Pilot study measuring the effects of therapeutic horseback riding on school-age children and adolescents with autism spectrum disorders

Robin L. Gabriels\textsuperscript{a,b,*}, John A. Agnew\textsuperscript{a,b}, Katherine D. Holt\textsuperscript{b}, Amy Shoffner\textsuperscript{c}, Pan Zhaoxing\textsuperscript{a,b}, Selga Ruzzano\textsuperscript{b}, Gerald H. Clayton\textsuperscript{a,b}, Gary Mesibov\textsuperscript{d}

\textsuperscript{a}University of Colorado Denver, 13001 E. 17th Place, Aurora, CO 80045, United States
\textsuperscript{b}Children’s Hospital Colorado, 13123 E. 16th Ave, B130, Aurora, CO 80045, United States
\textsuperscript{c}Colorado Therapeutic Riding Center, 11968 Mineral Road, Longmont, CO 80504, United States
\textsuperscript{d}University of North Carolina, Chapel Hill, 510 Meadowmont Circle, Suite 300, Chapel Hill, NC 27517, United States

Autism spectrum disorders (ASDs) are diagnosed based on three core symptoms involving impaired social interaction and communication abilities along with restricted, repetitive, and stereotyped behaviors and interests (American Psychiatric Association, 2000). Children with an ASD are at high-risk for behavioral disturbances that can impair their quality of life and be of great concern to their caregivers. National surveys and studies indicate higher rates of caregiver stress and frustration related to the care and management of the behaviors of children diagnosed with an ASD compared to other special needs populations (Lecavalier, Leone, & Wiltz, 2006; Schieve, Blumberg, Rice, Visser, & Boyle, 2007). The ASD population is a critical public health issue in the United States due to high prevalence rates, high estimates of health care utilization and costs (Croen, Najjar, Ray, Lotspeich, & Bernal, 2006; Guevara, Mandell, Rostain, Zhao, & Hadley, 2003; Liptak, Orlando, et al., 2006; Liptak, Stuart, & Auinger, 2006; Newacheck & Kim, 2005), and the significant distress these children and their families frequently experience. Clinically, symptoms of irritability are often reasons families seek alternative therapies, medication intervention trials, and hospitalizations for their ASD children.

This preliminary study examined the effects of 10 weekly lessons of therapeutic horseback riding (THR) on 42 participants diagnosed with an autism spectrum disorder (ages 6–16 years) compared to a subset (n = 16) of the total study population who were first evaluated before and after a 10-week waitlist control condition. All participants received baseline and post-condition assessments in the areas of self-regulation (Irritability, Lethargy, Stereotypic Behavior, and Hyperactivity), adaptive living skills, and motor skills. Participants who completed 10 weeks of THR demonstrated significant improvements on measures of Irritability, Lethargy, Stereotypic Behavior, Hyperactivity, expressive language skills, motor skills, and verbal praxis/motor planning skills. When compared to the pre- and post-assessments of participants from the waitlist control condition, the THR group still showed significant improvements in self-regulation behaviors. The THR-specific change from the baseline to post-assessments suggests that the improvements are related to the THR treatment.

© 2011 Elsevier Ltd. All rights reserved.

* Corresponding author at: Dept. Psychiatry and Behavioral Sciences, Children’s Hospital Colorado, 13123 E. 16th Ave, B130, Aurora, CO 80045, United States. Tel.: +1 720 777 3404; fax: +1 720 777 7313.

E-mail addresses: Robin.gabriels@childrenscolorado.org (R.L. Gabriels), John.agnew@colorado.edu (J.A. Agnew), katherineduncan.holt@gmail.com (K.D. Holt), amy@ctrcinc.org (A. Shoffner), Zhaoxing.pan@childrenscolorado.org (P. Zhaoxing), Selga1@aol.com (S. Ruzzano), Jerry.clayton@childrenscolorado.org (G.H. Clayton), Gary_mesibov@med.unc.edu (G. Mesibov).

1750-9467/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rasd.2011.09.007
Although it is known that ASDs are highly heritable (approximately 90%) (see Freitag, Staal, Klauck, Duceticis, & Waltes, 2010 for a literature review) and that biologic and brain irregularities have implications for the ASD behavior patterns and learning styles (Baron-Cohen, 2004; Bauman & Kemper, 2005; Belmonte et al., 2004; Eigsti & Shaprio, 2003; McAlonan et al., 2004; Rubenstein & Merzenich, 2003; Sparks et al., 2002; Waiter et al., 2005), there remain many unanswered questions as to the exact etiology of ASDs and the effectiveness of various interventions. This leads many families of ASD individuals to turn to a variety of treatment options to alleviate symptoms, including those that have little evidence base. Interventions, particularly those that are not well-researched, may be costly and not always successful, sometimes further complicating the ASD child’s behavioral presentation.

