistent threats to validity (see Tables 1 and 2), the results fall well short of the standards set forth for establishing empirical support for treatments (Chambless et al., 1998).

Conclusion and Summary

The development of new, more effective treatments for mental disorders is a noble goal. The desire to disseminate promising treatments to those in need is equally admirable. Nevertheless, practitioners and researchers sometimes encounter difficulties when the dissemination of a treatment extends beyond the empirical evidence supporting that treatment for a given population. Our goals were to systematically review the empirical literature on EAP and THR in the treatment of mental illnesses, to evaluate threats to the validity of reported findings, and present a cogent and concise summary of the degree to which practitioners are justified in embracing this approach and making broad claims for its efficacy and/or effectiveness. The results allow us to provide an unequivocal summary statement: The empirical literature on equine-related treatments for mental illness is limited in scope, the studies that exist are compromised by multiple methodological flaws, and there is no consistent evidence that the treatments afford benefits beyond those offered by the passage of time. Given the time and expense associated with ERT (and the dissemination of any new treatment), there appears to be scant justification at present for its use as a standalone or adjunctive treatment for any mental disorder.

It is possible that, in coming years, randomized controlled trials with assessments conducted by independent observers, conclusions based on intent-to-treat analyses, and results buffered by long-term follow-ups will provide a compelling evidence base for ERT for specific mental disorders. At that time, it might be appropriate to market these treatments and provide them to those in need, assuming that authors also demonstrate that the treatment is as cost-effective as other evidence-based approaches and is not associated with a greater frequency of problematic outcomes. With respect to the latter concern, it is plausible that the costs associated with treatments involving large animals would be higher than treatments that do not involve such animals. Nevertheless, such rigorous scientific standards are not yet close to being satisfied. Therefore, we have serious scientific and ethical concerns regarding the continued use and marketing of EAP and THR for mental illness or psychological maladjustment more broadly. The rationale for our concern becomes clear when considering that ERT is marketed widely for diagnoses (e.g., posttraumatic stress disorder, eating disorders; EAGALA, 2013a) and populations (e.g. military personnel; EAGALA, 2013b) for which insufficient research evidence exists.

In the case of the military, we found several instances of websites containing the following text or a slightly altered version of it: "The Equine Assisted Growth and Learning Association (EAGALA) evaluated treatment of members of the Georgia National Guard where deployments averaged two years or more. The study revealed that 100 percent of soldiers who completed equine assisted therapy had dramatically reduced stress levels" (e.g., Bona, 2011; Calico Junction New Beginnings Ranch, Inc., 2013; Featherston & Dodson, 2013; New Hope Equine Assisted Therapy, 2013). In no instance was a reference provided, and we were unable to find any source document supporting these claims.

The key concept of informed consent involves explaining to clients the evidence for and against available treatments for their presenting complaint(s). In the context of psychotherapy, this involves explaining whether there is empirical support for the use of a given treatment for a given set of problems and how that support compares with that of other treatments, particularly when efficacious techniques for this problem are already available. The American Psychological Association's (APA's) Ethical Principles of Psychologists and Code of Conduct note that in therapeutic situations "for which generally recognized techniques and procedures have not been established, psychologists inform their clients/patients of the developing nature of the treatment, the potential risks involved, alternative treatment that may be available, and the voluntary nature of their participation" (APA, 2010, p. 13). Given the dearth of empirical evidence for ERT and its continued use and promotion, it seems unlikely that full informed consent is routinely transpiring in treatment centers offering ERT. We urge all practitioners of ERT to inform their clients of the extremely limited evidence base for their intervention, as well as to educate them regarding alternative and better supported treatment options.

Belief in the healing potential of a specific treatment is understandable (Frank, 1993); however, in the absence of rigorous evidence, such beliefs can sometimes be dangerous for consumers, most of whom lack the scientific training to evaluate assertions regarding the treatment's efficacy and effectiveness. The danger of such beliefs rests primarily in opportunity costs. Consumers who come to believe in the effectiveness of an unsupported treatment run the risk of being directed away from effective care and towards an experimental approach less likely to yield beneficial results. As such, it becomes more likely that symptoms will persist longer and increase in severity, which in some cases could legitimately be conceptualized as life threatening (e.g., increased suicidality in prolonged depression). Furthermore, the financial burden on clients and their family members may increase due to the need to seek further interventions following suboptimal treatment.

