INPLASY

Efficacy of Birth Ball for Labour Pain: A Systematic Review and Meta-Analysis of Randomised Controlled Trials

INPLASY2025120065
doi: 10.37766/inplasy2025.12.0065
Received: 18 December 2025
Published: 18 December 2025

Tang, LB; Zhang, ZF.

Corresponding author:
Zhifu Zhang

zhangzf3195@163.com

Author Affiliation:
Shenzhen Longhua District People's Hospital.

ADMINISTRATIVE INFORMATION

Support - Not applicable.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY2025120065

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 18 December 2025 and was last updated on 18 December 2025.

INTRODUCTION

Review question / Objective This study assesses the efficacy of birth ball therapy as an adjunctive treatment for labour pain through a systematic review and meta-analysis.

Condition being studied Eight RCTs with 738 women were included.

METHODS

Participant or population Eight RCTs with 738 women were included.

Intervention n/a.

Comparator n/a.

Study designs to be included Statistical Analysis.

Eligibility criteria The inclusion criteria were constructed based on the PICOS principle, considering the five key elements: Participants, Intervention, Comparison, Outcome and Study design. In this study, the following criteria were set:
1) Pregnant women undergoing childbirth, including primiparous or multiparous women (required to be at low risk for obstetric intervention, between 37 and 42 weeks of gestation, with singleton pregnancies). Although the search strategy allowed for both vertex and breech presentations, no included study ultimately enrolled breech participants;
2) a control group comprising pregnant women receiving routine treatment without the use of birth balls;
3) outcome measures were set, including labour pain scores (e.g. Visual Analog Scale), duration of the first and second stages of labour, labour interventions (oxytocin use, induction, epidural anaesthesia), Apgar score of the newborn, mode of delivery and episiotomy;
4) inclusion of only randomised controlled trials (RCTs). The exclusion criteria primarily included the following situations: 1) duplicate data, extended

studies or identical research; 2) types of studies irrelevant to the topic, such as animal studies, case reports, literature reviews or conference abstracts; 3) studies with incomplete data or unreported outcome measures; and 4) studies using other interventions or control methods.

Information sources A total of 170 relevant records were obtained through our search strategy, with 38 from PubMed, 46 from Embase and 86 from Cochrane. After removing duplicates, both through Endnote X9 software and manual deduplication, 62 records were excluded. Subsequently, 96 records were removed based on reviewing titles and abstracts that were not relevant to the study topic.

Main outcome(s) First, the primary outcome observed was the effect of the birth ball on labour pain. Five studies reported pain scores following treatment. Among them, four studies reported pain scores when cervical dilation was 4–7 cm and four reported pain scores when cervical dilation exceeded 7 cm. Compared with the control group, women in the intervention group experienced a significant reduction in pain scores when cervical dilation exceeded 7 cm, with an MD of −1.82 (95% CI: −2.54 to −1.10; $p < 0.001$; I2 = 81%). However, there was no significant difference between the two groups in pain scores when cervical dilation was 4–7 cm, with an MD of −0.57 (95% CI: −1.95–0.82; $p = 0.42$; I2 = 90%) (Figure 3). Analysis of the first stage labour duration included 7 studies (n=432 participants) after excluding one study involving epidural analgesia. The birth ball was associated with a mean difference of -63.89 minutes (95% CI: -147.51 to 19.73; $p = 0.13$; I^2 = 93%) compared to routine care, though this result was not statistically significant. The substantial heterogeneity (I^2 = 93%) may be attributed to clinical diversity in populations (variations in parity, pain tolerance) and methodological differences (types of birth balls used, timing and duration of interventions). Meta-regression analysis did not identify significant effects of study design, parity, or ball type on this heterogeneity, suggesting other unmeasured factors may be contributing to the variability. For pain assessment, analysis of cervical dilation >7 cm included 4 studies (n=247 participants) with no epidural interventions. The birth ball group showed significantly lower pain scores (MD = -1.82, 95% CI: -2.54 to -1.10; $p < 0.001$) with substantial heterogeneity (I^2 = 81%). Analysis of cervical dilation 4-7 cm included 4 studies (n=242 participants), showing no significant difference between groups (MD = -0.57, 95% CI: -1.95 to.

Quality assessment / Risk of bias analysis For RCTs, the quality of the studies was assessed using the Cochrane Collaboration's risk of bias tool (Nordic Cochrane Centre, Cochrane Collaboration, Copenhagen, Denmark). This helped determine the internal validity of each study. When evaluating the risk of bias, several aspects were considered, including: 1) random sequence generation (selection bias); 2) allocation concealment (selection bias); 3) blinding of researchers and participants (performance bias); 4) blinding of outcome assessors (detection bias); 5) completeness of outcome data (attrition bias); 6) selective reporting (reporting bias); and 7) other sources of bias. The assessment of each aspect was categorised into three levels: 1) 'low risk', indicating that the method was correct and had no impact on outcome measurement; 2) 'high risk', indicating that the method was flawed and had an impact on outcome measurement; and 3) 'unclear', indicating that the study did not provide relevant information. In addition, funnel plots were used to assess potential publication bias, helping to identify and correct biases that may arise from small sample effects or selective reporting.