Equine assisted activities and therapies (EAAT), including therapeutic horseback riding (THR), represent an intervention approach sought out by the ASD population, but there are various forms of EAAT and few evidence-based published studies to guide clinicians and consumers. The therapeutic intervention strategies of the various forms of EAAT can vary. THR originated in Denmark in 1952 when Madame Lis Hartel, wheelchair dependent due to polio, rehabilitated herself and later won Olympic medals with her horse (King, 2007). In the 1960s, the idea of riding horses for therapeutic purposes spread from Denmark, Norway, and England to the U.S. and Canada. Since then, individuals in the U.S. with a variety of disabilities have used THR to enhance physical, psychosocial, and cognitive functioning. In order to standardize this growing industry, in 1969, the North American Riding for the Handicapped Association (NARHA) was established to develop national accreditation standards to ensure safe and effective practices for riding programs and instructors (Engel, 1997).

Today, NARHA is known as the Professional Association of Therapeutic Horsemanship International (PATH International) and has a network of over 700 member centers throughout the U.S. and Canada (Masini, 2010). The type and breadth of interventions has expanded into a growing field described as EAAT. Under the umbrella of EAAT is a group of therapeutic equine interventions incorporating horses into both activities and the therapeutic process (Freund, Brown, & Huff, 2010; Masini, 2010). Hippotherapy and equine assisted psychotherapy are listed among those interventions categorized as equine assisted therapies, while equine assisted activities include THR (the intervention targeted in this pilot study), interactive vaulting, therapeutic carriage driving, and other horse-related activities such as grooming and stable management (PATH International, 2011). More specifically, THR involves a therapeutic team consisting of volunteer(s), a horse, and a certified therapeutic riding instructor working with individuals who have special needs in a small group setting. The certified therapeutic riding instructor is qualified to teach riding skills related to the pursuit of educational, therapeutic, sport, and leisure goals. In contrast, hippotherapy is led by a medical professional such as an occupational, speech, or physical therapist (Freund et al., 2010). The therapist utilizes the movement of the horse in providing therapy to clients, usually in a one-on-one setting.

The majority of the literature evaluating EAAT has focused on the use of hippotherapy with individuals who have physical disabilities and many of these studies have had small sample sizes and qualitative results. MacKinnon (2007) reviewed publications between 1997 and 2003 and found that a total of 45 articles examining hippotherapy were published during that time period. Of those articles, most were case reports and only 18 were qualitative or quantitative research studies. The population characteristics of those 18 articles included sample sizes from 1 to 40, 66% had child participants, 38% involved participants with medical conditions, 22% involved participants with cerebral palsy, two involved participants diagnosed with autism, and one involved participants with learning disabilities. Only eight of the 18 studies were both quantitative studies and reported statistically significant results indicating improvements in physical strength, tone and endurance, memory, knowledge of the horse, self-confidence (sense of mastery and control), use of gestures to communicate, understanding of safety, eye contact, behavioral control, and response to physical contact. The few autism-specific EAAT studies reported improvements in speech, behavior, sensory modulation, and social-emotional responsiveness.

More recently, there has been a growing interest in quantifying the effects of THR with the ASD population. King (2007) examined the social and related behavioral benefits of 10 weeks (2 h per week) of THR in five children diagnosed with autism. Participants were administered THR in a small group setting led by a PATH International-certified instructor. The THR interventions included sensory activities, tasks, and exercises both on and off the horse. Participants’ social behaviors were evaluated pre- and post-THR intervention by a caregiver and teacher report measure developed by the investigators. Statistical analyses were not conducted due to the small sample size. However, qualitative information about participants revealed improvements in participants’ mood, eye contact, expressive language/conversation, animal care, cooperative social behaviors, and being more open to trying new things.

Bass, Duchowny, and Llabre (2009) published a study of a 12-week THR intervention with children diagnosed with an ASD (n = 19) compared to a waitlist control group (n = 15). Pre- and post-intervention parent report evaluations were conducted within 12 weeks after the conclusion of the conditions (THR or waitlist). The 1-h weekly THR intervention involved teaching: (a) a mounting and dismounting sequence, (b) stretching activities while mounted on the horse, (c) riding skills, (d) social-communication skills games while mounted on then horse, and (e) grooming activities. Results indicated significant improvements in the THR group on the total score and Social Motivation subscale of the Social Responsiveness Scale (SRS). Social motivation behaviors in the subscale include self-confidence, interest in others, and focused behaviors. Additional significant improvements were identified in the THR group on the Sensory Profile total score and subscales of Sensory Seeking, Attention/Distractibility, Sensory Sensitivity, and Sedentary. Compared to other studies in this line of research, this particular study was well-designed and had the largest sample of ASD participants. However, the exclusive use of parent report measures in this study was a limiting factor.
Results from the limited EAAT studies with children who have an ASD suggest that as few as 10 weeks of this intervention can have positive effects on improving mood/emotion regulation, adaptive and social behaviors and awareness, and motor coordination/planning; however, research designs and measurement instruments have had significant limitations. Studies in this area thus far have been limited by the use of caregiver report outcome measures, small sample sizes, a lack of comparative control conditions, and variability of equine intervention methods employed. The current pilot study adds to the sparse EAAT literature, specifically the use of THR with ASD individuals, by including a variety of outcome measures (both objective and caregiver report), a large sample size, and a waitlist control condition. The primary aim of this pilot study was to evaluate the effects of 10 1-h weekly lessons of THR in three core areas of functioning: Self-regulation behaviors, adaptive skills, and motor skills in school-age children and adolescents (ages 6–16 years) diagnosed with either autistic or Asperger’s disorder. The secondary aim of this pilot study was to compare any improvements made in self-regulation behaviors, adaptive skills, and motor skills in the THR group to those made in a group of participants evaluated before and after a 10-week waitlist control condition.