The lack of consistent positive results across studies is complicated by the extreme variability in the structure and content of the treatment. Indeed, we felt compelled to divide our description of ERT into two groups–ERT and THR–and even within those categories there was substantial heterogeneity in the structure of treatment. In this sense, the extent to which ERT can be legitimately conceptualized as a single intervention is debatable. This limitation highlights the lack of a consistent and coherent theory of change processes in ERT. As such, the manner in which equines might produce treatment outcomes via hypothesized mediators is unknown. Had the literature revealed consistent positive results for ERT in the absence of well-delineated treatment mechanisms, a case could still be made for administering the treatment. Nevertheless, the conjunction of unconvincing research support and the absence of a convincing theoretical rationale is problematic (see David & Montgomery, 2011, and Lilienfeld, 2011, for broader discussions) and speaks to the need for caution in implementing ERT in any setting.

The consistent presence of multiple serious threats to the validity of the results rendered several studies difficult or virtually impossible to interpret. We have already discussed problems related to manualization, treatment fidelity, and experimenter expectancies, but other important threats were common. Many studies included either no control treatment or a suboptimal one (e.g., a waitlist control). In the absence of a control condition, it is impossible to rule out the possibility that apparent positive effects are due to common factors (e.g., therapeutic alliance), the passage of time (e.g., spontaneous remittance), or regression to the mean (i.e., the tendency for extreme scores to become less extreme on retesting). Waitlist control conditions allay some of these concerns, but do not allow authors to demonstrate that their treatment outperforms evidence-based approaches. As a consequence, designs including such conditions cannot address

crucial questions of incremental efficacy (does the treatment yield benefits above and beyond extant treatments?) and cost-effectiveness.

Furthermore, no studies of ERT utilized random assignment, which is a crucial tool for reducing systematic bias in treatment allocation. Although randomization failures can and do occur in treatment studies, randomization minimizes the odds that individuals in one condition differ systematically from those in another condition on variables that can spuriously affect treatment outcome (e.g., initial severity of symptoms, comorbid diagnoses, demographics).

All but one study neglected to indicate how psychiatric diagnoses were established; the lone study that did neglected to report the inter-rater reliability for its diagnostic interviews. As a consequence, our understanding of the effects of ERT on specific diagnostic conditions is necessarily limited. Although the current diagnostic system is certainly fallible and may not be provide an ideal conceptual framework for some clients (e.g., Westen, Novotny, & Thompson-Brenner, 2004), this limitation does not obviate the need to explicate the boundary conditions for which the treatment does and does not apply.

We acknowledge, of course, that although randomized controlled trials are often considered the "gold standard" method by which a treatment acquires empirical support, other valid designs exist. APA Division 12's Task Force for Promotion and Dissemination of Psychological Procedures noted that well-designed single-case experiments can constitute the body of evidence for an empirically supported treatment (Chambless & Ollendick, 2001). Given that the time and financial constraints inherent in randomized controlled trials often make them difficult to undertake, many have called for an increase in single-case, idiographic, or patient-focused research as a way to refine existing treatment protocols and gather data efficiently on novel treatments (e.g., Barlow & Nock, 2009; Borckardt et al., 2008; Lambert, Hansen, & Finch, 2001). These methods, in which the target behavior is assessed repeatedly during a baseline period and then again during the intervention phase (AB design) or in which these phases are repeated (ABAB design), allow participants to serve as their own control. In our review, however, we were unable to locate any ERT single-case studies that utilized an experimental protocol.

In summary, although there is no evidence that ERT is iatrogenic, there is negligible evidence that it offers benefits to individuals with mental disorders or other psychological difficulties. The limited research conducted to date is marked by substantial methodological problems that render the (a) results difficult to interpret and (b) marketing of this expensive treatment ethically problematic. Specifically, because the treatment lacks sufficient support in well-designed trials to be considered empirically supported for any mental illness, the marketing and dissemination of ERT violates the goal of providing the best mental health care to those in need. ERT risks driving up opportunity costs, thereby diverting patients away from evidence-based care and towards experimental approaches with intuitive appeal that are devoid of compelling support.

Although research on animal-assisted therapies presents certain stubborn obstacles that are challenging to overcome (e.g., difficulties with developing a reasonable placebo condition; Chur-Hanson, Stern, & Winefield, 2010), others are more easily addressed (e.g., development of psychometrically strong measures of constructs of interest, use of random assignment and proper control conditions, follow-up data) and should be addressed in future studies. Indeed, we believe that research concerning ERT and other experimental treatment modalities should continue. At the same time, the quality of that research needs to improve, and results supportive of efficacy and effectiveness, incremental validity, and cost-effectiveness need to be obtained and replicated by multiple independent research teams prior to the dissemination of marketing of ERT to practitioners and the general public.

Given the strong evidence base for many treatments for a wide variety of diagnoses (e.g., Chambless et al., 1998), we recommend that, in view of the current evidence base, individuals in need of mental health services avoid seeking out ERT and treatment centers avoid practicing this approach. We further urge major organizations, such as the United States Armed Forces and United States Department of Veterans Affairs, to hold off on the implementation of ERT programs on a wide-scale basis unless and until a strong research foundation for this treatment emerges.

