2015

Hippotherapy and gross motor function in children with spastic cerebral palsy: A critical analysis of the available literature

Joe Oelfke
Pacific University

Hippotherapy and gross motor function in children with spastic cerebral palsy: A critical analysis of the available literature

CRITICALLY APPRAISED TOPIC

Title: Hippotherapy and gross motor function in children with spastic cerebral palsy: A critical analysis of the available literature

Clinical Scenario: A seven-year old male presents with a primary diagnosis of spastic quadriplegic cerebral palsy (CP) while attending a school for children with disabilities in the township of Gugulethu, Cape Town, South Africa. Impairments include decreased trunk and pelvic control, decreased lower extremity strength, increased lower and upper extremity tone, an uncoordinated flexed gait pattern, and a lack of independent ambulation. Prior to the introduction of hippotherapy, the patient was non-ambulatory. Medical treatment to date has involved general trunk and lower extremity strengthening exercises, wheelchair positioning, speech and occupational therapy for speech and swallowing problems, and gait training. Hippotherapy is suggested to be an effective intervention for children with cerebral palsy to assist in neuromuscular re-education of postural muscles and decreasing postural tone and is currently being implemented as a routine intervention in the care of this patient.

Brief introduction: CP is a non-progressive neurological disorder that results from a lesion to the developing cerebrum, either in utero, during the birthing process, or during the first years of life. There are two primary ways of classifying CP: region of the body that is affected and the type of CP. Affected areas include diplegia (upper extremities are affected to a greater extent than the lower extremities), hemiplegia (one side of the body is affected, involving both the upper and lower extremity), and quadriplegia (all four limbs and trunk are affected). Types of CP include spastic, dystonic, athetoid, ataxic, and mixed[18]. As the patient was diagnosed with spastic CP, the analysis will focus on this type. Spastic CP typically presents with increased tone in affected areas of the body with slow movements and resistance to movement. Management of CP varies depending on involvement of the disorder but can include stretching and strengthening of affected body parts, neuromuscular re-education, gait and balance training, botulinum toxin injections to hypertonic muscles, baclofen therapy, muscular release surgeries, and/or a selective dorsal rhizotomy procedure[3].

Clinically, CP is classified according to the Gross Motor Function Classification System (GMCFS). The GMFCS is a five-level classification system that describes children and youth with cerebral palsy according to self-initiated movements with an emphasis on sitting, walking, and wheeled mobility. Distinctions between levels are based on functional ability, the need for assistive technology, and quality of movement. The original version, published in 1997 by Palisano *et al.* describes children up to 24 months of age. An expanded version, the GMFCS E&R, was published and validated in 2008 by Palisano *et al.* to describe children and youth up to 18 years. Table 1 highlights the general distinctions between levels of the GMFCS[17].

Table 1. General distinctions between levels of the GMFCS & GMFCS E&R[17].

Level	Description
I	Walks without limitations
II	Walks with limitations
III	Walks using a hand-held mobility device
IV	Self-mobility with limitations; may use powered mobility
V	Transported in manual wheelchair

Hippotherapy and therapeutic horseback riding are the terms used to describe treatment strategies that utilize the movement of horse to improve postural control, balance, and general function and mobility. Hippotherapy is thought to be effective in managing patients with neurological disorders such as cerebral palsy because the pelvic movements of the horse are very similar, in biomechanical terms, to that of a human when walking[4]. By seating a patient on the horse's back, the movement of the horse, including tempo, gait, and cadence, are thought to have a carry-over effect to the patient to help facilitate neuromuscular re-education to improve sitting balance and overall gross motor function. This occurs because these movements encourage bilateral postural muscle activation to react to the horse's pelvic sway[10]. The body heat from the horse is also thought to improve blood circulation, which reduces abnormal muscle tone to aid in sitting balance[14].

While hippotherapy is growing in popularity as a treatment option for children with motor dysfunction secondary to CP, there is some controversy regarding the required time and costs as compared to other techniques[11,12,13,15], such as whether a hippotherapy simulator can provide the same benefits as hippotherapy itself[1,5,11,]. The primary controversy lies in the often misinterpreted distinction between hippotherapy and therapeutic horseback riding (THR). THR refers to the use of horseback riding to teach equestrian skills to and provide recreational activities for individuals with disabilities, whereas hippotherapy refers to the implementation of specific, targeted exercises performed while riding the horse[9]. Proper hippotherapy requires additional training regarding human movement and physiology to implement exercises that address impairments, functional limitations, or participation restrictions of these individuals[9]. For the purposes of my clinical question, I want to critically appraise the highest quality available research regarding the efficacy of hippotherapy in improving gross motor function and balance in children with cerebral palsy.

My Clinical question: Does physical therapy (PT) treatment involving hippotherapy lead to improved gross motor function and balance in children with cerebral palsy?

Clinical Question PICO:

Population – Children, ages 6-18 years, with spastic cerebral palsy (quadriplegia or diplegia)

Intervention - Hippotherapy

Comparison – Trunk and extremity strengthening and stretching exercises

Outcome – Gross Motor Functional Measure (GMFM)

Overall Clinical Bottom Line: Based on the results described by the authors identified in Table 2, there is low quality evidence supporting the use of hippotherapy as a gross motor function intervention for children with a primary diagnosis of cerebral palsy. Two studies were randomized-controlled trials while the other two utilized an A-B and A-B-A study design with moderate to poor internal validity. The large number of major threats to internal validity in all four articles dramatically decrease confidence in and ability to generalize findings from the study results. This suggests the need for more rigorous research with strict protocols and study design and larger, heterogeneous samples to

reach a definitive decision regarding the effectiveness of hippotherapy compared to strengthening and stretching exercises for the trunk and extremities in improving GMFM score as a measure of postural control.

Search Terms: Hippotherapy, dynamic postural exercise, equine-assisted therapy, physical therapy, therapeutic horseback riding, cerebral palsy

Appraised By: Joe Oelfke, SPT
School of Physical Therapy
College of Health Professions
Pacific University
Hillsboro, OR 97123
oelf7736@pacificu.edu

Rationale for my chosen articles:
Upon searching the PubMed database using the previously listed search terms, 43 articles were identified. Of these, four were selected that assessed the efficacy of hippotherapy for improving gross motor control in children with varying levels of involvement of spastic cerebral palsy. This selection was based primarily on how well the individual articles matched the clinical PICO based on my patient, followed by the quality of research design as measured by the PEDro score. Due to the lack of literature on this topic, three randomized-controlled trials and one quasiexperimental design trial were selected for this literature critique. These articles were selected based on their PEDro scores and most closely match the clinical PICO in terms of age and diagnosis for the children and the outcome measures employed. However, only Kang *et al* .article assessed hippotherapy in conjunction with traditional physical therapy as compared to physical therapy alone and a control. McGibbon *et al.* and Casady *et al.* utilized a repeated measures design. These two articles were selected as they provide evidence regarding the benefit of hippotherapy over a prolonged period of time. Finally, Herrero *et al.* assessed the efficacy of a hippotherapy simulator compared to barrel-sitting and the effect of each on gross motor function. Table 1 provides a brief summary of the clinical PICO and PEDro scores and Table 2 provides a comparison of the PEDro scores.

Table 2. Brief summary of the PICO and PEDro scores of appraised articles.

Author	PEDro Score	Population	Intervention	Comparison	Outcome Measures
Herrero *et al.* (2012)	8/10	38 children with CP	Hippotherapy simulator	Barrel sitting	GMFM & SAS
McGibbon *et al.* (2009)	7/10*	6 children with spastic CP	Hippotherapy	None	GMFM & Self-perception profiles
Kang *et al.* (2012)	5/10	45 children with CP	Hippotherapy	Hippotherapy + PT and control	Sitting balance test
Casady *et al.* (2004)	3/10	10 children with CP	Hippotherapy	None	GMFM & PEDI

Note. Abbreviations: CP (cerebral palsy), PT (physical therapy), GMFM (Gross Motor Function Measure), SAS (Sitting Assessment Scale), PEDI (Pediatric Evaluation of Disability Index).
*PEDro score is not appropriate for entirety of study. More details are provided in the individual article appraisal.

Table 3. Comparison of PEDro Scores

	Herrero, *et al.*	McGibbon, *et al.*	Kang, *et al.*	Casady, *et al.*
Random	Y	Y	Y	N
Concealed allocation	Y	Y	N	N
Baseline comparability	Y	Y	Y	Y
Blind Subjects	N	N	N	N
Blind Therapists	N	N	N	N
Blind Assessors	Y	Y	N	N
Adequate Follow-up	Y	Y	Y	N
Intention-to-Treat	Y	N	N	Y
Between Group	Y	Y	Y	N
Point Estimates & Variability	Y	Y	Y	Y
Total Score	8/10	7/10*	5/10	3/10

*PEDro score is not appropriate for entirety of study. More details are provided in the individual article appraisal.

Article: Herrero P, Gomez-Trullen EM, Asensio A, *et al.* A. Study of the therapeutic effects of a hippotherapy simulator in children with cerebral palsy: A stratified single-blind randomized controlled trial. *Clinical Rehabilitation,* 2012; 1105-1113.

