Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Heterogeneous Effects of Labor Analgesia on Labor Duration and Perineal Trauma: A Causal Forest Analysis of a Real-World Obstetric Cohort

JiaPing Lu
Jinhua People's Hospital
Heng Xu
xheng0248@sina.com
Jinhua People's Hospital
Qiulian Xu
Jinhua People's Hospital

Research Article

Keywords: epidural analgesia, labor duration, treatment effect heterogeneity, causal forest, episiotomy, cesarean delivery, oxytocin augmentation, real-world evidence

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11320016/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background.

Whether epidural analgesia prolongs labor and increases operative delivery remains debated, with randomized trials and real-world studies reaching conflicting conclusions. Both literatures estimate average effects over heterogeneous populations. We asked instead *whose* labor is prolonged by intrapartum interventions, and whether those women bear the downstream risks of perineal trauma and intrapartum cesarean.

Methods.

We analyzed a single-center prospective delivery registry of 2,111 singleton parturients who attempted vaginal birth (June 2025 onward). The primary exposure was maintained epidural analgesia versus Lamaze breathing-based analgesia (active comparator); epidural versus no analgesia served as sensitivity contrast. A secondary exposure was intrapartum uterotonic augmentation (oxytocin or artificial rupture of membranes) versus none. Adjusted average effects were estimated by cross-fitted double machine learning (continuous outcomes) and doubly robust estimation (binary outcomes). Heterogeneous effects (CATEs) on total labor duration were estimated with an honest causal forest using DML residualization, and validated by quartile calibration, calibration regression, and best linear projection; variable importance stability was assessed across five full-pipeline re-runs. Secondary obstetric outcomes were analyzed within CATE quartiles.

Results.

Epidural analgesia was associated with 125.0 minutes longer labor on average (SE 11.3, $P < 0.001$), but quartile-calibrated effects ranged from 58.2 (Q1) to 273.3 minutes (Q4; Q4–Q1 difference 215.1, $P < 0.001$). Maternal age, gestational age at delivery, and gestational age at prenatal registration were the dominant effect modifiers, ranking in the top three in all 15 stability re-runs. Women in the lowest CATE quartile had lower adjusted episiotomy rates than Lamaze controls (− 14.2 percentage points [pp]), whereas women in the highest quartile had higher rates (+ 13.1 pp; Q4–Q1 difference 27.3 pp, $P = 0.010$) and a trend toward more intrapartum cesarean (+ 7.0 pp, $P = 0.101$). An independent analysis of uterotonic augmentation reproduced the heterogeneity pattern (Q1 7.5 vs Q4 109.4 minutes, $P < 0.001$).

Conclusions.

The labor-prolonging association of intrapartum interventions is strongly heterogeneous and predictable from routine prenatal characteristics. The magnitude of prolongation itself stratifies downstream

perineal-trauma risk. Reframing intrapartum intervention research around heterogeneity rather than averages may inform individualized labor management; prospective validation is required.

Introduction

Neuraxial labor analgesia, primarily epidural analgesia, has become a cornerstone of modern obstetric care and is now offered to the majority of parturients in many settings [1]. Yet the relationship between epidural analgesia and labor outcomes remains one of the most persistently debated questions in obstetrics. Randomized trials and their meta-analyses generally report a modest prolongation of the second stage of labor but no increase in cesarean delivery [2], and trials of contemporary low-concentration regimens report still smaller effects [3], whereas observational studies in real-world settings repeatedly associate epidural use with longer labors, more operative vaginal deliveries, and higher rates of intrapartum cesarean [1, 4]. In China, where cesarean rates historically ranked among the highest worldwide and labor analgesia has been promoted nationally as part of cesarean-reduction efforts [5], the question is especially salient. This discordance is often attributed to confounding by indication—women with longer, more painful labors are more likely to request and receive analgesia—but it may also reflect something more fundamental: an average treatment effect, estimated over a heterogeneous obstetric population, may be the wrong summary of a deeply individual physiological response.

A growing methodological literature argues that for most clinical decisions, the policy-relevant quantity is not the average effect but the conditional average treatment effect (CATE)—how the effect of an intervention varies with patient characteristics [6–9]. This reframing is particularly apt for labor analgesia, because the clinical harm attributed to epidurals is mediated through prolonged labor: prolonged second stage is associated with maternal and neonatal morbidity [10], and longer first stage with postpartum hemorrhage [11, 12]. It is therefore the woman whose labor is substantially prolonged who plausibly faces increased risks of perineal trauma, operative delivery, and intrapartum cesarean, whereas a woman whose labor is minimally affected may enjoy the benefits of pain relief with little downstream cost. If effect heterogeneity exists and is predictable from routine prenatal characteristics, it could inform individualized intrapartum management—yet we are unaware of any study that has formally estimated and validated heterogeneity of the labor-prolonging effect of analgesia in a real-world cohort.