Strategy of data synthesis The meta-analysis was conducted using RevMan 5.4 (Nordic Cochrane Centre, Cochrane Collaboration, Copenhagen, Denmark). For continuous variables, mean differences (MD) with 95% confidence intervals (CIs) were employed as the statistical analysis indicators for effect sizes. For categorical variables, risk ratios (RRs) were used as the statistical analysis indicator for effect sizes. In terms of heterogeneity between studies, a random effects model was used to combine the effect sizes of each study. A comprehensive approach to heterogeneity assessment was implemented. To evaluate the heterogeneity of combined effect sizes, Q-statistic and I^2 statistics tests were conducted. In cases of substantial heterogeneity ($I^2 \geq 50\%$), we performed additional analyses including sensitivity analyses by sequential exclusion of individual studies and planned meta-regression to explore potential sources of heterogeneity. The random-effects model was consistently applied to account for between-study variability. To comprehensively assess heterogeneity, we employed multiple approaches. Beyond the standard Q-statistic and I^2 tests, we conducted meta-regression analyses using RevMan software to explore potential sources of substantial heterogeneity ($I^2 \geq 50\%$) for primary outcomes. Pre-specified covariates included study design (single-center vs. multi-center), parity (Primiparous vs. multiparous), type of birth ball (Swiss ball vs. peanut ball), and gestational age.

The significance level for meta-regression was set at α= 0.05. Additionally, we implemented the GRADE (Grading of Recommendations, Assessment, Development and Evaluations) framework to systematically evaluate the quality of evidence for all primary outcomes. Two reviewers independently assessed five domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. Evidence quality was categorized as high, moderate, low, or very low. The significance level for the meta-analysis was set at α = 0.05.

We employed a comprehensive approach to assess potential publication bias. In addition to visual inspection of funnel plots for asymmetry, we conducted formal statistical testing using Egger's regression test for outcomes with sufficient studies (n ≥ 10). For outcomes with fewer studies, we acknowledged the limited statistical power of these tests while still reporting the results for transparency.

To evaluate the potential influence of specific study characteristics or quality on the results of the meta-analysis, sensitivity analyses was conducted using RevMan software. This involved systematically excluding individual studies one at a time to observe changes in the combined effect sizes and heterogeneity. By assessing the stability of the results with and without each study, we aimed to identify any significant impact that individual studies might have on the overall findings.

Subgroup analysis Similar high heterogeneity was observed for pain scores at 4-7 cm dilation (I^2 = 90%) and second stage labour duration (I^2 = 96%), necessitating cautious interpretation of these findings, but there was only a marginal effect on the duration of the second stage of labour, with no statistically significant difference (MD = −10.57, 95% CI: −35.64 – 14.50; p =0.41; I2 = 96%) (Figure 4). The use of the birth ball did not lead to a significant difference in the use of oxytocin, induction or epidural anaesthesia compared with routine treatment, with Risk Ratio (RR): 0.97 (95% CI: 0.76–1.25;

p = 0.84, I2: 42%) and Risk Ratio (RR): 0.78 (95% CI: 0.32–1.89; p = 0.58, I2: 81%), respectively. The use of the birth ball also did not impact the risk of newborns having Apgar scores <7 (RR = 0.78 ,95% CI: 0.11–5.57; p = 0.81; I2 = 57%) (Figure 5). Regarding the mode of delivery, the use of the birth ball did not significantly affect spontaneous vaginal delivery (RR = 1.12, 95% CI: 0.90–1.38; p = 0.31; I2 = 37%), instrumental-assisted delivery (RR = 0.90, 95% CI: 0.48–1.66; p = 0.72; I2 = 26%), caesarean section (RR = 1.01, 95% CI: 0.21–4.80; p = 0.99; I2 = 65%) or episiotomy (RR = 0.82, 95% CI: 0.63–1.06; p = 0.13; I2 = 50%), with no statistically significant differences (Figure 6) . The GRADE assessment of evidence quality for all primary outcomes is summarized in Supplementary Table 3. Overall, the evidence was rated as low or very low quality, primarily due to risk of bias concerns, substantial unexplained heterogeneity, and imprecision in effect estimates.

Sensitivity analysis In sensitivity analyses, we systematically excluded each study sequentially to assess the robustness of our findings. The results demonstrated minimal changes in the overall effect estimates and their statistical significance, indicating that no single study exerted disproportionate influence on the pooled results. However, it is methodologically important to note that the substantial heterogeneity (I^2 values ranging from 81-96% for key outcomes) persisted throughout these analyses, with I^2 values remaining above 80% regardless of which study was excluded. This pattern suggests that the observed heterogeneity originates from multiple methodological and clinical sources distributed across the included studies rather than being driven by any single outlier. While the direction and magnitude of point estimates remained stable, the wide confidence intervals observed in all sensitivity analyses reflect the underlying heterogeneity and highlight the precision limitations in our current evidence base.

Country(ies) involved China.

Keywords Labour; Pain; Physical Therapy; Women.

Contributions of each author

Author 1 - Libin Tang.
Author 2 - Zhifu Zhang.