References

Achenbach, T. M., & Rescorla, L. A. (2001). Child behavior checklist for ages 6 to 18. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

Bachi, K., Terkel, J., & Telchman, M. (2011). Equine-facilitated psychotherapy for at-risk adolescents: The influence on self-image, self-control and trust. Clinical Child Psychology and Psychiatry, 17, 298–312. doi:10.1177/1359104511404177

Barlow, D. H., & Nock, M. K. (2009). Why can't we be more idiographic in our research? Perspectives on Psychological Science, 4, 19–21. doi:10.1111/j.1745-6924.2009.01088.x

Bass, M. M., Duchowny, C. A., & Llabre, M. M. (2009). The effect of therapeutic horseback riding on social functioning in children with autism. Journal of Autism and Developmental Disorders, 39, 1261–1267. doi:10.1007/s10803-009-0734-3

Bertoti, D. B. (1988). Effect of therapeutic horseback riding on posture in children with cerebral palsy. Physical Therapy, 68, 1505–1511.

Bizub, A. L., Joy, A., & Davidson, L. (2003). "It's like being in another world": Demonstrating the benefits of therapeutic horseback riding for individuals with psychiatric disability. Psychiatric Rehabilitation Journal, 26, 377–384. doi:10.2975/26.2003.377.384

Bona, E. (2011). War, post trauma, and animal assisted therapy. Retrieved from http://ezinearticles.com/?War,-Post-Trauma-and-Animal-Assisted-Therapy&id=6036618

Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as a natural laboratory for psychotherapy research: A guide to case-based time-series analysis. American Psychologist, 63, 77–95. doi:10.1037/0003-066X.63.2.77

Burgon, H. L. (2011). "Queen of the world": Experiences of "at-risk" young people participating in equine-assisted learning/therapy. Journal of Social Work Practice, 25, 165–183. doi:10.1080/02650533.2011.561304

Calico Junction New Beginnings Ranch, Inc. (2013). Equine therapy for (PTSD) post traumatic stress disorder. Retrieved from http://www.calicojunctionnewbeginningsranch.org/ptsd.html

Chambless, D. C., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P., … Woody, S. R. (1998). Update on empirically validated therapies, II. Clinical Psychologist, 51, 3–16.

Chambless, D. L., & Ollendick, T. H. (2011). Empirically supported psychological interventions: Controversies and evidence. Annual Review of Psychology, 52, 685–716. doi:10.1146/annurev.psych.52.1.685

Chur-Hanson, A., Stern, C., & Winefield, H. (2010). Gaps in the evidence about companion animals and human health: Some suggestions for progress. International Journal of Evidence-Based Healthcare, 8, 140–146.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin.

Corcoran, L. (n.d.). Straight from the horse's mouth: Equine assisted psychotherapy. Retrieved from http://www.bestselfatlanta.com/Sponsored-Topic/Life-Enrichment/Equine-Assisted-Psychotherapy/Straight-From-the-Horse%E2%80%99s-Mouth-Equine-Assisted-Psychotherapy.html

Corring, D., Lundberg, E., & Rudnick, A. (2013). Therapeutic horseback riding for ACT patients with schizophrenia. Community Mental Health Journal, 49, 121–126. doi:10.1007/s10597-011-9457-y

Cuypers, K., de Ridder, K., & Strandheim, A. (2011). The effect of therapeutic horseback riding on 5 children with attention deficit hyperactivity disorder: A pilot study. The Journal of Alternative and Complementary Medicine, 17, 901–908. doi:10.1089/acm.2010.0547

David, D., & Montgomery, G. H. (2011). The scientific status of psychotherapies: A new evaluative framework for evidence based psychosocial interventions. Clinical Psychology: Science and Practice, 18, 89–99.

Equine Assisted Growth and Learning Association. (2013a). Frequently asked questions. Retrieved from http://www.eagala.org/sites/default/files/attachments/EAGALAFAQs_3.3.11_FINAL_0.pdf

Equine Assisted Growth and Learning Association Military Services. (2013b). Retrieved from http://www.eagala.org/military

Ewing, C. A., MacDonald, P. M., Taylor, M., & Bowers, M. J. (2007). Equine-facilitated learning for youths with severe emotional disorders: A quantitative and qualitative study. Child and Youth Care Forum, 36, 59–72. doi:10.1007/s10566-006-9031-x

Featherston, J., & Dodson, M. (2013). Equine assisted psychotherapy. Paper presented at the Santa Clara Valley Brain Injury Conference, San Jose, CA.

Frank, J. D. (1993). Persuasion and healing: A comparative study of psychotherapy. Baltimore, MD: Johns Hopkins University Press.