Clinical Bottom Line: Based on the results of this article, treatment with a hippotherapy simulator does not appear to be an effective treatment for improving GMFM-66 score in children age 4-18 years with cerebral palsy. The treatment group received an individualized exercise program while sitting on a hippotherapy simulator for 15 minute treatment sessions, once per week for ten weeks. The control group received the same treatment protocol, but with the simulator switched off. The experimental group demonstrated a medium treatment effect for the SAS score [effect size = 0.59 (95% CI: -0.92 to 0.26)], but the 95% confidence interval crosses zero, suggesting that the true mean may actually favor the control group. In addition, the treatment group demonstrated a small treatment effect for the GMFM-66 score [effect size = 0.25 (95% CI: -0.10 to 0.60)], with the 95% confidence interval again crossing zero. When assessing dichotomous data, a number needed to treat of 9.5 (95% CI: 2.4 to -4.8) was calculated and when analyzing only the GMFCS level V patients, a number needed to harm of -2.81 (95% CI: 18.9 to -1.3) was calculated, indicating that the control group is more likely to improve; however, the confidence interval extends into infinity, thus creating a meaningless number. One minor and three major threats to internal validity were identified, dramatically decreasing confidence in the study results. While hippotherapy simulator treatment is a relatively low-cost, timely treatment option, poor internal and external validity of these results and feasibility issues suggest further research is necessary to reach a decision regarding the clinical PICO.

Article PICO:

> **Population—** 38 children, ages 4-18 years, with cerebral palsy
>
> **Intervention—** Hippotherapy simulator for 15 minute sessions, provided once per week for 10 weeks
>
> **Comparison—** Hippotherapy simulator turned off for 15 minute sessions, provided once per week for 10 weeks
>
> **Outcomes—** GMFM and the Sitting Assessment Scale (SAS)

Blinding: The authors state that this was a single-blind study in which the assessor was blinded to group allocation. Blinding of the subjects and therapists could not occur do to the nature of the study as it was readily apparent if the simulator was turned on or off. However, this is a minor threat to internal validity because, due to the children's age and level of involvement of CP, they likely would not be capable of manipulating their performance to influence the outcomes of the study.

Controls: The control group received the same 15 minute treatment sessions as the intervention group, once per week for 10 weeks, while sitting on the hippotherapy simulator; however, the simulator remained turned off for the duration of the treatment. As with the intervention group, the therapists also prescribed individualized therapeutic exercise during these sessions. This was an appropriate comparison group as the changes seen at completion of the study can most likely

be attributed to the rhythmic movements created by the simulator when turned on. However, as the therapists instructed the children to perform different motor activities according to their motor possibilities, there is the possibility that the individualized nature of the exercises could mean that observed differences between the groups may be inappropriately attributed to the intervention when in actuality they were due to an individualized treatment program. While there was an appropriate control group in this study, the methodology of treatment administration potentially threatens internal validity.

Randomization: Assignment of subjects to groups was randomized and was stratified. The stratification was performed according to the GMFCS level, which was appropriate as it ensured a similar number of subjects were allocated to each group, at each GMFCS level. The randomization was also concealed and successful with no statistically significant differences at baseline (see Table 4); however the sample was biased toward a higher level of disability of 50% of the children classified as GMFCS level V. This can potentially impact interpretation of the results as the sample is not normally distributed. As a result, the sample sizes for each GMFCS level are too small to carry out a meaningful analysis and only the children with the highest level of disability can be appropriately analyzed.

Table 4. Baseline comparability of groups, assuming an alpha level of 0.05 to determine significance.

Demographic & Outcome Measures	Control Group	Intervention Group	P-Value
Mean Age	9.05	9.95	0.504
Gender	10:9	14:5	0.179
GMFCS Level I	2	2	0.834
GMFCS Level II	2	1	-
GMFCS Level III	3	2	-
GMFCS Level IV	3	4	-
GMFCS Level V	9	10	-
Total GMFM Score	42.75	40.91	0.758
GMFM Dimension B	29.84	25.68	0.405
SAS Score	15.58	15.21	0.848

Note. Table reflects the author's reported data.

Study: This study was a single-blinded, stratified randomized-controlled trial with 19 subjects in both the control and intervention groups. The inclusion and exclusion criteria used were from a previously published study[5]. Subjects were included if they were between 4 and 18 years old with a primary diagnosis of CP. Subjects were excluded if they had undergone a selective dorsal rhizotomy, had active convulsions not controlled by medication, were allergic to electrode adhesives, had a visual impairment not corrected with glasses, had an associated illness or circumstance that may interfere with the results or be detrimental to treatment, or were unable to attend intervention sessions and/or refused to participate.

After baseline measurements were taken by an independent assessor and sociodemographic information was obtained with a questionnaire, subjects were stratified into GMFCS levels as described above and then randomized into treatment and control groups. During the treatment

sessions, the children were placed on the hippotherapy simulator and were instructed to maintain the sitting position for 15 minutes. This included active extension of the trunk, stabilization of the pelvis, and abduction of the hips while the simulator produced a rhythmic and repetitive movement similar to the walking of a horse. During the 15 minutes sitting on the simulator, the therapist instructed the children to perform various activities that were individualized according to their motor possibilities. Treatment sessions for the control group were conducted in the same manner but with the simulator turned off. In this way, the control group received the same treatment without the rhythmic effects produced by the simulator.

Outcome Measures: The outcome measures used in this study were the GMFM and the SAS, both of which are directly relevant to the clinical question. These outcome measures were assessed immediately following randomization and stratification, after the full 10 weeks of treatment, and after a three month follow-up. The reliability and validity of these outcome measures were not addressed in the study; however, previous studies have demonstrated high levels of inter- and intra-rater reliability[6] and construct validity[7]. A minimal clinically important difference (MCID) of 0.8 on the GMFM-66 has been suggested[16]; however the population used in this calculation was ambulatory children. As the patient of interest was not ambulatory prior to the introduction of hippotherapy, this MCID is not valid for the clinical scenario. The authors also did not address clinimetric properties for the SAS. A previous study has cited excellent reliability values ($p < .0.001$) for the SAS[8], but validity values and an MCID could not be located.

Study Losses: Following randomization and group allocation four subjects dropped out from the study, three from the control group and one from the treatment group. Two of these dropouts occurred during the treatment sessions, and the other two occurred after the end of treatment prior to the three-month follow-up. However, all subjects were included in the data analysis in an intention-to-treat fashion with the previous data points being carried forward. All subjects were analyzed in the groups to which they were originally randomized.

Summary of internal validity: One minor threat and three major threats to internal validity were identified in this study. The first threat is maturation; cerebral palsy is a non-progressive disorder which means that there is a possibility that any changes seen from the pre-test to the three-month follow-up could be due to natural history of the disorder. Because both groups received treatment, improvements in outcome measure scores cannot necessarily be the direct result of the hippotherapy simulator. This, however, is a minor threat as three months is a relatively short time to see significant changes in function in children with CP across the GMFCS levels. The second threat is in study design and reliability of treatment implementation; although not possible in this study, the therapists were not blinded to group allocation. This may inherently result in bias in perceived treatment potential, possibly affecting the therapy provided as it was on an individualized basis rather than a standardized protocol. This is a major threat because as there was no true control group that received no treatment. Differences in individualized therapy provided between groups may have a significant effect on the outcome of the study and result in an incorrect interpretation of observed differences between groups as they may be attributed to the intervention when in actuality they were due to the individualized program. The third threat is the possibility of inadequate power. The authors do not state how the sample size of 38 subjects was established nor did they perform a power analysis, suggesting that there is a strong possibility that an adequate sample size was not obtained. This is a major threat because it can result in either a Type I (observing statistical significance when in fact it does not exist) or Type II (no statistical significance is observed when in fact it does exist) error

and can affect the interpretation of statistical outcomes. The fourth threat is inter-subject differences at baseline. The authors state that the groups were similar at baseline, however following stratification the sample was heavily biased toward children with higher levels of disability; 19 children (50% of the sample) were classified as GMFCS level V. This is a major threat because if there is not a normal distribution of patient presentation across the sample, it can affect the sensitivity of statistical tests to demonstrate true changes over time. Overall, this article has poor internal validity, dramatically decreasing confidence in the study results.

Evidence: Both outcome measures are relevant to the clinical question. However, the results of the SAS are not helpful to the clinical question as it was largely insensitive to changes over the treatment and follow-up periods. The authors focused primarily on dimension B of the GMFM (sitting); however, because the patient of interest had deficits in more functional areas than sitting, the analysis will include the full GMFM score. In addition, as the patient of interest was being treated with hippotherapy for a full year, this analysis will focus on the long-term effect of hippotherapy (baseline vs. three-month follow-up). Results for the GMFCS level V subgroup will also be analyzed due to the sample bias. The following tables include data reported by Herrero *et al*.; effect sizes were calculated by the appraisal author based on mean differences from baseline to three-month follow-up at the conclusion of the study, as the authors did not report statistical significance for any of the observed differences. As an appropriate MCID could not be located, lacking identification of statistical significance limits the ability to interpret the clinical relevance of any observed treatment effects, regardless of the effect size.

Table 5. Results of the SAS for the entire sample (n=38).

Group (n)	Baseline Mean (SD)	3-Month Follow-Up Mean (SD)	Mean Difference (SD)	Effect Size (95% Confidence Interval)	
				Within-Group	Between-Group
Experimental (19)	15.21 (5.93)	15.00 (5.82)	-0.21 (0.92)	0.72	0.59 (-0.06 to 1.24)
Control (19)	15.58 (5.81)	15.84 (5.70)	0.26 (0.65)	0.72	

According to the above calculations, the change in SAS score from baseline to follow-up for both groups correlates with a medium effect size. Additionally, between-group comparison reveals a medium effect size in favor of the experimental group. However, the 95% confidence interval crosses zero, suggesting the true mean may actually favor the control group rather than the experimental group, although this is somewhat unlikely given that the low end of the confidence interval is -0.06. The authors do not report if the baseline to three-month follow-up changes were clinically significant, so definitive and clinically meaningful conclusions from the within-group effect sizes cannot be made.