In this study we apply honest causal forests [7, 13], combined with doubly robust cross-fitted nuisance estimation, to a prospective single-center cohort of women who attempted vaginal birth. Causal forests are an ensemble, nonparametric method designed specifically to estimate CATEs without prespecifying subgroups, and their honest sample-splitting construction guards against overfitting-driven spurious heterogeneity. We address three questions. First, does the effect of epidural analgesia (relative to an active comparator, Lamaze breathing-based analgesia) on total labor duration vary substantially across parturients? Second, which routine prenatal characteristics—available before any intrapartum intervention—drive this heterogeneity? Third, do women with the largest predicted labor-prolonging effect

also bear the highest adjusted risk of perineal trauma (episiotomy) and intrapartum cesarean? As an independent corroboration, we repeat the analysis for intrapartum uterotonic augmentation (oxytocin or artificial rupture of membranes versus none), asking whether heterogeneity in response to intrapartum interventions is a general phenomenon or specific to analgesia.

Methods

Study design and data source

We conducted a retrospective analysis of a prospective obstetric delivery registry at Jinhua People's Hospital, Jinhua, China, including all deliveries recorded from June 2025 onward. The registry captures maternal demographics, prenatal risk assessment (five-tier color-coded risk classification per Chinese national maternal safety guidelines [14]), intrapartum management, and delivery outcomes. The study was reported following the STROBE statement for observational studies [15] and the PATH statement for predictive treatment-effect heterogeneity analyses [6]. The study protocol was approved by the Ethics Committee of Jinhua People's Hospital; the requirement for individual informed consent was waived given the use of de-identified routinely collected data.

Study population

From 3,586 deliveries recorded in the study period, we defined the analytic cohort as women who attempted vaginal birth (spontaneous vaginal delivery, instrumental delivery, or breech-assisted delivery, plus women converted to intrapartum cesarean after labor onset). Women undergoing planned (elective) cesarean without labor were excluded because intrapartum interventions and labor-duration outcomes are not defined for them. Multiple gestations (n = 11) were excluded to avoid effect heterogeneity attributable to twin physiology, yielding a final analytic sample of 2,111 singleton parturients who attempted vaginal birth (Fig. 1).

Exposures

Primary exposure (epidural analgesia). Labor analgesia was recorded as maintained epidural analgesia, Lamaze breathing-based analgesia, transcutaneous electrical nerve stimulation (TENS, n = 2), or none. The primary contrast compared maintained epidural analgesia (n = 1,732) with Lamaze breathing (n = 223). Lamaze was chosen as the primary comparator because both groups entered labor with intent to use an analgesic modality, making the comparison less susceptible to confounding by labor speed than a comparison with no analgesia (no-analgesia parturients in our registry were predominantly precipitous labors; propensity scores for the no-analgesia group were concentrated above 0.9). A contrast of epidural versus no analgesia (n = 123) was retained as a sensitivity analysis. Women receiving TENS were excluded from both contrasts.

Secondary exposure (intrapartum augmentation). Intrapartum uterotonic augmentation was defined as oxytocin infusion or artificial rupture of membranes during established labor (n = 1,156), compared with no augmentation (n = 636), consistent with common intrapartum practice [16, 17]. Women receiving cervical ripening agents (prostaglandin E2 or balloon catheter, n = 169) were excluded because cervical ripening represents a distinct, pre-labor intervention with a different causal structure.

Outcomes

The primary outcome was total labor duration (minutes from labor onset to delivery, recorded for women completing vaginal delivery; n = 1,603 in the primary epidural contrast and n = 1,501 in the augmentation analysis), interpreted against contemporary labor-curve norms [18]. Secondary outcomes were episiotomy (including mediolateral and median), second-degree or greater perineal laceration, intrapartum cesarean conversion, and intrapartum-plus-2-hour postpartum estimated blood loss, with definitions consistent with WHO intrapartum-care recommendations [19]. Because converted cesareans do not contribute a completed labor duration, analyses of cesarean conversion used all parturients in each contrast with the CATE model trained on labor duration (see below).

Covariates

All effect-modification and confounding variables were restricted to characteristics available **before the intrapartum period**: maternal age (years), gravidity (G), parity (P), gestational age at delivery (weeks + days/7), gestational age at prenatal registration (weeks), five-tier risk classification (ordinal 0–4: green/yellow/orange/purple/red), scarred uterus, accompanied labor (continuous intrapartum support [20]), hepatitis B surface antigen positivity, syphilis antibody positivity, and non-local household registration (province other than Zhejiang). We deliberately did not adjust for intrapartum variables (amniotic fluid appearance, cord status), which may lie on the causal pathway between intrapartum interventions and outcomes and controlling for which could absorb part of the true effect. Missing covariate values were imputed with the sample median.

Statistical analysis

Adjusted average treatment effects. We first estimated adjusted ATEs using double machine learning (DML [21]; partially linear model) for continuous outcomes and doubly robust (AIPW) estimation for binary outcomes, with 5-fold cross-fitting. Nuisance functions (outcome regression and propensity score) were fit with random forests (400 trees, minimum leaf size 20).