1. Methods

1.1. Participants

Forty-two participants (36 male, 6 female) between the ages of 6 and 16 years (mean age: 8.7 years) and a nonverbal IQ (NVIQ) range of 44–139 (mean: 95.2) diagnosed with either autistic or Asperger’s disorder participated in the THR intervention aspect of this pilot study. Thirty-eight percent of participants’ caregivers reported participants having co-morbid psychiatric conditions in addition to an ASD and 33% reported participants taking psychoactive medications. Sixteen of these participants also participated in a 10-week waitlist control condition before starting the THR intervention. A description of specific participant demographics is outlined in Table 1.

1.2. Screening process

After approval from the Institutional Review Board, participants were recruited from the community via study fliers and a television news story. Parents/legal guardians interested in the study were provided a brief summary of the study over the phone and asked several questions specific to study inclusion and exclusion criteria. Inclusion criteria involved chronological ages 6–16 years, a diagnosis of autistic or Asperger’s disorder from a psychologist or psychiatrist in the community, and a combined score of at least 11 on the Irritability and Stereotypy scales of Aberrant Behavior Checklist-community (ABC-C; Aman, Burrow, & Wolford, 1995; Aman, Singh, Stewart, & Field, 1985). The ABC-C is commonly used in psychopharmacological studies with the ASD pediatric population (Arnold et al., 2000). Exclusion criteria involved having previous exposure to THR within the past three years, having ridden a horse for more than two weeks in the past three years, or having a history of animal abuse or phobia to horses.

Participants who met initial study criteria were then scheduled for two study screening visits; one at the hospital clinic and one at the riding center. During the hospital screening visit, consent and assent forms were obtained from parents/legal guardians and participants. Participants’ DSM-IV diagnosis of autism or Asperger’s was confirmed by a licensed clinical psychologist (first author) using information from the Social Communication Questionnaire (SCQ) (Rutter, Bailey, & Lord, 2003) and the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 1989). Participants’ nonverbal intelligence (NVIQ) was evaluated using the Leiter-R Brief IQ (Roid & Miller, 1997) if participants had not already received a NVIQ assessment within the previous three years and after they had been seven years of age. Participants with standard NVIQ scores less than 40 were excluded due to concerns that such individuals might not meet the riding center criteria for participation in a group lesson. Before participants were screened at the riding center, their physicians gave permission for the child to participate in mounted horseback riding activities. Participants then met with an advanced-level therapeutic riding instructor at the riding center. This screening included a review of participants’ medical history to screen for precautions (e.g., history of seizures) and contraindications (e.g., sensory issues such that the child could not wear a helmet) as outlined by PATH International industry standards. Further, parents/legal guardians were interviewed regarding participants’ physical, cognitive, communication, behavioral, social, and sensory processing functioning and needs. This screening also included assessing participants while mounted on a horse to observe horsemanship skills and physical and cognitive functioning levels in relation to an ability to ride a horse. As a result of this screening session, each participant was

<table>
<thead>
<tr>
<th></th>
<th>All (n = 42)</th>
<th>Waitlist (n = 16)</th>
<th>Non-waitlist (n = 26)</th>
<th>p-Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (range) at enrollment</td>
<td>8.7 (6–16)</td>
<td>8.8 (6–14)</td>
<td>8.6 (5–16)</td>
<td>0.867</td>
</tr>
<tr>
<td>Gender</td>
<td>Male: 36</td>
<td>Male: 15</td>
<td>Male: 21</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td>Female: 6</td>
<td>Female: 1</td>
<td>Female: 5</td>
<td></td>
</tr>
<tr>
<td>Mean nonverbal IQ</td>
<td>95 (44–139)</td>
<td>98 (52–139)</td>
<td>91 (44–119)</td>
<td>0.324</td>
</tr>
<tr>
<td>Seizures</td>
<td>Yes: 2; No: 40</td>
<td>Yes: 0; No: 16</td>
<td>Yes: 2; No: 24</td>
<td>0.517</td>
</tr>
</tbody>
</table>

*p-Values represent t-test for continuous variables and Fisher’s exact test for categorical variables.
assigned to an appropriate THR lesson group level (beginner, intermediate, advanced) or excluded from the study if assessed as unable to ride or approach a horse.