Table 6. Results of the SAS for GMFCS level V (n=19).

Group (n)	Baseline Mean (SD)	3-Month Follow-Up Mean (SD)	Mean Difference (SD)	Effect Size (95% Confidence Interval)	
				Within-Group	Between-Group
Experimental (10)	11.56 (5.90)	11.11 (5.30)	-0.4 (1.33)	0.63	0.74 (-0.19 to 1.67)
Control (9)	11.00 (5.52)	11.44 (5.63)	0.44 (0.88)	0.63	

According to the above calculations, the change in SAS score from baseline to follow-up for the GMFCS level V subgroup correlates with a medium within-group effect size for both groups and a medium between-group effect size in favor of the experimental group. The 95% confidence interval also crosses zero, suggesting that the true mean may actually favor the control group rather than the experimental group. In addition, the authors do not report if the baseline to three-month follow-up changes were clinically significant, so definitive and meaningful conclusions from the within-group effect sizes cannot be made.

Table 7. Results of the GMFM-66 for the entire sample (n=38).

Group (n)	Baseline Mean (SD)	3-Month Follow-Up Mean (SD)	Mean Difference (SD)	Effect Size (95% Confidence Interval)	
				Within-Group	Between-Group
Experimental (19)	40.91 (17.50)	43.53 (17.16)	2.63 (5.75)	0.20	0.25 (-0.39 to 0.89)
Control (19)	42.75 (19.02)	44.24 (19.76)	1.50 (2.78)	0.20	

According to the calculations presented in Table 7, the change in GMFM-66 scores for the entire sample correlates with a small within-group effect sizes for both groups and a small between-group effect size in favor of the experimental group. However, the confidence interval again crosses zero, decreasing the ability to draw any definitive conclusions from this data. In addition, the authors do not report if the baseline to three-month follow-up changes were significant, so definitive and meaningful conclusions from the within-group effect sizes cannot be made.

Table 8. Results of the GMFM-66 for GMFCS level V (n=19).

Group	Baseline Mean (SD)	3-Month Follow-Up Mean (SD)	Mean Difference (SD)	Effect Size (95% Confidence Interval)	
				Within-Group	Between-Group
Experimental (10)	28.43 (12.07)	31.11 (8.12)	2.68 (7.25)	0.33	0.42 (-0.49 to 1.33)
Control (9)	28.14 (14.68)	28.46 (13.82)	0.32 (2.70)	0.33	

According to the calculations presented in Table 8, the change in GMFM-66 for the GMFCS level V subgroup correlates with small within-group effect sizes for both groups and a small between-group effect size in favor of the experimental group. The 95% confidence interval again crosses zero, and the authors do not report if the baseline to three-month follow-up changes were significant. As a result, definitive and meaningful conclusions from the within-group effect sizes cannot be made.

Table 9. Number of subjects (n=38) who improved at follow-up in the total GMFM-66 score.

		Improvement from baseline		
		Yes	No	Total
Group	Experimental	10	9	19
	Control	12	7	19
	Total	22	16	38

Herrero *et al.* defined 'improvement from baseline' as an improvement in score on the GMFM-66 total score, but does not take into consideration statistical or clinical significance associated with those changes. Using the data presented in Table 7, the calculated number needed to treat (NNT) is 9.5 (95% CI: 2.4 to -4.8). This suggests that 10 patients would need to be treated with

the hippotherapy simulator in order to see one patient improve on the GMFM-66 total score. While the effect size associated with this data demonstrated a medium positive between-group effect of 0.71 (see Table 7), the 95% confidence interval crosses zero at the upper limit, suggesting that the NNT may be as high as infinity. Again, no definitive conclusions can be drawn from this data due to the confidence interval and no clinical correlation of 'improvement' on the GMFM-66.

Table 10. Number of subjects (n=19) who improved on the GMFM-66 at follow-up in the GMFCS level V subgroup.

		Improvement from baseline		
		Yes	No	Total
Group	Experimental	8	2	10
	Control	4	5	9
	Total	12	7	19

Based on the data presented in Table 8, the calculated number needed to harm (NNH) is 2.81 (95% CI: 18.9 to -1.3). Calculating a NNH suggests that three people would need to be treated with the control in order to see one improvement, again suggesting that the control may actually be more beneficial than the intervention. While the effect sizes associated with this data demonstrate a small within-group effect of 0.22 (see Table 8), the 95% CI again crosses zero at the upper limit, suggesting that the NNH may be as high as infinity. As discussed previously, no definitive conclusions can be drawn from this data.

Applicability of study results:

Benefits vs. Costs: Based on the results of this study, few benefits were identified. Time commitment for both the patient and therapist is very minimal as treatment sessions lasted only 15 minutes and were conducted once per week. No adverse events were reported. The equipment used in this study was the Core Trainer Exercise Equipment, JOBA produced by Panasonic. The cost of this specific device could not be located, however similar equipment costs $220-250, which does not present a significant financial burden on a facility.

Despite the low cost for a single hippotherapy simulator, the minimum purchase order for similar equipment as used in this study is 45 units and as a result presents a significant financial and feasibility issue for a facility. Training for proper operation the various settings of the device is also required of the therapist. In addition, while both groups received the same amount of intervention time and resources, neither demonstrated a significant change from baseline in either the SAS or GMFM-66 scores at conclusion of the study, as evidenced by previously reported effect sizes.

For these reasons, and the lack of significant benefit for the patients, the additional financial and time costs for a hippotherapy simulator are not justified.

Feasibility of treatment: The procedures used in this study were not described sufficiently for reproduction, although a previous study by the same authors outlined the equipment settings used during treatment[5]. In addition, the individualized exercises taught to the children were not described, so this could not be replicated. The requirements of equipment, expertise, and time,

however, would be readily available in a clinical setting. The treatment is pain-free and adherence to a home exercise program (HEP) is not required. Ten treatment sessions is reasonable for a plan of care and is also likely within what is allowed by insurance companies, especially for children in a school district setting who are typically seen at least once per week for the full school year. However, as the presented data indicates, hippotherapy simulator treatment does not appear to be an effective treatment for children with cerebral palsy, which creates a large feasibility issue for the therapist, patient, and insurer.

Summary of external validity: While this study matches the clinical PICO very well in terms of age and diagnosis, it compares a hippotherapy simulator (versus true hippotherapy) with an individualized exercise program to barrel-sitting with the same individualized exercise program. A direct comparison of true hippotherapy to "traditional" physical therapy treatment (e.g. stretching and strengthening exercises) for CP would be more beneficial to make an informed clinical decision regarding the usefulness and effectiveness of hippotherapy. In addition, the threats to internal validity compromise the ability to generalize these results to a clinical setting due to a potentially inappropriate sample size, study design and reliability of treatment intervention, and uneven distribution of patient presentation in the two groups. However, the sample is similar to patients treated in a school district for children with CP.

Article: McGibbon NH, Benda W, Duncan BR, Silkwood-Sherer D. Immediate and long-term effects of hippotherapy on symmetry of adductor muscle activity and functional ability in children with spastic cerebral palsy. *Archives of Physical Medicine and Rehabilitation,* 2009; 966-974.

Clinical Bottom Line: Based on the results of this article, hippotherapy intervention appears to be an effective treatment for improving GMFM-66 score in children age 4-16 years with a primary diagnosis of spastic CP, although the high number of threats to internal validity dramatically compromises the results of this study and the ability to make definitive conclusions. This study utilized an A-B design with no control group, therefore changes observed at the conclusion of the study cannot be directly attributed to the intervention. Statistical analysis with a Friedman test and Nemenyi's post-hoc test demonstrated significant differences between T3, T4, and both baseline measures (X^2-calculated = 14.6, X^2-critical = 7.82, $p<0.05$). One minor and six major threats to internal validity were identified, drastically decreasing confidence in the study results and compromising external validity. While hippotherapy appears to be an effective and timely intervention, very poor internal and external validity of these results as well as feasibility issues and a high amount of additional training required of the therapist suggest that further, higher quality research is necessary to reach a decision regarding the potential benefits of hippotherapy.

Article PICO:

Population— Six children with a diagnosis of spastic CP, age 4-16 years

Intervention— 30-minute hippotherapy sessions, delivered once per week for 12 weeks

Comparison— Baseline GMFM-66 measurements

Outcomes— GMFM-66, adductor muscle activity measured via surface electromyography (EMG), the Self-Perception Profile for Children, ages 8 to 13, and the Pictorial Self-Perception Profile for Young children, ages 4 to 7

Blinding: As the authors utilized an A-B-A design where the subjects served as their own control, blinding of the subjects, therapists, and assessors was not possible. This, however, is not a significant threat as the children likely were not able to manipulate their performance to influence the results of the study. In addition, as there was no true comparison group, therapist administration of the treatment and assessor evaluation of the treatment effect would not have a significant effect on the study outcome.