Heterogeneous treatment effects. CATEs were estimated with an honest causal forest implemented following Athey & Imbens [7] and generalized random forests [13], adapted to observational data via DML residualization: within each tree, the outcome and treatment were replaced by their residuals from cross-fitted random-forest nuisance models, and leaf-level effects were estimated as the residual-on-

residual regression coefficient. Each tree was grown on a 70% subsample, split into a structure half (used only to select splits, maximizing between-leaf effect heterogeneity) and an honest estimation half (used only to estimate leaf effects). We fit 400 trees with minimum leaf size 60, maximum depth 6, and 4 randomly selected splitting variables per node; individual CATEs were averaged across trees.

Validation of heterogeneity. We assessed estimated heterogeneity using three diagnostics. (i) *Quartile calibration*: parturients were ranked by CATE into quartiles, and within each quartile the treatment effect was re-estimated non-parametrically using cross-fitted AIPW pseudo-outcomes; heterogeneity was tested as the Q4–Q1 difference. (ii) *Calibration regression*: AIPW pseudo-outcomes regressed on CATE; a significant positive slope indicates the CATE ranking carries signal (slopes > 1 reflect honest-forest shrinkage of CATE point estimates and do not invalidate the ranking). (iii) *Best linear projection (BLP)*: AIPW pseudo-outcomes regressed on CATE plus all covariates; a joint F-test of covariate coefficients tests whether observed heterogeneity is fully captured by the CATE ranking. Variable importance was quantified as the proportion of splits on each covariate across all trees.

Stability of variable importance. To assess whether the identified effect modifiers depend on random seed, the entire pipeline (nuisance estimation and forest growth) was re-run with five independent random seeds; we report the mean and standard deviation of variable importance and the frequency with which each variable ranked in the top three.

CATE-stratified secondary outcomes. To test whether labor-prolonging effect size predicts downstream obstetric risk, the primary CATE model (trained on labor duration) was used to rank all parturients in each contrast (including converted cesareans); within CATE quartiles, episiotomy and intrapartum-cesarean effects were estimated with doubly robust methods, and the Q4–Q1 difference tested for effect modification.

Analyses were performed in Python 3.12 (scikit-learn 1.7, statsmodels 0.14); the honest causal forest was implemented in-house with full source code provided as supplementary material. Two-sided $P < 0.05$ was considered statistically significant. Following established causal-inference guidance for observational data [22], all effect estimates are interpreted as adjusted associations; because the data are observational and intrapartum interventions are indication-driven, causal language is used with the explicit caveat of potential residual confounding (notably the absence of body mass index and cervical dilation at admission).

Results

Study population

Of 3,586 deliveries recorded during the study period, 2,122 women attempted vaginal birth; after excluding 11 multiple gestations, 2,111 singleton parturients formed the analytic cohort (Fig. 1). For the primary epidural analysis, 1,955 women received maintained epidural analgesia (n = 1,732) or Lamaze breathing-based analgesia (n = 223); for the sensitivity contrast, 1,855 women received epidural or no

analgesia (n = 123). For the augmentation analysis, 1,792 women received intrapartum oxytocin or artificial rupture of membranes (n = 1,156) or no augmentation (n = 636).

Baseline imbalance between exposure groups was consistent with confounding by indication (Table 1). In the primary epidural contrast, epidural users were substantially less parous (mean parity 0.49 vs 1.15; SMD = 0.81), of lower gravidity (SMD = 0.53), more often accompanied during labor (SMD = 0.52), and less often non-locally registered (SMD = 0.30); the unadjusted labor-duration difference was very large (SMD = 1.16), underscoring the need for adjusted estimation. In the augmentation contrast, exposed women had lower parity (SMD = 0.39) and higher accompanied-labor rates (SMD = 0.43).

Table 1
Baseline characteristics and unadjusted outcome differences by exposure group (standardized mean differences, SMD)

Variable	Epidural (n = 1,732)	Lamaze (n = 223)	SMD	Augmentation (n = 1,156)	No augmentation (n = 636)	SMD
Maternal age, y	30.12	30.96	-0.17	30.08	30.51	-0.09
Gravidity	2.11	3.01	-0.53	2.07	2.60	-0.27
Parity	0.49	1.15	-0.81	0.49	0.80	-0.39
Gestational age at delivery, wk	39.38	39.30	0.06	39.44	39.11	0.21
Gestational age at registration, wk	11.78	11.89	-0.04	11.71	11.95	-0.09
Risk score (0–4)	0.84	0.99	-0.20	0.85	0.89	-0.05
Scarred uterus	0.02	0.01	0.13	0.03	0.02	0.02
Accompanied labor	0.76	0.52	0.52	0.79	0.59	0.43
HBsAg positive	0.03	0.05	-0.11	0.03	0.05	-0.10
Syphilis antibody positive	0.01	0.01	-0.04	0.01	0.01	0.00
Non-local registration	0.25	0.38	-0.30	0.25	0.30	-0.11
Outcomes (unadjusted):						
Total labor duration, min	414.1	230.6	1.16	421.5	320.6	0.56
Blood loss, mL	323.2	314.3	0.06	326.7	308.3	0.12
Episiotomy	0.38	0.17	0.47	0.40	0.24	0.35
>=2nd-degree laceration	0.42	0.55	-0.26	0.41	0.50	-0.19
Intrapartum cesarean	0.05	0.01	0.26	0.07	0.02	0.24