1.3. Measures

All parent/legal guardian report measures were rated by a consistent person for each participant. All measures were used to assess self-regulation, adaptive, and motor functioning within one month of study entry and one month after the THR intervention. The 16 participants that participated in the 10-week waitlist control condition were evaluated within one month prior to the waitlist control period and these evaluations were used as these participants’ baseline measurement for statistical analyses. At the end of the waiting period, these 16 participants were again evaluated and these post-waiting condition evaluations also served as these participants’ pre-THR evaluations, because they occurred within one month prior to starting the THR intervention. All motor skill measures were administered by a registered occupational therapist.

1.3.1. Self-regulation

The Aberrant Behavior Checklist-community (ABC-C, Aman et al., 1985) is a 58-item symptom checklist for assessing problem behaviors of children and adults with developmental disabilities in community settings. The items resolve into five scales: Irritability/Agitation, Lethargy/Social Withdrawal, Stereotypic Behavior, Hyperactivity, and Inappropriate Speech. The ABC-C has been validated and normed for raters including parents, special educators, psychologists, direct caregivers, nurses, and others with knowledge of the person being assessed (Aman et al., 1985). Extensive psychometric assessment of the ABC-C indicates that its subscales have high internal consistency, adequate reliability, and established validity (Aman et al., 1985).

1.3.2. Adaptive skills

The Vineland Adaptive Behavior Scales – Interview Edition, Survey Form (VABS-II, Sparrow, Cicchetti, & Balla, 2005) is a clinician interview with the parent/legal guardian and measures child adaptive functioning, providing raw and standard scores in the domains of Communication, Daily Living Skills, and Socialization Skills. The VABS-II has good to excellent split-half and test–retest reliability and modest concurrent validity (Sparrow et al., 2005). For the current study, administration and scoring reliability on the VABS-II was achieved among three graduate student research assistants, achieving at least 80% scoring reliability on three consecutive VABS-II interview assessments. Additionally, every effort was made to ensure that each participant was evaluated by the same person for both the baseline and post-condition assessments.

1.3.3. Motor skills

The Bruininks-Oseretsky Test of Motor Proficiency (BOT-2, Short Form, Bruininks & Bruininks, 2005) assesses the motor proficiency of able-bodied patients, as well as patients with serious motor dysfunction and developmental handicaps. The Bruininks-Oseretsky Test provides a separate measure of gross and fine motor skills, making it possible to obtain meaningful comparisons of performance in two areas. The Complete Battery includes 46 items. However, for this study, short form subtests were used: Fine Motor Precision, Fine Motor Integration, Manual Dexterity, Balance, Running Speed and Agility, Upper-Limb Coordination and Strength. The psychometric assessment of the BOT-2 indicates that its subscales have high internal consistency and test–retest reliability, very consistent inter-rater reliability, and established validity (Bruininks & Bruininks, 2005). For the current study, administration and scoring reliability on the BOT-2 was achieved among three occupational therapists, achieving at least 80% scoring reliability on three consecutive BOT-2 assessments. Additionally, every effort was made to ensure that each participant was evaluated by the same person for both the baseline and post-condition assessments.

Sensory Integration and Praxis Test (SIPT, Ayres, 1989) is made up of 17 tests that measure sensory processes and praxis (the ability to organize, plan, and perform an action). For the aims of this study, only two subtests (Praxis on Verbal Command and Postural Praxis) were used. Test–retest reliability is rated as good for the praxis subtests. For this study, administration and scoring reliability on the SIPT was achieved among the same three occupational therapists who also administered the BOT-2, achieving at least 80% scoring reliability on three consecutive SIPT assessments. Again, every effort was made to ensure that each participant was evaluated by the same person for both the baseline and post-condition assessments.

1.4. Procedure

All participants received a battery of baseline and post-condition evaluations previously described in the Measures section. A final subset (n = 16) of participants entered into the study were evaluated before the start of a 10-week waiting period, during which time they continued their normal activities and did not have interactions with study personnel or study horses. This waitlist group of 16 participants did not differ significantly from the total study population in terms of age, gender, NVIQ, presence of seizures or use of psychoactive medications. These 16 participants did differ from the total study population in that fewer members of this subgroup had co-morbid psychiatric diagnoses.

The baseline and post-evaluation measures included the Aberrant Behavior Checklist-community (ABC-C) (Aman et al., 1995), the Short Form of the Bruininks-Oseretsky Test of Motor Proficiency (BOT-2) (Bruininks & Bruininks, 2005), the Verbal...
Praxis and Postural Praxis subscales of the Sensory Integration and Praxis Test (SIPT) (Ayres, 1989) and the Vineland Adaptive Behavioral Scales – II Interview Edition, Survey Form (VABS-II) (Sparrow et al., 2005). Participants traveled to the riding center for 10 consecutive weeks of 1-h THR lessons. Participants who missed more than two lessons were withdrawn from the study. During each weekly THR lesson, parents/legal guardians completed the ABC-C questionnaire to describe their child’s behavior during the preceding week. At that same time, parents/legal guardians documented any changes during that week in their child’s medications or interventions.