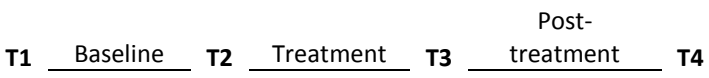
Controls: There was no true control group in this study and the six subjects served as their own control. This was not an appropriate comparison group because the observed differences between groups cannot be attributed to the intervention as there is a large risk of maturation due to the long time periods between assessment sessions.

Randomization: Randomization did not occur in the second, long-term portion of the study, described below. To establish the study sample, the authors recruited the first six participants from the first phase to respond and who met the eligibility criteria.

Study: The study was a pretest/posttest randomized controlled trial conducted at an outpatient hippotherapy center accredited by the North American Riding for the Handicapped Association (now the Professional Association of Therapeutic Horsemanship International [PATHI]) and was conducted under the direction and supervision of a physical therapist certified as a Hippotherapy Clinical Specialist by the American Hippotherapy Certification Board. This study was divided into two phases. Phase I investigated the immediate effects of 10-minutes of hippotherapy on symmetry of adductor muscle activity and functional ability in children with spastic cerebral palsy. Phase II investigated the long-term effects of 12 weeks of weekly 30-minute hippotherapy treatment sessions on adductor activity, self-concept, and gross motor function. Only Phase II is applicable to the clinical scenario, and thus is the only portion that will be analyzed by the appraisal author. To determine the population for Phase II, the authors enrolled the first six children from Phase I who agreed to the study conditions and met the eligibility criteria. Inclusion criteria were as follows: diagnosis of spastic CP made by a pediatric neurologist, age 4-16 years, ability to walk independently with or without an assistive device, ability to comply with the study protocol and follow verbal directions, and sufficient hip abduction to sit astride a horse or barrel with no evidence or report of hip dislocation. Exclusion criteria were as follows: previous history of selective dorsal rhizotomy, tonic clonic seizures uncontrolled by medication, known allergy to horses, dust, or electrode adhesive, surgical procedures, botox injections, lower extremity casting within six months prior to testing, and hippotherapy or horseback riding experience within six months prior to testing. Participants were also required to obtain a complete physician's referral and evaluation prior to participation in the 12 weeks of treatment.

Phase II was a 36-week repeated-measures design divided into three 12-week sections: baseline, treatment, and post-treatment. No intervention was provided on the baseline and post-treatment days. All children were assessed on four separate occasions throughout the 36-week period: T1 (EMG data for Phase I marked the beginning), T2 (12 weeks later immediately prior to the start of treatment), T3 (immediately after completion of 12 weeks of treatment), and T4 (12 weeks after termination of treatment). Figure 1 provides a graphic representation of the assessment scheme. All six children received one weekly session of hippotherapy during the treatment period.

Figure 1. Phase II assessment intervals.



Note: T=test.

The treatment sessions followed a previously published protocol[19], which was modified to address each child's treatment goals. Sessions began with an initial warm-up in which the child was placed on the horse and became relaxed and adjusted to the rhythmic movement of the horse's pelvis and maintaining balance during a dynamic centered sitting position. Once the warm-up was completed, the horse began to walk in straight lines and at gentle curves while the child maintained a central seated position on the horse's back. The challenge to the child was gradually increased via modifications to the horse's movement and walking patterns: introduction of figure eights, circles, or serpentines challenge lateral weight shift and midline postural control whereas lengthening the stride of the horse transmits greater movement

amplitude through the child's pelvis and acceleration/deceleration of walking speed challenges anticipatory postural reactions[3]. Specific exercises were also incorporated into the horseback riding as appropriate. These exercises included changing position on the horse, core stability, and motor planning (i.e. backward sitting, supine to sit, trunk rotation, and side sitting). As the study progressed, a few children were able to use a saddle and stirrups to facilitate lower extremity strengthening, midrange segmental control, and coordination with exercises such as partial stand in the stirrups and a sit-stand-sit sequence. The sessions concluded with a short cool-down phase as needed, but detail of the cool-down process was not detailed by the authors.

Outcome Measures: The outcome measures used in this study included adductor muscle activity using surface EMG, the GMFM-66, the Self-Perception Profile for Children, ages 8 to 13, and the Pictorial Self-Perception Profile for Young Children, ages 4 to 7. However, the only outcome measure of interest to help answer the clinical question is the GMFM-66. Clinimetric characteristics of this measure are as previously discussed. The reported MCID of 0.8 was calculated using ambulatory children[16], and will be used for data analysis as all participants in this study were ambulatory. However, as the patient of interest was not ambulatory prior to the introduction of hippotherapy, the ability to make draw meaningful conclusions from the results will be limited.

Study Losses: This study did not experience any losses. All subjects completed assessment at each testing period.

Summary of internal validity: Seven major threats to internal validity were identified in this study. The first threat is maturation. Due to the duration of the study (36 weeks) there is a possibility that, because CP is a non-progressive disorder, any observed differences at the post-treatment testing period may be due to natural history of the disorder rather than the intervention and that these children would have improved regardless of treatment. The authors did not account for this threat by using a control group. The second threat is selection bias as randomization in selection of participants from the entire sample from Phase I did not occur. Because the authors enrolled the first six children who agreed to the study conditions and met inclusion criteria, there is a strong possibility that these six children were not a heterogeneous sample and did not accurately represent children with spastic CP as a whole. The third is the threat of repeated testing. The children in this study were tested four times during Phase II and as a result, may have become familiar with the GMFM-66, leading to improved scores independent of treatment effect. Because cerebral palsy can present with or without cognitive impairments, this is a major threat. The fourth threat to internal validity is inadequate power. A power analysis was performed for Phase I of this study; however, a repeat power analysis was not performed for Phase II and, as a result, the Phase II sample size is very low. This is a major threat because without a power analysis an appropriate sample size cannot be determined, which can result in a Type I or Type II error as previously discussed. The fifth threat is rater bias. Due to the A-B study design, blinding of the assessors was not possible and creates the possibility that the rater could subconsciously alter the GMFM-66 score given to the children based on the desire to see improvement at the T4 testing session. The sixth threat is inappropriate statistical tests. The authors do not state how the results of the GMFM-66 were analyzed to determine significance of observed changes, and do not provide the results of statistical tests ran on the data set. This is also a major threat because meaningful conclusions cannot be determined without statistical evidence of relevant observed differences. The

seventh threat is lack of a strict protocol. While the hippotherapy treatment sessions followed a similar overall pattern (warm-up, straight-line walking, curved walking, figure eights, cool-down), the authors incorporated individualized therapeutic exercises during horse-back riding as appropriate for the child. As a result, there is the possibility that any observed differences are inappropriately attributed to the intervention when in actuality they were due to an individualized program. In addition, the PEDro score given for the entirety of this study is not appropriate for Phase II on its own due to the large number of internal validity threats. Overall, Phase II of this study has poor internal validity, dramatically decreasing confidence in study results.

Evidence: Three outcome measures were utilized by the authors in Phase II of this study, but as stated previously only the GMFM-66 outcomes are relevant to the clinical question. As such, this appraisal will focus on the two assessment intervals: T1-T2 to determine baseline consistency, and T1-T4 to highlight the long-term effect of hippotherapy on gross motor function. As the patient of interest was being treated for a full year of hippotherapy, this long-term effect is most relevant to the clinical scenario. All statistical calculations were completed by the appraisal author based on the provided individual data by McGibbon *et al*. Table 11 highlights the individual GMFM-66 scores at each testing interval used in the data analysis.

Table 11. Phase II GMFM-66 scores at each interval, as presented by McGibbon *et al*.

Subject	T1	T2	T3	T4
1	56.62	56.86	63.33	65.33
2	51.85	50.62	54.38	53.62
3	47.68	47.09	50.62	52.62
4	65.63	65.33	70.39	68.86
5	58.56	59.86	63.63	64.98
6	44.31	46.32	49.21	50.09

Prior to carrying out any data analysis, the score data highlighted in Table 11 must be converted to rank data. Because the subjects in Phase II were not allocated in a randomized fashion, the small sample size, and the potential lack of a normally distributed sample, analyzing the data using a one-way ANOVA would violate the inherent requirements for use of such tests. As a result, a within-group analysis utilizing the Friedman test will be used for data analysis. Table 12 highlights the Phase II GMFM-66 scores converted to rank data for use in the analysis.

Table 12. Phase II GMFM-66 scores at each interval, adjusted for rank order analysis using a Friedman test. Rank Differences were determined with a Nemenyi's Post-Hoc Test.

Subject	T1	T1 - Rank	T2	T2 - Rank	T3	T3 - Rank	T4	T4 - Rank
1	56.62	1	56.86	2	63.33	3	65.33	4
2	51.85	2	50.62	1	54.38	4	53.62	3
3	47.68	2	47.09	1	50.62	3	52.62	4
4	65.63	2	65.33	1	70.39	4	68.86	3
5	58.56	1	59.86	2	63.63	3	64.98	4
6	44.31	1	46.32	2	49.21	3	50.09	4
Rank Differences		9		9		20		22

Results of the Friedman test demonstrated a X^2-calculated value of 14.6 and a X^2-critical value of 14.6, which suggests that a significant difference in GMFM-66 mean score exists between the four time intervals at an alpha level of 0.05. To determine where significant differences exist within the data set, a Nemenyi's Post-Hoc Test was used, which yielded rank differences highlighted in Table 12. Post-hoc testing also demonstrated a critical value of 12.5, indicating that any rank differences greater than this value are significant at an alpha level of 0.05.

