Supplemental Information

METHODS

Eligibility Criteria

As described in detail in our published protocol (Häusler M, Heussen N. Protocol for a systematic review and meta-analysis on the effect of hippotherapy and related equine-assisted therapies on motor capabilities in children with cerebral palsy. *Syst Rev*. 2020;9(1):48), randomized clinical trials and nonrandomized trials including parallel-group and crossover designs were included. We included children with perinatally acquired or congenital CP aged <18 years, studies comparing usual care with different equine-assisted treatments mostly combined with usual care and studies comparing living and artificial horse–based treatments. We restricted the language of publication to English, German, French, or Spanish.

Information Sources

We identified trials through systematic searches of PubMed, Embase, Web of Science, and Cochrane Central Register of Controlled Trials databases from their inception to the present, not imposing any restriction on publication status. We also searched ClinicalTrials.gov, the WHO International Clinical Trials Registry Platform Search Portal, and the international prospective register for systematic reviews, PROSPERO.

Search and Study Selection

We conducted our first search in May 2019 and updated our search in February 2022 before the final analyses. Based on a preliminary search we developed the following final search strategy: (hippotherapy OR "horseback riding") OR ("CP" AND [saddle OR horse OR ride OR riding OR equine OR hippotherapeutic]), which was adapted to each database. Both authors independently screened the full texts, identified studies for inclusion, and identified and recorded reasons for exclusion of the ineligible studies.

Data Collection and Data Items

We independently extracted study characteristic and outcome data from included studies, resolved disagreement by consensus, and contacted authors for missing data. We documented the study design, the total duration of the study, the number of study centers and location, the study setting, and the date of the study. We extracted data on the number of randomized, lost to follow-up, withdrawn, and analyzed patients. Moreover, we noted the distribution of age, sex, and severity of the disease as well as the eligibility criteria for our review.

We searched within each study for data on the primary outcome motor function. For this, we extracted data on the GMFM total as well as domains A to E, SAS, PBS, modified functional reach test, electromyography, force plate testing, range of motion tests for different joints, body asymmetry scores, gait analysis, (modified) Ashworth scale, and Formetric analysis. For the secondary outcome QoL, we extracted data on the PEDI, CHQ, KIDSCREEN, and the CP-QoL tool.

Risk of Bias in Individual Studies

According to the guidance document of the current version of the RoB 2 tool and the ROBINS-I tool (https://sites.google.com/site/riskofbiastool/), risk of bias for each included study was independently evaluated by the 2 reviewers. We assessed the risk of bias in randomized trials as bias arising from the randomization process, because of deviation from intended interventions, because of missing outcome data, in measurement of the outcome, and in selection of the reported result. We rated each study within each domain as having low or high risk of bias or some concerns. Risk of bias in nonrandomized trials was evaluated as bias because of confounding, in selection of participants into the study, in classification of interventions, because of deviations from intended interventions, because of missing data, in measurement of outcomes, and in selection of the reported result. Each study was graded as having low, moderate, serious, or critical risk of bias or providing no information.

Summary Measures

Continuous data describing the primary and secondary outcomes are analyzed by the MD or SMD with 95% CIs as effect estimate. If studies reported median and range, we estimated the mean and SD according to Hozo et al.[48] If studies used effect sizes instead of mean or median and SD or range, reported effect sizes were converted to SMD and corresponding standard errors according to Borenstein et al.[49] Data from trials that reported different subscores of an outcome were combined by computing a composite score as proposed by Borenstein et al[49] assuming a correlation of 0.5 among the different subscores. To include data from crossover trials in our analyses, we used the generic inverse variance approach of RevMan 5.3.3 to obtain a summary effect for studies with crossover and parallel-group designs.

Synthesis of Results

To take the unit of analysis issue in crossover trials into account, we estimated the treatment effect as mean crossover difference or standardized mean crossover difference according to Curtin et al.[50] Given the clinical heterogeneity across trials on patients with CP and their differences in comorbidities and comedications, we used random-effects meta-analyses to calculate an overall summary of average treatment effect across trials. Results are presented as the pooled meta-analytic effect estimate with corresponding 95% CI and *P* value, the estimates of the between-study variability τ^2, and I^2 to describe the percentage of the variability in effect estimates that is owing to heterogeneity.

Risk of Bias Across Studies

We assessed for each included study if a registered study protocol was available and whether the protocol was registered before the study was initialized. If available, we compared the reported outcomes against the outcomes documented in the protocol to evaluate potential reporting bias.

We intended to create and examine funnel plots to explore possible publication bias for the primary outcome by assessing funnel plot asymmetry visually and by using Egger's test at a significance level of 5% if we were able to pool more than 10 trials.

Additional Analyses

As described in our protocol as subgroup analyses, we compared the primary outcome motor function in subgroups of studies with "real" hippotherapy versus further equine-assisted treatments. Furthermore, the primary outcome was compared in subgroups of studies (1) using true hippotherapy, therapeutic riding, or artificial horses, respectively; (2) using living or artificial horses; (3) distinguishing by treatment duration of 8 weeks maximum or more; or (4) distinguishing between the GMFM-66 and GMFM-88 form of the GMFM tool.

In the case of substantial heterogeneity defined as (a) $I^2 >$ 50% and (b) inconsistency between trials in the direction or magnitude of effects (judged visually), or a *P* value $<.10$ in the Q-test for heterogeneity, we intended to perform subgroup analyses to identify possible clinical or methodological heterogeneity causing the statistical heterogeneity.

Sensitivity analyses were performed for the primary outcome by limiting analyses to studies at low risk or some concerns of bias.

SUPPLEMENTAL REFERENCES

48. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol.* 2005; 5(1):13
49. Borenstein MH. L.V., Higgins, J.P.T, Rothstein, H.R. Multiple outcomes or time-points within a study. In: Borenstein M, ed. *Introduction to Meta-Analysis. 1.* Chichester, West Sussex: John Wiley and Sons, Ltd; 2009
50. Curtin F, Altman DG, Elbourne D. Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes. *Stat Med.* 2002;21(15): 2131–2144