Adjusted average treatment effects

In the primary contrast, epidural analgesia was associated with a mean 125.0-minute longer total labor duration (SE 11.3; P < 0.001; Table 2). Adjusted average effects on episiotomy (+ 4.8 percentage points [pp], P = 0.33), estimated blood loss (+ 3.4 mL, P = 0.75), and intrapartum cesarean (+ 3.0 pp, P = 0.079)

were small and not statistically significant. In the sensitivity contrast (epidural vs no analgesia), the labor-duration association was similar (+ 117.3 min, P < 0.001), but intrapartum cesarean was paradoxically lower with epidural (− 12.0 pp, P = 0.015), consistent with strong selection of fast laborers into the no-analgesia group and reinforcing the choice of Lamaze as the primary comparator.

Table 2
Adjusted average treatment effects (ATE) by contrast and outcome (5-fold cross-fitted DML for continuous outcomes, doubly robust for binary outcomes)

Contrast	Outcome	n	ATE	SE	P
Epidural vs Lamaze (primary)	Labor duration, min	1,603	+ 125.0	11.3	< 0.001
Epidural vs Lamaze	Blood loss, mL	1,854	+ 3.4	10.9	0.752
Epidural vs Lamaze	Episiotomy	1,955	+ 0.048	0.049	0.331
Epidural vs Lamaze	>=2nd-degree laceration	1,955	-0.088	0.052	0.091
Epidural vs Lamaze	Intrapartum cesarean	1,955	+ 0.030	0.017	0.079
Epidural vs no analgesia (sensitivity)	Labor duration, min	1,520	+ 117.3	14.5	< 0.001
Epidural vs no analgesia	Blood loss, mL	1,746	+ 26.4	14.3	0.065
Epidural vs no analgesia	Episiotomy	1,855	+ 0.114	0.059	0.052
Epidural vs no analgesia	>=2nd-degree laceration	1,855	+ 0.039	0.057	0.494
Epidural vs no analgesia	Intrapartum cesarean	1,855	-0.120	0.049	0.015
Augmentation vs none	Labor duration, min	1,501	+ 53.3	8.7	< 0.001
Augmentation vs none	Blood loss, mL	1,704	+ 16.4	7.8	0.036
Augmentation vs none	Episiotomy	1,792	+ 0.067	0.024	0.005
Augmentation vs none	>=2nd-degree laceration	1,792	-0.059	0.026	0.022
Augmentation vs none	Intrapartum cesarean	1,792	+ 0.038	0.009	< 0.001

Intrapartum augmentation was associated with a 53.3-minute longer labor (SE 8.7; P < 0.001), more episiotomy (+ 6.7 pp, P = 0.005), more blood loss (+ 16.4 mL, P = 0.036), fewer second-degree-or-greater lacerations (− 5.9 pp, P = 0.022), and a higher intrapartum cesarean rate (+ 3.8 pp, P < 0.001).

Heterogeneity in the labor-prolonging effect of epidural analgesia

The honest causal forest estimated CATEs centered near the ATE (mean 129.6 min, SD 12.9) with meaningful spread (1st–99th percentile: 103–157 min; Fig. 2A). Quartile calibration confirmed large, monotone heterogeneity: within-CATE-quartile doubly robust effects increased from 58.2 min (95% CI −12.8 to 129.2) in Q1 to 158.0, 183.1, and 273.3 min (95% CI 238.6–308.0) in Q4 (Fig. 2B). The Q4–Q1 difference was 215.1 min (SE 40.3; z = 5.34; P < 0.001). Calibration regression on CATE showed a significant positive slope (β = 6.29, P < 0.001; slopes > 1 reflect honest-forest shrinkage of point estimates with preserved ranking), and the BLP joint F-test was non-significant (P = 0.265), indicating that the CATE ranking adequately captured the observable heterogeneity.

Variable importance identified three dominant effect modifiers: maternal age (27.3% of splits), gestational age at delivery (27.3%), and gestational age at prenatal registration (25.9%), followed by gravidity (11.8%); risk classification, parity, and all other covariates contributed little (Table 3). Parturients in the lowest-CATE quartile were younger (mean 28.2 y), of lower parity, and registered earliest for prenatal care (mean 10.7 weeks), whereas those in the highest quartile registered latest (12.3 weeks).