The sample baseline GMFM-66 scores did not change from T1 to T2 (rank difference = 9; $p>0.05$). However, significant differences were observed between T3 and T4 (rank difference = 20 and 22, respectively; $p<0.05$), and T3 and T4 were significantly different from the two baseline intervals. The results of this analysis suggest that hippotherapy resulted in improvement in GMFM-66 score from baseline to immediate cessation of treatment, and from both those intervals to the 12-week follow-up following cessation of treatment.

Applicability of study results:

Benefits vs. Costs: Following the appraisal of this study, few benefits were identified. Time commitment for the therapist, child, and child's family is relatively low as treatment sessions were 30 minutes in duration. No adverse events were reported. While these benefits are supported by significant decrease in GMFM-66 scores from baseline to T4, the number of threats to internal validity suggests that the benefits are not justified when considering the costs and other potential treatment options. All participants received the same amount of treatment time and direct time with the therapist.

While there are over 850 certified hippotherapy centers around the world, availability of a center largely varies based on geographic location will affect feasibility of treatment. For example, over 100 locations exist within the United States, but there is only one PATHI-certified center in Canada[25]. Additional training by the therapist is also required, including a board certification to become a credentialed hippotherapy instructor, which is potentially a large barrier preventing access to appropriate hippotherapy treatment. In addition, while the groups received the same total time with a therapist and number of treatment sessions, the individualized nature of therapeutic exercise prescription compromises the ability to make definitive decisions regarding the reported outcomes of the study.

For these reasons and the large number of threats to internal validity, the additional costs and feasibility issues regarding implementation of hippotherapy do not appear to be justified.

Feasibility of treatment: The procedures utilized in this study are described sufficiently for reproduction, including the primary individualized therapeutic exercises used during hippotherapy sessions. The requirements of equipment (access to a PATHI-certified center and clinical specialization) are likely not readily available in a clinical setting; however, the duration of hippotherapy sessions are within the timeframe allowed by most insurance and physical therapy companies. This is especially relevant for a school-district setting where children are typically seen once per week for a full year. Treatment was reported to be pain-free and adherence to an HEP was not required. However, the large number of threats to internal validity compromises the ability to make clinical decisions based on the outcome of this study, which creates a large feasibility issue for the patient, therapist, and insurer.

Summary of external validity: While this study matches the clinical PICO very well in terms of patient age, primary diagnosis, both length and type of intervention, and typical population seen in a school district for children with CP, it utilizes an A-B study design. The lack of a true control group and comparison to "traditional" physical therapy intervention in conjunction with the large number of major threats to internal validity further limit the ability to make informed clinical decisions regarding the efficacy of hippotherapy over other treatment options and ability to extrapolate the results to other populations or types of CP.

Article: Kang H, Jung J, Yu J. Effects of hippotherapy on the sitting balance of children with cerebral palsy: a randomized controlled trial. *Journal of Physical Therapy Science*, 2012; 833-836.

Clinical Bottom Line: Based on the results of this article, hippotherapy plus traditional physical therapy appears to be an effective and safe treatment option to improve sitting balance in children age 6-10 years with a primary diagnosis of severe hemiplegic or diplegic CP. However, one minor and four major threats to internal validity, as well as poor feasibility, reduce confidence in and generalizability of the results.

This study was a randomized controlled trial where the both the hippotherapy and physical therapy group received traditional stretching and strengthening exercises for 30 minute sessions, semi-weekly, for eight weeks. The hippotherapy group additionally received physical therapy treatment on the same schedule, and the control group received no intervention. Statistical analysis demonstrated that, while the physical therapy group improved in sitting balance from pre- to post-treatment ($p < 0.05$, effect size range: 0.04 - 0.13), the hippotherapy group demonstrated much larger effect sizes for the same time frame ($p <0.05$, effect size range: 0.92 - 1.17). The hippotherapy group also demonstrated large between-group effect sizes that favor hippotherapy over physical therapy or control ($p < 0.05$, effect size range: 0.37 - 2.07). While hippotherapy plus traditional physical therapy appears to be an effective and timely intervention, internal and external validity and feasibility issues suggest that further research is necessary to reach a decision regarding the clinical PICO. Future research should include a more strict study protocol and design and larger sample size to reduce the threats to internal validity.

Article PICO:

Population— 45 children with hemiplegic or diplegic CP, age 6-10 years

Intervention— Hippotherapy plus physical therapy (HTG)

Comparison— Physical therapy (PTG) and no intervention (CON)

Outcomes— Sitting balance as measured by center of pressure sway on a force plate

Blinding: Blinding of the therapists to group allocation was not possible due to the nature of this study; it was readily apparent to which group participants were allocated. However, subject blinding was utilized as participants in the HTG and PTG groups were kept separate at all times. The HTG received treatment in the morning while the PTG received treatment in the afternoon, and participants were forbidden to speak with one another regarding the treatment and study. The authors do not state whether the assessor was blinded to group allocation. This is a potentially significant threat as the assessor may either intentionally or unintentionally favor one group over another.

Controls: There were two comparison groups in this study: PTG, which only received physical therapy, and CON, which received no treatment. These both were appropriate comparison

groups as the sole difference between HTG, PTG, and CON was the type of intervention applied, meaning differences between groups can most likely be attributed to the intervention.

Randomization: Participants were randomly assigned to the three groups using a table of random sampling numbers. The authors do not state whether or not randomization was concealed; however, randomization was successful as the authors report no statistical differences in any variables between groups prior to therapy. As the subjects included in this study had either hemiplegic or diplegic CP, randomization should have been stratified due to the significant differences in motor ability between these two populations.

Study: This was a single-blind randomized controlled trial with a total of 45 participants in three groups, carried out by the K equestrian team and staff from S hospital in Korea. Following explanation of the study and obtaining consent, participants were allocated to one of three groups (HTG, PTG, and CON) as stated previously. Inclusion and exclusion criteria were as follows: independent gait ability of less than 10 minutes, no horse riding-related experience in previous two years, no internal or neurological surgery in previous two months, and no specific medical problems including psychological problems.

Following randomization, a baseline sitting balance was assessed using a force plate placed on a wooden box. Subjects sat in the center of the force plate and were asked to look at a 10 cm diameter circle placed one meter ahead. The center of upper body weight was traced for 30 seconds for three repetitions with rest allowed to relieve fatigue. Following baseline assessment, the HTG and PTG were provided with traditional physical therapy comprised of strengthening and stretching exercises for 30-minute sessions, semi-weekly for eight weeks. This treatment was performed by two expert physical therapists but was not described further. The HTG also received hippotherapy, in which a leader pulled the reins of a horse at the front and two side-walkers held the legs of the children to prevent them from falling. The hippotherapy consisted of sitting and standing in the saddle, manipulating objects such as a bar, ball, ring, and toy, and maintaining posture while the horse moves. No other specific exercises or protocol was provided. The CON received no treatment.

Following eight weeks of intervention, a post-test was performed and the results were analyzed utilizing a paired t-test to assess within-group differences in sitting balance from baseline to post-treatment and a one-way ANOVA to determine between-group differences from baseline to post-treatment. Post-hoc calculations were performed using the Bonferroni correction, and statistical significance for all tests was accepted with an alpha-value of 0.05.

Outcome Measures: While not directly relevant to gross motor function, the only outcome measure utilized in this study is sitting balance as measured by center of pressure excursion on a force plate. Reliability and validity was not addressed by the authors, and an MCID was not provided. However, these clinimetric properties have been determined in previous studies. Kyvelidou *et al.* (2012) found that center of pressure excursion measurements on a force plate system demonstrated high intra- and inter-session intraclass correlation coefficients (ICC), and these reliability values increased as sitting posture improved[20]. In addition, center of pressure excursion has been found to be sensitive in discernment of balance performance in children with cerebral palsy[21-24]. It would also appear this method of quantifying sitting balance has face validity as any change in center of pressure excursion while sitting on a force plate would be the direct result of postural sway.

Study Losses: One child each in the HTG and CON withdrew from the study prior to post-treatment assessment; however, no further details were reported regarding the reasons for withdrawal. However, all subjects were analyzed in the groups to which they were randomized and an intention-to-treat analysis was not indicated as study losses totaled less than 15% of both the total sample size and individual group sample sizes.

Summary of internal validity: One minor threat and three major threats to internal validity were identified in this study. The first threat is extraneous variables, such as patient motivation or activities outside of the experimental treatment sessions. This is a minor threat, however; as the study population was comprised of severe CP, extra-experimental activities likely would not cause a large effect on sitting balance in this population. The second threat is rater bias. This is a major threat, as a lack of blinding of the assessors can cause a subconscious tendency to favor the HTG over PTG and CON. The second major threat is a lack of statistical power. The authors do not state if a power analysis was performed prior to randomization of subjects and data collection, and an insufficient sample size has the potential to cause a Type I or Type II error, as described previously. The third major threat is the Hawthorne Effect. Participants tend to act differently when they know they are being studied and may exhibit a change in baseline to post-treatment outcome measure scores simply due to interaction with a therapist or researcher. As it is not clear if all groups received the same amount of treatment time or if CON had direct interaction with a researcher, there is the possibility that observed treatment effects could be due to differences in direct interaction with a clinician or researcher. Overall, this study has poor internal validity, dramatically decreasing confidence in the study results.

Evidence: The only outcome measure used in this study is sitting balance as measured by center of pressure excursion and velocity on a force plate. While this is not a direct measure of gross motor function, there is a sitting component of the GMFM and as such, this outcome measure is considered relevant to the clinical scenario. Significant differences between groups that are identified below were calculated by the study authors. All effect size calculations were performed by the appraisal author.