1.5. THR lessons

The THR intervention took place at a long-standing (30 years) Premier PATH International Accredited Riding Center that has both an outdoor arena and a heated indoor arena. Premiere PATH International centers have to comply with standards and site visits every five years to assess adherence to those standards in areas including: safety (e.g., rider safety, horse selection and training, risk management, volunteer screening and training, and creating distraction-free and safe environments for horses, riders, and their families); ethics (e.g., horse care, record keeping, confidentiality practices); and effectiveness (e.g., instructor certification and continuing education). PATH International also provides three levels of certification for therapeutic riding instructors. The advanced level certification involves demonstrated competency in the areas of equine management, horsemanship, riding instruction, teaching methodology and knowledge of disabilities.

The THR lessons were taught in small groups of three to four participants and were led by a PATH International Advanced Instructor with trained volunteers present to work with each participant and the horse. Every effort was made to ensure that the horses and volunteers assigned to each participant remained the same throughout the 10-week THR lessons to facilitate relationship-building and maintain consistency. Each participant had one to three volunteers and the number of volunteers assigned to each rider depended upon the participant’s level of functioning demonstrated during the initial screening assessment. All volunteers participated in the center’s required 3-h training and volunteers had access to the day’s lesson plan before participating in the THR session.

Each THR lesson had a two-part focus: therapeutic and horsemanship. Instructors set individual therapeutic goals and horsemanship goals based on the initial evaluation of each participant’s developmental ability level and horsemanship skill level. The lesson plans consisted of activities and exercises that addressed physical, psychological, cognitive, and social skills as well as horsemanship skills. Every THR lesson followed the same routine: (a) put riding helmet on, (b) sit and wait on the bench, (c) mount horse, (d) THR activities, (e) dismount horse, (f) groom horse, and (g) put away equipment. This routine was presented to all participants at the start of each lesson in the form of a picture schedule. THR lessons lasted 1 h and participants rode their horse for a minimum of 45 min during that hour. The mounted portion of each lesson consisted of mounting, warm up, review of skills from the previous week, teaching a new skill, a game or activity to practice the new skill, a lesson review followed by a cool down period and dismount. After riding, participants led their horse to the tacking area, learned skills of un-tacking and grooming, thanked their horse and volunteers, and put their equipment away.

1.6. Data analysis

Two-sample Student’s t-test and Fisher’s exact test were used to compare the waitlist control group participants with the THR intervention group participants for continuous and categorical demographic and clinical variables. Paired t-tests were used to assess the change between the baseline and the post-THR intervention evaluations. Analysis of covariance (ANCOVA) was used to compare the post-evaluations from the waitlist control group \( (n = 16) \) to the post-evaluations from the entire THR intervention group \( (n = 42) \). In this ANCOVA, participants’ baseline evaluation scores, age at enrollment, and NVIQ were adjusted as covariates. In order to analyze the time course of weekly changes of the ABC-C scores during the THR intervention period, a linear mixed effects model with a compound symmetry covariance structure was used to perform the repeated measures analysis of variance.

2. Results

2.1. Demographic analyses

Table 1 lists participants’ baseline demographic and clinical characteristics including age, gender, co-morbid psychiatric diagnoses, psychoactive medications, and NVIQ. Table 2 provides average scores at study entry (baseline) and study completion (post-THR) for all outcome measures reported here and also provides participant numbers and \( p \)-values for all outcome measures. Participant numbers are fewer than 42 for some measures because not all participants were able to complete all measures.

2.2. Self-regulation behaviors

Comparing baseline THR evaluations to those conducted post-THR intervention, participants demonstrated significant improvements on the ABC-C Irritability \( (p < 0.001) \), Lethargy \( (p < 0.001) \), Stereotypic Behavior \( (p < 0.001) \), Hyperactivity \( (p < 0.001) \) and Inappropriate Speech \( (p = 0.05) \) scales, as shown in Fig. 1.
From repeated measures ANOVA analysis using a linear mixed-effects model, analysis of the weekly parent/legal guardian reports for the ABC-C demonstrated that the behaviors measured improved over the course of the THR intervention. For the Irritability, Lethargy, Stereotypic Behavior and Hypersensitivity scales, statistically significant improvement was noticed as early as the third week of THR (Fig. 2). Measures at every subsequent week were significantly different from baseline. The improvement reached the maximum level around week five and stabilized over the rest of the THR intervention period.

2.2.1. Adaptive skills

The VABS-II adaptive composite standard score was used for the analysis, as were the raw scores from the Social, Communication and Daily Living domains. Raw scores were used because previous research has suggested that raw scores better reflect the incremental adaptive skill changes of individual with ASDs than do the scaled scores (Gabriels, Ivers, Hill, Agnew, & McNeill, 2007). Comparing baseline and post-THR assessments for the VABS-II, significant improvements were observed for the Adaptive Total score (p = 0.001), the Communication raw score (p = 0.035), the Social raw score (p = 0.016) and the Daily Living raw score (p = 0.011). Based on the significant improvement observed on the VABS Communication raw score, it was thought to be important to examine the subscales of this measure. The VABS-II Communication domain assesses both expressive and receptive language abilities. These domain raw scores were examined using paired t-tests and it was found that participants displayed significant improvement on the measure of expressive language (p = 0.005), but there was only a trend toward improvement on the measure of receptive language (p = 0.061, n.s.).