Table 3
Variable importance (split frequency) and stability across five full-pipeline re-runs

Variable	Primary contrast: mean (SD); top-3 hits	Sensitivity contrast: mean (SD); top-3 hits	Augmentation: mean (SD); top-3 hits
Gestational age at delivery	0.272 (0.011); 5/5	0.304 (0.017); 5/5	0.311 (0.006); 5/5
Maternal age	0.270 (0.005); 5/5	0.256 (0.009); 5/5	0.254 (0.011); 5/5
Gestational age at registration	0.266 (0.006); 5/5	0.279 (0.013); 5/5	0.277 (0.009); 5/5
Gravidity	0.116 (0.004); 0/5	0.098 (0.008); 0/5	0.101 (0.005); 0/5
Risk score	0.047 (0.003); 0/5	0.047 (0.004); 0/5	0.037 (0.004); 0/5
Parity	0.019 (0.003); 0/5	0.010 (0.002); 0/5	0.013 (0.002); 0/5
Accompanied labor	0.007 (0.002); 0/5	0.006; 0/5	0.004; 0/5
Non-local registration	0.004 (0.001); 0/5	0.004; 0/5	0.004; 0/5
Scarred uterus, HBsAg, syphilis	~ 0.000; 0/5	~ 0.000; 0/5	~ 0.000; 0/5

CATE-stratified downstream obstetric risk

Stratifying all parturients by their predicted labor-prolonging CATE revealed strong effect modification for secondary outcomes (Fig. 3). For episiotomy, the adjusted epidural effect increased monotonically

across quartiles: −14.2 pp (95% CI − 28.7 to 0.3) in Q1, + 10.0 pp in Q2, + 12.0 pp in Q3, and + 13.1 pp in Q4 (Q4–Q1 difference 27.3 pp; z = 2.59; P = 0.010). For intrapartum cesarean, effects were near null in Q1–Q2 (− 0.6 and − 1.0 pp) but positive in Q3–Q4 (+ 6.1 and + 6.4 pp; Q4–Q1 difference 7.0 pp; z = 1.64; P = 0.101). Thus, the same analgesic exposure was associated with lower perineal-trauma risk in women whose labors were minimally prolonged, but higher episiotomy risk—and a trend toward higher cesarean risk—in women whose labors were most prolonged.

Sensitivity contrast

In the epidural-versus-no-analgesia contrast, CATE heterogeneity was weaker (Q4–Q1 = 66.7 min, P = 0.037) and the BLP test indicated residual unexplained heterogeneity (P = 0.013), consistent with the poorer covariate overlap of the no-analgesia comparator (propensity scores concentrated above 0.9). Variable importance rankings were nevertheless concordant (gestational age at delivery 30.4%, registration timing 27.9%, age 25.6%).

Independent corroboration: intrapartum augmentation

The augmentation analysis reproduced the heterogeneity pattern (Fig. 4). CATEs averaged 54.0 min (SD 9.6); quartile-calibrated effects increased from 7.5 min (Q1) to 109.4 min (Q4) (Q4–Q1 = 101.9 min; z = 4.10; P < 0.001), with a non-significant BLP test (P = 0.755). Gestational age at delivery was the strongest modifier (30.6% of splits), followed by registration timing (26.8%) and age (24.5%). The augmentation effect on intrapartum cesarean, however, showed no significant modification by CATE quartile (Q4–Q1 = − 0.7 pp; P = 0.81).

Stability of variable importance

Across five independent full-pipeline re-runs per analysis (15 forests in total), the top-three modifiers were identical in every run—gestational age at delivery, maternal age, and registration timing—although their within-trio order occasionally swapped (Table 3). Coefficients of variation for the top variables were 1.8–5.5%; gravidity ranked fourth in all runs.

Discussion

Principal findings

In this real-world cohort of women attempting vaginal birth, the association of epidural analgesia with labor duration was strongly heterogeneous: validated, cross-fitted effect estimates ranged from approximately one hour in the least-affected quartile to more than four and a half hours in the most-affected quartile. Three routine prenatal characteristics—maternal age, gestational age at delivery, and timing of prenatal registration—robustly identified the strata, with rankings stable across repeated re-runs of the full estimation pipeline. Critically, the magnitude of the labor-prolonging effect itself predicted downstream obstetric risk: women whose labors were minimally prolonged under epidural had lower episiotomy rates than comparably selected Lamaze controls, whereas women whose labors were most

prolonged had higher episiotomy rates and a trend toward more intrapartum cesarean. A parallel, independently validated heterogeneity pattern for intrapartum uterotonic augmentation suggests that individualized response to intrapartum interventions is a general phenomenon rather than an artifact of analgesia.

Comparison with previous work

Randomized trials and meta-analyses of neuraxial analgesia have generally concluded that epidurals modestly prolong the second stage without increasing cesarean delivery [2, 3]. These estimates are averages over trial populations that are, by eligibility and consent, comparatively homogeneous; trials were neither designed nor powered to characterize response heterogeneity. Observational studies, conversely, have repeatedly reported larger prolongations and more operative delivery [1, 4], findings typically discounted as confounded by indication. Our results reconcile these literatures in a specific sense: the average adjusted association in our real-world cohort (+ 125 min) exceeds trial averages, but the distribution of individual effects spans values compatible with both bodies of evidence. The debate about “the” effect of epidurals may therefore be unresolvable as posed; the estimable and clinically meaningful question is whose labor is prolonged. To our knowledge, no prior study has formally estimated and validated CATEs for labor analgesia using causal-forest methods, nor linked the predicted effect size to perineal-trauma risk.