Table 13. Excursion and velocity of excursion data for each group, from pre-test to post-test. Data represented as group mean ± standard deviation.

Item	Direction	Pre-test			Post-test		
		HTG (n=15)	PTG (n=15)	CON (n=15)	HTG (n=14)	PTG (n=15)	CON (n=14)
Excursion (cm)	M/L	49.1 ±30.3	56.8 ± 21.6	59.1 ± 27.0	21.3 ± 9.2	55.9 ± 23.1	88.5 ± 52.1
	A/P	112.4 ± 56.2	93.4 ± 40.0	115.3 ± 51.3	48.3 ± 13.0	82.9 ± 22.4	104.9 ± 47.2
	Total	132.7 ± 69.1	120.8 ± 43.5	141.6 ± 60.6	56.4 ± 16.5	110.1 ± 34.1	158.6 ± 71.1
Velocity (cm/sec)	M/L	2.0 ± 1.2	2.3 ± 0.8	2.3 ± 1.0	0.8 ± 0.3	2.2 ± 0.9	3.5 ± 2.0
	A/P	4.6 ± 2.3	3.7 ± 1.5	4.6 ± 2.0	1.9 ± 0.5	3.3 ± 0.9	4.2 ± 1.9
	Total	5.4 ± 2.8	4.9 ± 1.7	4.9 ± 1.7	2.2 ± 0.6	4.4 ± 1.3	6.4 ± 2.8

Note. M/L: medial/lateral, A/P: anterior/posterior.

Individual data and p-value values were not reported by the authors; however, based on the reported outcomes of an ANOVA and t-test, within-group comparison before and after therapy demonstrated that all variables in the HTG were significantly decreased following treatment ($p < 0.05$). PTG demonstrated significant differences in M/L excursion and velocity ($p < 0.05$), and the

CON demonstrated no significant differences.

Table 14. Within-group comparisons, calculated by the appraisal author. Data represented as group mean ± standard deviation.

Item	Direction	HTG		PTG		CON	
		Mean Difference	Effect Size	Mean Difference	Effect Size	Mean Difference	Effect Size
Excursion (cm)	M/L	27.8 ± 19.75	0.92	0.9 ± 22.35	0.04	29.4 ± 39.55	1.09
	A/P	64.1 ± 34.6	1.14	10.5 ± 31.2	0.26	7.4 ± 49.25	0.14
	Total	76.3 ± 42.3	1.13	10.7 ± 38.8	0.25	17 ± 57.3	0.39
Velocity (cm/sec)	M/L	1.2 ± 0.75	1	0.1 ± 0.82	0.13	1.2 ± 1.5	1.2
	A/P	2.7 ± 1.4	1.17	0.4 ± 1.2	0.27	0.4 ± 1.95	0.2
	Total	3.2 ± 1.7	1.14	0.5 ± 1.5	0.29	1.5 ± 2.25	0.88

Note. HTG: pre-test n=15, post-test n=14. PTG: pre-test n=15, post-test n=15.

The within-group effect sizes for HTG correlate with large effect sizes for all variables, suggesting that hippotherapy plus traditional physical therapy is a very effective treatment option to improve sitting balance in this population. However, analysis of the PTG demonstrated very small effect sizes for M/L excursion and velocity. This suggests that while statistically significant decreases were observed in these variables, the difference may not be clinically relevant.

Between-group analysis and post-hoc testing demonstrated that excursion and velocity were significantly decreased in the HTG compared to PTG and CON ($p < 0.05$). PTG demonstrated significant decreases in only M/L and total excursion and velocity ($p < 0.05$).

Table 15. Between-group comparisons, calculated by the appraisal author. Effect sizes calculated based on mean differences reported in Table 14.

Item	Direction	HTG vs PTG	HTG vs CON	PTG vs CON
		Effect Size (95% CI)	Effect Size (95% CI)	Effect Size (95% CI)
Excursion (cm)	M/L	1.27 (0.47 - 2.07)	0.05 (-0.69 - 0.79)	0.90 (0.13 - 1.66)
	A/P	1.63 (0.79 - 2.47)	1.33 (0.51 - 2.15)	0.08 (-0.65 - 0.80)
	Total	1.62 (0.78 - 2.46)	1.18 (0.38 - 1.98)	0.13 (-0.60 - 0.86)
Velocity (cm/sec)	M/L	1.40 (0.59 - 2.21)	0.00 (-0.74 - 0.74)	0.92 (0.15 - 1.69)
	A/P	1.77 (0.91 - 2.63)	1.35 (0.53 - 2.18)	0.00 (-0.73 - 0.73)
	Total	1.13 (0.34 - 1.91)	0.85 (0.08 - 1.63)	0.53 (-0.21 - 1.27)

Regarding HTG vs PTG, large effect sizes were observed in all variables except total velocity. In addition, 95% confidence intervals do not cross zero for any measures. Regarding HTG vs CON, all variables demonstrated large effect sizes and 95% confidence intervals except M/L excursion and M/L velocity, however, on average these effect sizes are not as large as HTG vs PTG, which is likely due to significantly larger confidence intervals that also cross zero on multiple occasions. Regarding PTG vs CON, M/L excursion and velocity demonstrated large effect sizes. All other variables demonstrate small or medium effect sizes, however the confidence intervals cross zero.

The results of this study demonstrate that while PTG had within-group improvement in sitting balance from baseline to post-treatment, as well as significant improvements as compared to control, HTG also experienced significant improvements and large effect sizes in both within- and between-group comparisons to both PTG and CON (effect sizes range: 0.37 - 2.66; $p < 0.05$). This suggests that HTG (hippotherapy plus traditional physical therapy) is more effective than physical therapy alone and no treatment to improve sitting balance in children, ages 6-10 years, with spastic hemiplegic and diplegic CP.

Applicability of study results:

Benefits vs. Costs: Following the completion of the study appraisal, few benefits were identified. Time commitment for the therapist, patient, and family was low as treatment sessions were 30 minutes in duration, performed semi-weekly for eight weeks. While not directly addressed by the study authors, financial costs also appear relatively low as no additional equipment was required for treatment other than access to the equestrian team. The results of the study also suggest that, despite a low time and financial commitment, hippotherapy appears to be an effective treatment option for this population, further justifying these benefits.

Access to a certified hippotherapy center or equestrian team is a large cost for this type of treatment, as mentioned previously. In addition, while the HTG and PTG received the same number of treatment sessions, total treatment time appears to be different as both groups received a 30-minute session of physical therapy outside of the hippotherapy component of the HTG. It is also not readily apparent if CON received any direct time with the therapists throughout this study. In addition, extra training is required of the therapists to become educated in hippotherapy treatment.

Feasibility of treatment: The procedures and experimental parameters utilized in this study are not described well enough for reproduction. Details of exercises provided during the hippotherapy sessions and physical therapy sessions are not provided, and it is not clear what intervention, if any, was received by CON. While requirements of clinician expertise and time commitment are what would likely be available in a PT setting, ease of locating a certified hippotherapy center presents a large barrier to prevent clinicians from utilizing hippotherapy. The number and duration of PT sessions in this study are within the realm of that allowed by insurance companies, however, and an HEP was not required. Treatment was feasible for the patients as well, as treatment was not painful, no adverse events were reported, and the data suggests that hippotherapy in conjunction with physical therapy is an effective treatment option for this population.

Summary of external validity: This study most accurately matches the patient age, primary diagnosis, both length and type of intervention, and typical population seen in a school district for children with CP. In addition, this study utilizes the treatment intervention schedule that is most relevant to the clinical scenario by comparing hippotherapy plus physical therapy to physical therapy alone and control. However, the number of major threats to internal validity compromise the ability to generalize results to a larger population.

Article: Casady RL & Nichols-Larsen DS. The effect of hippotherapy on ten children with cerebral palsy. Pediatric Physical Therapy, 2004; 165-172.

Clinical Bottom Line: Based on the results of this article, hippotherapy appears to be an effective and safe treatment option to improve gross motor function in children age 2-7 years with a primary diagnosis of CP. However, eight major threats to internal validity were identified, which, despite adequate feasibility of treatment, dramatically reduce confidence in and generalizability of the results. This study utilized an A-B-A design where all ten children received 45-minute hippotherapy treatment sessions once per week for ten weeks. Analysis with analysis-of-variance (ANOVA) and a Tukey post-hoc test for honest significant difference (HSD) demonstrated a significant improvement in the crawling/kneeling dimension (mean difference = 2.85, HSD = 2.64), and the total GMFM-88 score (mean difference = 12.35, HSD = 9.34). While hippotherapy appears to be an effective and timely intervention, internal and external validity issues suggest further research is necessary to reach a decision regarding the clinical PICO, which should include a strict treatment and study protocol and design and a larger sample size.

Article PICO:

Population— 11 children with a primary diagnosis of CP, age 2.3-6.8 years.

Intervention— Hippotherapy

Comparison— Baseline GMFM measurements

Outcomes— PEDI and GMFM

Blinding: As the authors utilized an A-B-A design where the subjects served as their own control, blinding of the subjects, therapists, and assessors was not possible. This, however, is not a significant threat as the children likely were not able to manipulate their performance to influence the results of the study. In addition, as there was no true comparison group, therapist administration of the treatment and assessor evaluation of the treatment effect would not have a significant effect on the study outcome.