Gabriels, R. L., Agnew, J. A., Holt, K. D., Shoffner, A., Zhaoxing, P., Ruzzano, S., . . . Mesibov, G. (2012). Pilot study measuring the effects of therapeutic horseback riding on school-age children and adolescents with autism spectrum disorders. Research in Autism Spectrum Disorders, 6, 578–588. doi:10.1016/j.rasd.2011.09.007

Garb, H. N. (1998). Studying the clinician: Judgment research and psychological assessment. Washington DC: American Psychological Association.

Jenkins, S. R., & DiGennaro Reed, F. D. (2013). An experimental analysis of the effects of therapeutic horseback riding on the behavior of children with autism. Research in Autism Spectrum Disorders, 7, 721–740. doi:10.1016/j.rasd.2013.02.008

Kaiser, L., Spence, L. J., Lavergne, A. G., & Vanden Bosch, K. L. (2004). Can a week of therapeutic riding make a difference? A pilot study. Anthrozoos, 17, 63–72. http://dx.doi.org/10.2752/089279304786991918

Kendall, P. C., Flannery-Schroeder, E. C., & Ford, J. D. (1999). Therapy outcome research methods. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), Handbook of research methods in clinical psychology (pp. 330–363). New York, NY: John Wiley & Sons.

Klontz, B. T., Bivens, A., Leinart, D., & Klontz, T. (2007). The effectiveness of equine-assisted experiential therapy: Results of an open clinical trial. Society and Animal: Journal of Human and Animal Studies, 15, 257–267. doi:10.1163/156853007X217195

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. Journal of Consulting and Clinical Psychology, 69, 159–172. doi:10.1037/0022-006X.69.2.159

Lilienfeld, S. O. (2011). Distinguishing scientific from pseudoscientific psychotherapies: Evaluating the role of theoretical plausibility, with a little help from Reverend Bayes. Clinical Psychology: Science and Practice, 18, 105–112.

Lutter, C., & Smith-Osborne, A. (2011). Exercise in the treatment of eating disorders: An alternative view. Best Practices in Mental Health: An International Journal, 7, 42–59.

Marino, L., & Lilienfeld, S. O. (1998). Dolphin-assisted therapy: Flawed data, flawed conclusions. Anthrozoos, 11, 194–200.

Marino, L., & Lilienfeld, S. O. (2007). Dolphin-assisted therapy: More flawed data and more flawed conclusions. Anthrozoos, 20, 239–249. doi:10.2752/089279307X224782

New Hope Equine Assisted Therapy. (2013). Equine therapy for (PTSD). Retrieved from http://newhopeequine.tumblr.com/page/2

Professional Association of Therapeutic Horsemanship International. (2013a). What is equine-facilitated psychotherapy and equine-facilitated learning. Retrieved from http://www.pathintl.org/resources-education/resources/eaat/201-what-is-efpl

Professional Association of Therapeutic Horsemanship International. (2013b). Current fact sheet. Retrieved from http://www.pathintl.org/images/pdf/about-narha/documents/2011-fact-sheet.pdf

Schultz, P. N., Remick-Barlow, G. A., & Robbins, L. (2007). Equine-assisted psychotherapy: A mental health promotion/intervention modality for children who have experienced intra-family violence. Health and Social Care in the Community, 15, 265–271.

Selby, A., & Smith-Osborne, A. (2013). A systematic review of effectiveness of complementary and adjunct therapies and interventions involving equines. Health Psychology, 32, 418–432. doi:10.1037/a0029188

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

Snider, L., Komer-Bitensky, N., Kammann, C., Warner, S., & Saleh, M. (2007). Horseback riding as therapy for children with cerebral palsy: Is there evidence for effectiveness? Physical and Occupational Therapy in Pediatrics, 27, 5–23. doi:10.1080/J006v27n02_02

SouthSource: A Publication of South University. (2013). Can horses help with counseling? Retrieved from http://source.southuniversity.edu/can-horses-help-with-counseling-22974.aspx

Trotter, K. S. (n.d.). Resources. In Harnessing the power of equine assisted counseling. Retrieved July 22, 2013, from http://harnessingthepower.wordpress.com/resources/.

Trotter, K. S., Chandler, C. K., Goodwin-Bond, D., & Casey, J. (2008). A comparative study of the efficacy of group equine assisted counseling with at-risk children and adolescents. Journal of Creativity in Mental Health, 3, 254–284. doi:10.1080/15401380802356880

True Balance PLLC. (2013). Frequently asked questions. Retrieved from http://www.truebalancefarm.com/faq.html

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. Psychological Bulletin, 130, 631–663. doi10.1037/0033-2909.130.4.631

Copyright of Journal of Clinical Psychology is the property of John Wiley & Sons, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.