Interpretation of the effect modifiers

The three dominant modifiers are biologically and organizationally plausible. Greater gestational age at delivery implies a more ripened cervix and higher myometrial oxytocin-receptor expression [23], plausibly limiting any analgesia-related slowing; conversely, less-mature uteri may be more susceptible to disruption of uterine activity. Maternal age tracks declining myometrial contractile efficiency and higher dystocia rates [24]. The role of registration timing is more organizational than physiological: women who register earlier have more antenatal contacts, more opportunity for risk optimization, birth planning, and education [25], and—within the Chinese maternal safety system—earlier entry into tiered management [14]. That registration timing ranked alongside the biological variables in all 15 forest re-runs, and in both exposure analyses, suggests it carries independent signal rather than proxying for a single confounder. If confirmed prospectively, this would constitute an additional, previously unquantified argument for early prenatal registration.

Clinical implications

First, individualized counseling is feasible with information available before labor: age, gestational age, and registration timing jointly stratify the expected labor-prolonging effect over a clinically meaningful range (roughly 1–4.5 hours). Second, the predicted effect size is itself an intrapartum risk marker. Because episiotomy in contemporary Chinese practice is frequently a provider decision made to expedite delivery in prolonged second stage [26, 27], the monotone episiotomy gradient we observed is consistent with perineal risk being mediated through labor prolongation rather than caused by analgesia per se; women with high predicted prolongation may warrant closer labor-progress surveillance,

conservative second-stage management, and restraint in episiotomy, in line with consensus efforts toward safe prevention of the first cesarean [28] and perineal-trauma prevention [29]. Third, for augmentation, the absence of effect modification for cesarean conversion—despite strong heterogeneity in labor duration—suggests that prolonged labor under augmentation is tolerated differently than under analgesia, a distinction that merits mechanistic study. We caution that the CATE model is not yet a clinical tool: point estimates are shrunken by honest estimation, and prospective external validation is required before any decision-support use.

Strengths

The study combines several methodological safeguards rarely used together in obstetric observational research: an active-comparator design that structurally reduces confounding by labor speed; restriction of the adjustment set to pre-intrapartum variables, avoiding adjustment for mediators; doubly robust cross-fitted estimation; honest sample splitting in the causal forest, which protects against spurious heterogeneity; three complementary heterogeneity diagnostics (quartile calibration, calibration regression, BLP); and explicit stability testing of variable importance over repeated full-pipeline re-runs. The sensitivity contrast with no analgesia served as an internal negative-design check: its paradoxical cesarean estimate (− 12 pp) made visible exactly the selection structure that the primary design was built to avoid.

Limitations

First, residual confounding is likely. We lacked body mass index, cervical dilation at admission, analgesia timing [30], and augmentation indications; although effect-modification rankings and stratified patterns are generally more robust to residual confounding than absolute effect sizes, our estimates must be interpreted as adjusted associations. Second, CATE point estimates are shrunken (calibration slopes > 1): the quartile gradient is the reliable output, not individual predictions. Third, intrapartum cesarean was uncommon (n = 112), leaving the stratified cesarean analysis underpowered; the observed Q3–Q4 elevation is hypothesis-generating. Fourth, this is a single-center cohort with high epidural penetration and center-specific practice patterns (notably episiotomy thresholds), limiting external validity. Fifth, labor duration was unavailable for converted cesareans, so the primary CATE model was trained on completed labors; while this mirrors the clinical question (management during ongoing labor), it conditions on a post-initiation event. Finally, blood loss beyond 2 hours and maternal satisfaction outcomes were not captured.

Conclusions

In a real-world cohort, the labor-prolonging association of epidural analgesia—and of intrapartum augmentation—varied widely and predictably across parturients. Routine prenatal characteristics (age, gestational age, registration timing) identified strata in which analgesia was associated with minimal prolongation and lower perineal-trauma risk, and strata in which prolongation exceeded four hours with higher episiotomy and cesarean propensity. Reframing intrapartum intervention research around

heterogeneity rather than averages may move a decades-old debate toward actionable, individualized labor management; prospective validation of the CATE-based stratification is the necessary next step.

Declarations

Ethics approval and consent to participate.

The study protocol was approved by the Ethics Committee of Jinhua People's Hospital. The requirement for individual informed consent was waived by the ethics committee given the retrospective use of de-identified routinely collected data. The study was conducted in accordance with the Declaration of Helsinki.

Consent for publication.

Not applicable.

Competing interests.

The authors declare that they have no competing interests.

Funding.

Not applicable.

Author Contribution

JL: conceptualization, data curation, formal analysis, writing—original draft. HX: supervision, methodology, writing—review and editing, project administration. QX: data curation, investigation, validation. All authors read and approved the final manuscript.

Acknowledgement

The authors thank the obstetric nursing and registry staff of Jinhua People's Hospital for their work in maintaining the delivery registry.