Controls: There was no true control group in this study and the 11 subjects served as their own control. This was not an appropriate comparison group because the observed differences between groups cannot be attributed to the intervention as there is a large risk of maturation due to the long time periods between assessment sessions and overall 30-week study duration.

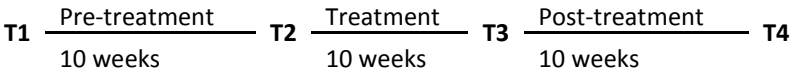
Randomization: Due to study design, randomization was not possible. To establish the study sample, the authors utilized convenience sampling via physician and therapist recruiting of appropriate subjects.

Study: This study utilized a single-subject, A-B-A design and was conducted at an NARHA (now PATHI) therapeutic riding center under the supervision of a physical therapist with 18 years of practical experience and certification in hippotherapy and a NAHRA-registered therapeutic riding instructor. Study authors obtained a convenience sample by contacting local physicians and therapists and to identify eleven subjects who were appropriate for the study. Inclusion and exclusion criteria were as follows: primary diagnosis of CP and have never had hippotherapy

treatment. Subjects were also required to obtain medical clearance to ride a horse and a referral for physical therapy from their primary physician. Any concurrent events that may have affected development and study outcomes were reported by the parent.

A time-series, quasi-experimental research design was utilized to show trends in development over a 30-week period. Four testing dates were implemented at which time all outcome measures were assessed. These testing dates are as follows: pre-test 1 (T1), pre-test 2 (T2), post-test 1 (T3), and post-test 2 (T4), and a 10-week time interval between test dates was held constant. Following T2, the treatment phase began, and concluded with T3. Figure 2 provides a graphic representation of the treatment and assessment scheme. All eleven subjects received one weekly session of hippotherapy during the treatment period.

Figure 2. Assessment intervals.



Assessment of subjects was completed via a videotape review. Subjects were recorded on each test date while performing the items on the GMFM that were appropriate to their current skill level. These recordings were later watched and rated by two scorers who had no contact with one another during the study. One scorer was a PT with 26 years of pediatric experience; the other was an OT with 20 years of pediatric experience. Reliability was established between the two scorers using videotapes of three children who did not participate in the study performing the GMFM test items. Previous research has demonstrated reliability values ranging from 0.77-0.88 using videotaped assessments viewed by therapists with no previous GMFM experience[6]. The average absolute agreement between the two scorers was 94.6% and interrater reliability was excellent ($ICC_{[3,1]}$ = 0.990-0.999).

Individual hippotherapy appointments were 45 minutes in length, although the authors state that the actual amount of time on the horse was typically 20-30 minutes. The physical therapists selected the appropriate horse for each subject and the horses typically worked during sessions with an experienced horse handler walking behind the horse. The physical therapist and a volunteer side walker walked along either side of the horse to assist the subjects with postural control as necessary, including initially riding tandem with the subjects. The horse's movement was modified during treatment sessions depending on the needs and responses of the subject, which included direction to walk, halt, alter tempo, or change the pattern of the horse's movement. While the horse walked, subjects were instructed to maintain postural alignment and symmetry of the head, trunk, and lower extremities with as little assistance as possible. If necessary, a U-shaped pillow was placed around the waist of subjects in the event that external support was needed.

Outcome measures: The outcome measures used in this study were the PEDI and GMFM-88; however, the only outcome measure of interest to help answer the clinical question is the GMFM-88. Clinimetric characteristics of this measure are as previously discussed. The reported MCID of 0.8 was calculated using ambulatory children[16]. However, as the patient of interest was not ambulatory prior to the introduction of hippotherapy and 60% of the subjects in this study were non-ambulatory, this MCID will not be used for data analysis.

Study losses: This study experienced a loss of one subject, which occurred prior to T2 and it is not clear why this loss occurred. An intention-to-treat analysis was not indicated.

Summary of internal validity: Eight major threats were identified in this study, reducing confidence in the authors' findings. The first two threats are related to lack of blinding, the first being the Hawthorne Effect. Without subject blinding, repeated testing and observation with a video camera, and introduction of a new therapy intervention, the subjects may behave differently when they know they are being studied. Rater bias is also a major threat as a lack of assessor blinding may result in a subconscious tendency to rate the subjects higher on the GMFM-88 as the study progresses, thus inappropriately influencing study results.

Threats also exist because of the procedures used by the authors for subject recruitment and group assignment. Without the inclusion of a control group, the threat of maturation exists. As CP is a non-progressive disorder and the study duration was 30-weeks in length, there is the possibility that changes in GMFM-88 scores seen at T3 and T4 may be due to natural history of the disorder rather than the intervention. Also, because the authors utilized convenience sampling to locate appropriate subjects for the study, a selection bias threat is introduced and the study sample may not be heterogeneous nor be an accurate representation of children with CP as a whole, which dramatically affects the external validity of the results. Furthermore, inadequate power is a threat. With a small sample size and no power analysis performed, either a Type 1 or Type II error may exist.

The final three threats relate to the study procedures carried out by the authors. Children in this study were assessed four times throughout the study and, as a result, may have become familiar with the testing procedures for the GMFM-88, leading to improved scores independent of treatment effect. Because CP can present with or without cognitive impairments, this is a major threat. The authors also did not control for extraneous variables. Because concurrent events were not restricted during the study duration, some subjects continued to receive standard physical, occupational, and/or speech and language therapy. In addition, one subject failed to wear a new AFO, another began oral baclofen during pre-treatment, another had bilateral otitis media during treatment, and others needed to be treated with hyperbaric oxygen during various phases of the study. This is a major threat because these events, particularly the continued therapy for some subjects, could lead to changes in GMFM-88 scores independent of treatment effects related to the hippotherapy. The final threat to internal validity is the lack of a strict treatment protocol. Because hippotherapy treatment was not standardized across the sample, there is the possibility that improvements in GMFM-88 scores seen at T3 and T4 may be due to individually tailored treatment sessions rather than hippotherapy treatment effect. Overall, this study has poor internal validity, dramatically decreasing confidence in study results.

Evidence: The outcome measure that is applicable to the clinical scenario is the GMFM-88. As a result, data and calculations discussed in this section will only analyze the GMFM-88 scores. The authors do not report individual or group mean data at each assessment period, therefore all reported statistical analysis was conducted by the study authors. Effects of hippotherapy were analyzed with an ANOVA, and the results are represented in Table 16.

Table 16. Results of GMFM-88 ANOVA, calculated by study authors.

GMFM Dimension	df	F	p-value
Lying/rolling	1.26	3.47	0.082
Sitting	1.34	10.99	0.004*
Crawling/kneeling	1.75	18.21	< 0.000*
Standing	1.36	12.05	0.003*
Walk/run/jump	1.07	6.69	0.027*
Total	1.26	32.62	< 0.000*

*Note. Significance value set at p < 0.05; * denotes statistically significant difference.*

ANOVA demonstrated that four of the five dimensions (sitting, crawling/kneeling, standing, and walk/run/jump) in the GMFM-88, as well as the total GMFM-88 score, demonstrated a significant change in score, as represented by a p-value < 0.05. Because this analysis includes change in three different phases of the study, the study authors conducted a Tukey post-hoc test to determine in which phases these significant differences occur. A Tukey HSD was calculated for each dimension of and the total GMFM-88 scores, and results are represented in Table 17.

Table 17. Results of Tukey post-hoc test.

GMFM Dimension	Tukey HSD	Mean Difference		
		T1-T2	T2-T3	T3-T4
Lying/rolling	--	--	--	--
Sitting	3.76	2.85	3.20	1.15
Crawling/kneeling	2.64	2.35	2.85*	1.55
Standing	1.99	1.20	1.65	1.25
Walk/run/jump	4.98	1.95	2.80	2.90
Total	9.34	9.00	12.35*	8.25

*Note. * denotes statistically significant difference relative to Tukey HSD. Data unavailable for lying/rolling due to p > 0.05.*

Following the Tukey post-hoc test, a significant difference was found in the crawling/kneeling dimension and the total GMFM-88 score, represented by a mean difference of greater value than the calculated Tukey HSD. This suggests that hippotherapy yielded a significant change in gross motor function for children with a primary diagnosis of CP. The mean difference of T1-T2 did not exceed the Tukey HSD, suggesting no significant change in baseline assessments prior to the introduction of hippotherapy. Additionally, the mean difference of T3-T4 did not exceed the Tukey HSD, suggesting that improvements in gross motor function were maintained following cessation of treatment with hippotherapy.

Applicability of study results:

Benefits vs. Costs: Following completion of the study appraisal, a number of benefits were identified. Time commitment for the therapist, patient, and family was low as treatment sessions were 30 minutes in duration and performed one time per week for ten weeks. An HEP is not required for this type of treatment. While not directly addressed by the study authors, financial costs also appear relatively low as no additional equipment was required for treatment

other than access to a certified hippotherapy center. Additionally, no adverse events were reported. Based on the results of this study, these benefits are justified as hippotherapy appears to be an effective treatment option for this population.

Access to a certified hippotherapy center and equestrian team is a large cost for this type of treatment due to potential transportation costs and general access, as discussed previously. Additional training is also required of the therapists to become educated in hippotherapy treatment.