2.2.2. Motor skills

Participants demonstrated improvements between the baseline and post-THR intervention group evaluations on the Short Form of the BOT-2 (p < 0.001) and SIPT Verbal Praxis (p < 0.001) and SIPT Postural Praxis (p = 0.009) measures. Raw scores on these measures were used in the analyses.

2.2.3. Comparison to waitlist control group

The presence of the 16-participant waitlist control group enabled further analysis of the data to determine if the improvements previously noted were due specifically to the THR intervention or if they might be due to a different, unidentified factor. To address this question, ANCOVA analyses were performed using baseline scores on the assessment Table 2

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>N</th>
<th>Baseline (mean ± SD)</th>
<th>Post-THR (mean ± SD)</th>
<th>Difference (mean ± SD)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC-C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irritability</td>
<td>42</td>
<td>20.2 ± 8.9</td>
<td>12.9 ± 8.5</td>
<td>-7.3 ± 8.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lethargy</td>
<td>42</td>
<td>12.4 ± 7.7</td>
<td>6.3 ± 7.1</td>
<td>-6 ± 6.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stereotypy</td>
<td>42</td>
<td>6 ± 4.2</td>
<td>3.3 ± 3.5</td>
<td>-2.7 ± 3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>42</td>
<td>23.7 ± 9.9</td>
<td>17.1 ± 11.6</td>
<td>-6.6 ± 8.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Inappropriate speech</td>
<td>42</td>
<td>4.3 ± 2.6</td>
<td>3.5 ± 3.1</td>
<td>-0.8 ± 2.7</td>
<td>0.055</td>
</tr>
<tr>
<td>Adaptive skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw social score</td>
<td>40</td>
<td>104.9 ± 29.9</td>
<td>113.2 ± 27.4</td>
<td>8.3 ± 20.9</td>
<td>0.016</td>
</tr>
<tr>
<td>Raw communication score</td>
<td>40</td>
<td>143.6 ± 24.9</td>
<td>149 ± 24.8</td>
<td>5.4 ± 15.6</td>
<td>0.035</td>
</tr>
<tr>
<td>Raw daily score</td>
<td>40</td>
<td>110.6 ± 35.1</td>
<td>117.4 ± 32.6</td>
<td>6.8 ± 15.9</td>
<td>0.011</td>
</tr>
<tr>
<td>Adaptive total score</td>
<td>40</td>
<td>75.5 ± 10.4</td>
<td>79.2 ± 11.3</td>
<td>3.7 ± 5.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Motor skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOT 2</td>
<td>38</td>
<td>45.5 ± 15.5</td>
<td>53.4 ± 15.2</td>
<td>8 ± 9.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SIPT: verbal score</td>
<td>38</td>
<td>16 ± 7.2</td>
<td>18.8 ± 7</td>
<td>2.8 ± 3.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SIPT: postural score</td>
<td>38</td>
<td>19.5 ± 7.4</td>
<td>22.9 ± 7.1</td>
<td>3.4 ± 7.5</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Fig. 1. Mean scores of ABC subscales depicting the mean change from baseline to post-THR for all participants.
Fig. 2. Data collected weekly from parents using the ABC-C. Data presented are the mean ± 2 SE.
measures, participants’ age, and NVIQs as covariates of interest. This analysis then compared the change from baseline to post-evaluations for the two conditions (waitlist and THR intervention) for all of the assessment measures. This analysis found that four of the changes were significant: Irritability ($p = 0.004$), Lethargy ($p = 0.007$), Stereotypic Behavior ($p = 0.020$) and Hyperactivity ($p = 0.008$) scales of the ABC-C. The VABS-II Communication raw score showed a trend toward statistical significance ($p = 0.090$, n.s.). See Table 3 for the least squares analysis of these variables. All other measures showed no significant differences on this ANCOVA analysis. The presence of a significant difference on the ANCOVA analysis suggests that the degree of change from the baseline evaluations to the post-evaluations differs significantly between the waitlist and the THR conditions and, therefore, may be due to the treatment itself. Lack of significant results for certain measures on this ANCOVA analysis could mean that any differences observed on the $t$-tests described previously may be due to developmental changes, lack of statistical power to identify a significant difference using the ANCOVA analysis or other factors that cause fluctuations between testing sessions, such as poor test–retest reliability. For the published and standardized measures used in this pilot study, test–retest reliability is generally high (Aman, Singh, & Turbott, 1987; Ayres, 1989; Bruininks & Bruininks, 2005; Sparrow et al., 2005), so this last factor is likely not the cause of the lack of significant results on the ANCOVA analysis. Significant results from this analysis are presented in Fig. 3.