Data Availability

The de-identified analytic dataset and the complete analysis code supporting the findings of this study are provided as supplementary files. Further inquiries can be directed to the corresponding author.

References

1. Callahan EC, Lee W, Aleshi P, George RB. Modern labor epidural analgesia: implications for labor outcomes and maternal-fetal health. Am J Obstet Gynecol. 2023. https://doi.org/10.1016/j.ajog.2022.06.017.
2. Anim-Somuah M, Smyth RMD, Cyna AM, Cuthbert A. Epidural versus non-epidural or no analgesia for pain management in labour. Cochrane Database Syst Rev. 2018;5CD000331. https://doi.org/10.1002/14651858.CD000331.pub4.
3. Wang TT, Sun S, Huang SQ. Effects of epidural labor analgesia with low concentrations of local anesthetics on obstetric outcomes: a systematic review and meta-analysis of randomized controlled trials. Anesth Analg. 2017;124(5):1571–83. https://doi.org/10.1213/ANE.0000000000001709.
4. Rimaitis K, Klimenko O, Rimaitis M, Morkunaite A, et al. Labor epidural analgesia and the incidence of instrumental assisted delivery. Med (Kaunas). 2015;51(2). https://doi.org/10.1016/j.medici.2015.02.002.
5. Li HT, Hellerstein S, Zhou YB, et al. Trends in cesarean delivery rates in China, 2008–2018. JAMA. 2020;323(1):89–91. https://doi.org/10.1001/jama.2019.17595.
6. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. Ann Intern Med. 2020;172(1):35–45. PMID: 31711134.
7. Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. Proc Natl Acad Sci USA. 2016;113(27):7353–60. PMID: 27382149.
8. Kunzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. Proc Natl Acad Sci USA. 2019;116(10):4156–65. https://doi.org/10.1073/pnas.1804597116.
9. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. J Am Stat Assoc. 2018;113(523):1228–42.
10. Aboukhater D, Hayasaka M, Furukawa N, et al. Evaluating the impact of the second-stage and pushing duration on maternal and neonatal outcomes: a systematic review and meta-analysis. Am J Obstet Gynecol. 2025. https://doi.org/10.1016/j.ajog.2025.07.014.
11. Ladfors LV, Liu X, Sandstrom A, et al. Risk of postpartum hemorrhage with increasing first stage labor duration. Sci Rep. 2024;14:22361. https://doi.org/10.1038/s41598-024-72963-2.
12. Huang CR, Xue B, Gao Y, et al. Incidence and risk factors for postpartum hemorrhage after vaginal delivery: a systematic review and meta-analysis. J Obstet Gynaecol Res. 2023;49(5):1276–91. https://doi.org/10.1111/jog.15654.
13. Athey S, Tibshirani J, Wager S. Generalized random forests. Ann Stat. 2019;47(2):1148–78.
14. National Health Commission of. the People's Republic of China. Standard for pregnancy risk assessment and management. Beijing: NHC; 2017.

15. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. Lancet. 2007;370(9596):1453–7. PMID: 18064739.
16. Costley PL, East CE. Oxytocin augmentation of labour in women with epidural analgesia for reducing operative deliveries. Cochrane Database Syst Rev. 2013;7CD009241. https://doi.org/10.1002/14651858.CD009241.pub3.
17. Smyth RMD, Alldred SK, Markham C. Amniotomy for shortening spontaneous labour. Cochrane Database Syst Rev. 2013;1CD006167. https://doi.org/10.1002/14651858.CD006167.pub3.
18. Zhang J, Landy HJ, Branch DW, et al. Contemporary patterns of spontaneous labor with normal neonatal outcomes. Obstet Gynecol. 2010;116(6):1281–7. https://doi.org/10.1097/AOG.0b013e3181fdef6e.
19. World Health Organization. WHO recommendations: intrapartum care for a positive childbirth experience. Geneva: WHO; 2018.
20. Bohren MA, Hofmeyr GJ, Sakala C, et al. Continuous support for women during childbirth. Cochrane Database Syst Rev. 2017;7CD003766. https://doi.org/10.1002/14651858.CD003766.pub6.
21. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J. Double/debiased machine learning for treatment and structural parameters. Econom J. 2018;21(1):C1–68.
22. Hernan MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2020.
23. Wathes DC, Borwick SC, Timmons PM, et al. Oxytocin receptor expression in human term and preterm gestational tissues prior to and following the onset of labour. J Endocrinol. 1999;161(1):143–51. https://doi.org/10.1677/joe.0.1610143.
24. Waldenstrom U, Ekeus C. Risk of labor dystocia increases with maternal age irrespective of parity: a population-based register study. Acta Obstet Gynecol Scand. 2017;96(8):1008–14. https://doi.org/10.1111/aogs.13167.
25. World Health Organization. WHO recommendations on antenatal care for a positive pregnancy experience. Geneva: WHO; 2016.
26. Ye J, Chen Y, Yang H, Chen Q, et al. A nationwide cross-sectional survey of episiotomy practice in China. Lancet Reg Health West Pac. 2021;18:100345. https://doi.org/10.1016/j.lanwpc.2021.100345.
27. Jiang H, Qian X, Carroli G, Garner P. Selective versus routine use of episiotomy for vaginal birth. Cochrane Database Syst Rev. 2017;2CD000081. https://doi.org/10.1002/14651858.CD000081.pub3.
28. Spong CY, Berghella V, Wenstrom KD, Mercer BM, Saade GR. Preventing the first cesarean delivery: summary of a joint Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-Fetal Medicine, and American College of Obstetricians and Gynecologists workshop. Obstet Gynecol. 2012;120(5):1181–93. PMID: 23090537.
29. Okeahialam NA, Sultan AH, Thakar R. The prevention of perineal trauma during vaginal birth. Am J Obstet Gynecol. 2023. https://doi.org/10.1016/j.ajog.2022.06.021.