Feasibility of treatment: The procedures and experimental parameters utilized in this study are not described well enough for reproduction. Details of treatment session protocol are not provided as treatment appears to have been individualized. This number and duration of treatment sessions is well within what is typically covered by insurance companies, especially for children who typically receive treatment in a school district setting. Additionally, the treatment appears to be feasible for patients as a home exercise program is not required for benefits to be attained, treatment was not painful, and no adverse events were reported. However, because this study was conducted at an external hippotherapy center, there is the potential that children must be taken out of school to attend treatment sessions, depending on hours of operation of the hippotherapy center and therapist availability.

Summary of external validity: This study accurately matches the patient age, primary diagnosis, length and type of intervention, and a typical population seen in a school district. In addition, the subjects in the study presented with a variety of types of CP, which are discussed previously. However, the large number of major threats to internal validity and lack of a sufficiently described treatment protocol greatly compromise the ability to generalize results to a larger population.

Synthesis/Discussion: The purpose of this critically appraised topic was to compare the efficacy of hippotherapy as a treatment option for children with a primary diagnosis of cerebral palsy to 'traditional' physical therapy consisting of stretching and strengthening exercises using the highest quality available research.

The PEDro scores for these chosen studies ranged from 8/10 to 3/10. PEDro scores rated above 5/10 are considered to have moderate to high quality research design as reported by the PEDro database. Two articles met this criterion. However, the PEDro score for McGibbon *et al*. was reported by the database with a score of 7/10, but this score does not apply to the entirety of the study, as discussed in the individual article appraisal.

Based on the four studies located that address this topic, there is relatively consistent evidence supporting the use of hippotherapy as an intervention for children with CP. While Hererro *et al*. demonstrated that a hippotherapy simulator does not provide statistically or clinically meaningful improvements in gross motor function for this population, McGibbon *et al*., Kang *et al*., and Casady *et al*. found that hippotherapy is an effective means to improve gross motor function. This discrepancy in outcomes is likely due to the treatment effect being studied and multiple methodological flaws. Additionally, only McGibbon *et al*. assessed hippotherapy plus traditional physical therapy compared to physical therapy alone, which most accurately matches the clinical scenario. However, due to multiple study design and major internal validity threats, the ability to draw meaningful conclusions from all four articles is very limited.

Notable methodological flaws are as follows. None of the four articles utilized blinding of the subjects or therapists, and Kang *et al*. and Casady *et al*. did not blind assessors. These are major threats to internal validity because of the potential subconscious tendency to not only rate subjects higher on outcome measures in a manner that supports the hypothesis but also to provide treatment interventions that bias the experimental group. McGibbon *et al*. and Casady *et al*. also utilized an A-B and A-B-A study design, respectively. The lack of a true control group in these studies, combined with the lack of a strict treatment protocol, limits the ability to conclude that observed treatment effects were actually due to the intervention. Additionally, McGibbon *et al*. utilized inappropriate statistical testing, as the manner in which data was collected violates the inherent assumptions of score data tests (i.e. t-test or ANOVA). As a result, the appraisal author converted data to rank order data and was assessed with a Friedman test. This discrepancy further limits the ability to draw meaningful conclusions from the reported results.

Other differences between studies were apparent, including the lack of a long-term follow-up by Hererro *et al*. As the clinical scenario necessitates the identification of long-term effects of hippotherapy, this limits the external validity of the study. Additionally, McGibbon *et al.*, Hererro *et al*., and Casady *et al*. did not utilize a power analysis prior to any data collection, which suggests the sample size may be too small and could result in a Type I or Type II error. This further limits the ability to draw meaningful conclusions from the results as a more appropriate sample size may yield different results. Finally, Kang *et al*. did not utilize an outcome measure that directly addresses gross motor function. However, as one dimension of the GMFM-66 and -88 addresses sitting balance, postural sway excursion and velocity as measured by a force plate appears to have face validity, is sensitive in discernment of balance performance in this population[21-24], and would theoretically correlate with findings on the GMFM.

In conclusion, while it appears that hippotherapy is an effective and timely intervention to improve gross motor function for children with cerebral palsy, poor study design and multiple methodological flaws suggest that further high quality research is necessary to reach a definitive conclusion regarding the clinical PICO.

References:

1. Herrero P, Gomez-Trullen EM, Asensio A, Garcia E, Casas R, Monserrat E, Pandyan A. Study of the therapeutic effects of a hippotherapy simulator in children with cerebral palsy: a stratified single-blind randomized controlled trial. *Clinical Rehabilitation,* 2012; 1105-1113.
2. Kang H, Jung J, Yu J. Effects of hippotherapy on the sitting balance of children with cerebral palsy: a randomized controlled trial. *Journal of Physical Therapy Science*, 2012; 833-836.
3. McGibbon NH, Benda W, Duncan BR, Silkwood-Sherer D. Immediate and long-term effects of hippotherapy on symmetry of adductor muscle activity and functional ability in children with spastic cerebral palsy. *Archives of Physical Medicine and Rehabilitation,* 2009; 966-974.
4. Casady RL & Nichols-Larsen DS. The effect of hippotherapy on ten children with cerebral palsy. *Pediatric Physical Therapy,* 2004; 165-172.
5. Herrero P, Asensio A, Garcia E, Marco A, Olivan B, Ibarz A, Gomez-Trullen EM, Casas R. Study of the therapeutic effects of an advanced hippotherapy simulator in children with cerebral palsy: a randomised controlled trial. *BMC Musculoskeletal Disorders*, 2010.
6. Nordmark E, Haaglund G, Jarnio GB. Reliability of the gross motor function measure in cerebral palsy. *Scandanavian Journal of Rehabilitation Medicine,* 1997; 25-28.
7. Lundkvist JA, Jarnio GB, Gummesson C, Nordmark E. Longitudinal construct validity of the GMFM-88 total score and goal total score and the GMFM-66 score in a 5-year follow-up study. *Physical Therapy,* 2009; 342-350.
8. Myrh U, Von Wendt L, Sandberg KW. Assessment of Sitting in Children with Cerebral Palsy from Videofilm: A Reliability Study. *Physical and Occupational Therapy in Pediatrics,* 1993; 21-35.
9. Casady RL, Nichols-Larsen DS. The effect of hippotherapy on ten children with cerebral palsy. Pediatric Physical Therapy, 2004; 165-172.
10. Fleck, C. Hippotherapy: Mechanics of human walking and horseback riding. Rehab with the Aid of the Horse: A Collection of Studies, 1992; 153-175.
11. Kuczynski M, Slonka K. Influence of artificial saddle riding on postural stability in children with cerebral palsy. Gait and Posture, 1999; 154-160.
12. Quint C, Toomey M. Powered saddle and pelvic mobility: an investigation into the effects on pelvic mobility of children with cerebral palsy of a powered saddle which imitates the movements of a walking horse. Physiotherapy, 1998; 376-384.
13. Sterba JA. Does horseback riding therapy or therapist-directed hippotherapy rehabilitate children with cerebral palsy? Developmental Medicine and Child Neurology, 2007; 68-73.
14. Sterba JA, Rogers BT, France AP. Horseback riding in children with cerebral palsy: effect on gross motor function. Developmental Medicine and Child Neurology, 2002; 301-308.
15. Zadnikar M, Kastrin A. Effects of hippotherapy and therapeutic horseback riding on postural control or balance in children with cerebral palsy: a meta-analysis. Developmental Medicine and Child Neurology, 2011; 684-691.
16. Oeffinger D, Bagley A, Rogers S, Gorton G, Kryscio R, Abel M, Damiano D, Barnes D, Tylkwowski C. Outcome tools used for ambulatory children with cerebral palsy:

responsiveness and minimum clinically important differences. *Developmental Medicine and Child Neurology*, 2008; 918-925.
17. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology,* 1997; 214-223.
18. Effgen, SK. *Meeting the physical therapy needs of children.*2nd ed. 2012, Philadelphia: F.A. Davis Company.
19. McGibbon NH, Andrade CK, Widener G, Cintas HL. Effect of an equine movement therapy program on gait, energy expenditure, and motor function in children with spastic cerebral palsy: a pilot study. *Developmental Medicine and Child Neurology*, 1998; 754-762.
20. Kyvelidou A, Harbourne RT, Shostrom VK, and Stergiou, N. Reliability of center of pressure measures for assessing the development of sitting postural control in infants with or at risk of cerebral palsy. *Archives of Physical Medicine and Rehabilitation*, 2012; 1593-1601.
21. Liao HF, Mao PJ, Hwang AW. Test-retest reliability of balance tests in children with cerebral palsy. *Developmental Medicine and Child Neurology*, 2001; 180-186.
22. Cherng RJ, Su FC, Chen JJ, *et al*. Performance of static standing balance in children with spastic diplegic cerebral palsy under altered sensory environments. *American Journal of Physical Medicine and Rehabilitation*, 2007; 336-343.
23. Liao SF, Yang TF, Hsu TC, *et al*. Differences in seated postural control in children with spastic cerebral palsy and children who are typically developing. *American Journal of Physical Medicine and Rehabilitation,* 2003; 622-626.
24. Rojas VG, Rebolledo GM, Munoz EG, *et al*. Differences in standing balance between patients with diplegic and hemiplegic cerebral palsy. *Neural Regeneration Research*, 2013; 2478-2483.
25. Find a Center. (2014, January 1). Retrieved November 19, 2014, from http://www.pathintl.org/path-intl-centers/find-center.
26. Alotaibi M, Long T, Kennedy E, *et al*. The efficacy of GMFM-88 and GMFM-66 to detect changes in gross motor function in children with cerebral palsy (CP): a literature review. *Disability Rehabilitation*, 2014; 617-627.