### Identification of subpopulations

In an attempt to determine if there was a specific subgroup of age or NVIQ that responded particularly well or particularly poorly to this 10-week THR intervention, data were analyzed using a linear calibration model. Unfortunately, due to the relatively low number of participants and the relatively high variability in their assessment scores, no meaningful results were obtained.

### Table 3

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Least square means adjusted for age, IQ and baseline value</th>
<th>Difference</th>
<th>DF</th>
<th>$t$-Value</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irritability</td>
<td>18.8</td>
<td>12.5</td>
<td>6.4</td>
<td>49</td>
<td>3.00</td>
</tr>
<tr>
<td>Lethargy</td>
<td>11.6</td>
<td>6.6</td>
<td>5.0</td>
<td>49</td>
<td>2.81</td>
</tr>
<tr>
<td>Stereotypic movement</td>
<td>5.6</td>
<td>3.7</td>
<td>1.8</td>
<td>49</td>
<td>2.40</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>24.0</td>
<td>16.9</td>
<td>7.1</td>
<td>49</td>
<td>2.78</td>
</tr>
<tr>
<td>VABS communication domain</td>
<td>142</td>
<td>150</td>
<td>−7</td>
<td>48</td>
<td>−1.75</td>
</tr>
</tbody>
</table>

---

Fig. 3. ABC-C subscales with significant results on ANCOVA analysis. Open markers represent data from the THR treatment and filled markers represent data from the wait-list treatment.
3. Discussion

This pilot study provides preliminary evidence that a 10-week therapeutic horseback riding (THR) intervention with children diagnosed with an ASD can result in significant improvement. Specifically, participants in the THR intervention group made significant improvements from baseline to post-evaluations on measures of self-regulation (Irritability, Lethargy, Stereotypic Behavior, and Hyperactivity), adaptive expressive language skills, motor skills, and verbal praxis/motor planning skills. The THR group’s improvements on the Irritability, Lethargy, Stereotypic Behavior, and Hyperactivity subscales of the ABC-C were significant when comparisons were made to the same baseline to post-evaluations of the waitlist control group. This study did not indicate that there was a specific profile (e.g., younger or older participants or higher or lower intellectual ability) that best predicted the significant improvements made as a result of the 10-week THR intervention.

Self-regulation problems, including irritability, hyperactivity, and lethargy are well-documented behavior problem areas in the ASD literature and are often target symptoms for medication intervention trials (Arnold et al., 2003). Symptoms such as irritability and hyperactivity are critical behaviors that can impact the ASD child’s ability to function successfully in home and school environments. Horses may help organize or provide input to the ASD child’s sensory system. This factor may contribute to helping the child feel calm. Stoner (2007) posits that a horse’s muscles generate warmth while moving, which can calm and relax the rider. However, what remains unclear is how or why this effect may continue after the child is no longer riding the horse. The following comments made by the parents/legal guardians during the study exit interview are reflective of the self-regulation improvements measured by this study.

“The day of riding my child was quiet, calm, and more peaceful.”
“The day of riding we see very different behaviors. Very calm. This morning he struggled getting into the building, but once on the horse he was fine.”
“He has been very happy, calm, and has had less school anxiety. He has been in a great mood”
“When my child returned to school after riding he was more focused according to his teacher.”
“She doesn’t act like popcorn as much (jumping out of bed) and has much fewer meltdowns”

The increased expressive communication behaviors observed in this study were unexpected in this pilot study, and may have been influenced by the human–horse interaction engaging and motivating experience inherent in the THR intervention. For example, if the child says “Walk on”, the horse responds. Volunteer handlers, who were present during the THR intervention to ensure the safety of the horse and rider, added to the social-communication experience of the THR intervention. In this study, the THR instructors routinely encouraged participants to verbalize instructions to their horses, which might help explain the observed improvements in expressive language. While ASD is characterized by a lack of human social understanding, horses are highly social animals who will respond to subtle human cues (Grandin, 1997). Such responsiveness is important for the cause-and-effect concrete learning styles of individuals with an ASD. A horse’s immediate response to the behaviors (however subtle) of an ASD child can be used in treatment to help the child better understand or become more aware of the impact of his or her social-communication behavior. The following comments made by the parents/legal guardians during the study exit interview are reflective of the adaptive communication improvements measured by this study.

“She seems to be asking more instead of just doing.”
“Overall, he has been more conversational. He has been sharing and compromising more” “Better with turn-taking, decreased anger, and better expressing himself”
“She seems to be using more words and is a little calmer”
“Decreased negativity. His teacher reported that he paid her a compliment on how nice she looked”

Motor skill deficits are often reported in the ASD literature (De Giacomo et al., 2009; Dowell, Mahone, & Mostofsky, 2009) and these deficits may contribute to the frequently reported ASD problems of communication (e.g., oral motor skills) (Gernsbacher, Sauer, Geye, Schweigert, & Hill Goldsmith, 2008) and imitation/praxis (Vanvuchelen, Roeyers, & De Weerdt, 2007). Horseback riding may help improve motor coordination, organization, and planning (sequencing ability, coordination, and multitasking) skills in children with an ASD. For example, the continuous adjustments to the horse’s movements during riding, “… involves the (rider’s) use of muscles and joints, leading to increased muscle strength, tone, bilateral control, balance, and a range of motion” (King, 2007, p. 122). In addition, when a rider is mounted on a horse, balance is critical. The gentle movements of the horse demand that a rider constantly adjusts his or her weight to stay upright. The occupational therapy literature has discussed the idea that for the rider, the horse provided “perceptual sensory excitement through all the receptors in the joints, muscles, tendons, ocular motor systems, and skin” … and that “this perpetual stimulation provided an instant feedback that constantly adjusted the movement systems, thus contributing to the maintenance of posture and allowing the client to become a participant involved in purposeful activity” (Engel & MacKinnon, 2007, p. xxi).

Although this study included multiple evaluators of outcome variables, the study results are limited by rater bias because raters were not blinded to participants’ intervention condition and by the use of a single parent/legal guardian report for the ABC-C self-regulation measure. An additional limitation is the lack of a more powerful randomized controlled design and a standardized THR intervention protocol. The preliminary nature of this investigation also leaves a number of questions
unanswered that could affect conclusions drawn from these data. Results reported above have not been corrected for multiple comparisons using a Bonferroni or similar correction. Because this was a pilot study, the primary outcome measure was the ABC-C, but there were several exploratory measures included. Due to the preliminary nature of this study, it was not clear at the time of design how many measures one should correct for, and so it was decided to not apply any corrections. The reader, therefore, is advised to interpret the statistical results accordingly. It is conceivable that the documented effects may not only be induced by THR therapy, but also may be the result of regular intensive interactions in a small group activity setting or that merely interacting with animals in a regular controlled environment could result in similar behavioral improvements. It is also conceivable that the modality that truly elicits improvement in behavior is the mechanical movement (i.e., sensory feedback) that occurs during riding. Of further note, this study does not address the long-term therapeutic effects of THR interventions. These questions require further study in order to fully evaluate the specificity and lasting effects of THR interventions. The authors of this paper are currently conducting a randomized control study to investigate changes in self-regulation behaviors, motor skills, language ability, and social skills that includes a larger sample size, more extensive outcome measures, and raters blinded to intervention condition.

Determining how the basis of the human–animal interaction via THR is helpful to individuals with an ASD has far-reaching implications for the quality of life in this population. Further examination of the impact of THR on skills such as communication and social deficits in children with an ASD is warranted. Other future study considerations include using a standardized THR intervention protocol and a fidelity measurement tool. A clearly defined and empirically supported THR curriculum manual could inform the over 700 PATH International member centers and other providers who work with ASD children. Such a curriculum would sharpen the delivery of THR, allowing THR programs to intervene in a manner that could elicit significant and, perhaps, long-term change for this ASD population.

This study is one of the few to quantify and systematically examine the benefits of THR on children with an ASD and helps provide validity to the anecdotal reports that THR can be effective for the ASD population. However, there remains little understanding of the importance of the human–animal function or its effectiveness. Dr. Temple Grandin, an adult with autism and Professor of Animal Science at Colorado State University, observes that animals are visual-based thinkers, a common learning style exhibited by ASD individuals, and suggests that people with autism are not unsocial, but rather that they relate to other people in ways more similar to the ways that animals do (Grandin & Johnson, 2005). Grandin’s assertions suggest that there may be something unique about animal interactions or assisted therapies with the ASD population that is worthy of further scientific exploration.

Acknowledgements

The authors would like to acknowledge Greenspring Christian Church for funding this study through The Children’s Hospital Colorado Research Institute. The authors would also like to thank the families and children who participated in this study as well as our colleagues who assisted with this project: Heather Bosler, Laurie Burnside, Briar Dechant, Amanda Drazin, Paul Fennessey, Anne Hanson, Rebecca Howard, Jeff Magouirk, Syd Martin, Julia Runde, Kate Trujillo, Sharen Trunnel, Marianne Wamboldt, and the staff of the Colorado Therapeutic Riding Center, particularly Carol Heiden, Jody Howard, Heather McLaughlin, Mary Mitten, Penelope Powell and Sharon Van Boven. Portions of the research described in this study have been previously presented at the International Meeting for Autism Research in May 2009 (Chicago, IL, USA) and 2010 (Philadelphia, PA, USA), at the Annual Spring Pediatrics Poster Session, Children’s Hospital Colorado in May 2009 (Denver, CO, USA), at the Graduate School of Professional Psychology Research Symposium University of Denver in May 2009 (Denver, CO, USA), at the PATH International Region 7 Pre-Conference Workshop in April 2010 (Baldwin City, KS, USA), at the University of Colorado Coleman Institute conference in October 2010 (Denver, CO, USA), and at the PATH International National Conference and Annual Meeting in November 2010 (Denver, CO, USA).

References