30. Sng BL, Leong WL, Zeng Y, et al. Early versus late initiation of epidural analgesia for labour. Cochrane Database Syst Rev. 2014;10CD007238. https://doi.org/10.1002/14651858.CD007238.pub2.
31. Li HT, Luo S, Trasande L, et al. Geographic variations and temporal trends in cesarean delivery rates in China, 2008–2014. JAMA. 2017;317(1):69–76. PMID: 28030701.

Figures

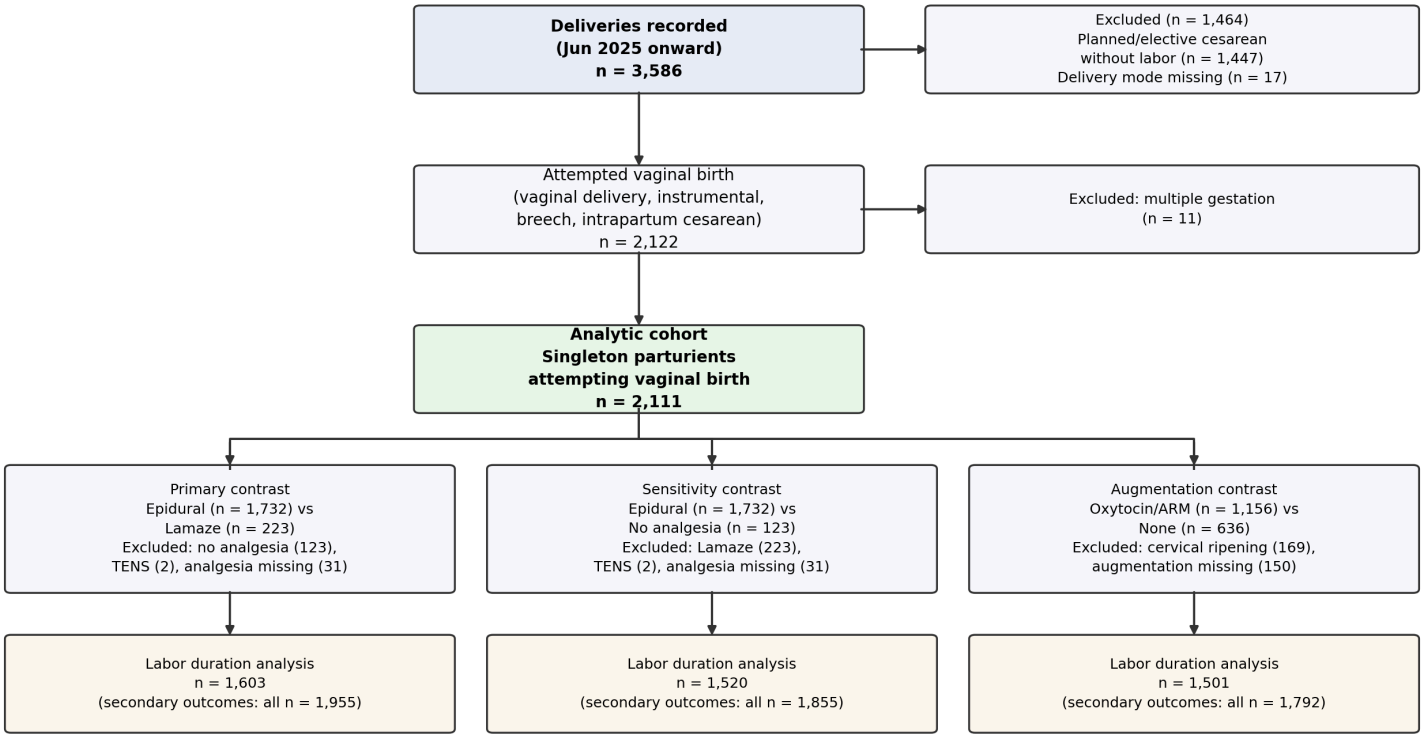


Figure 1. Study flow diagram. ARM = artificial rupture of membranes; TENS = transcutaneous electrical nerve stimulation.

Figure 1

• Study flow diagram.

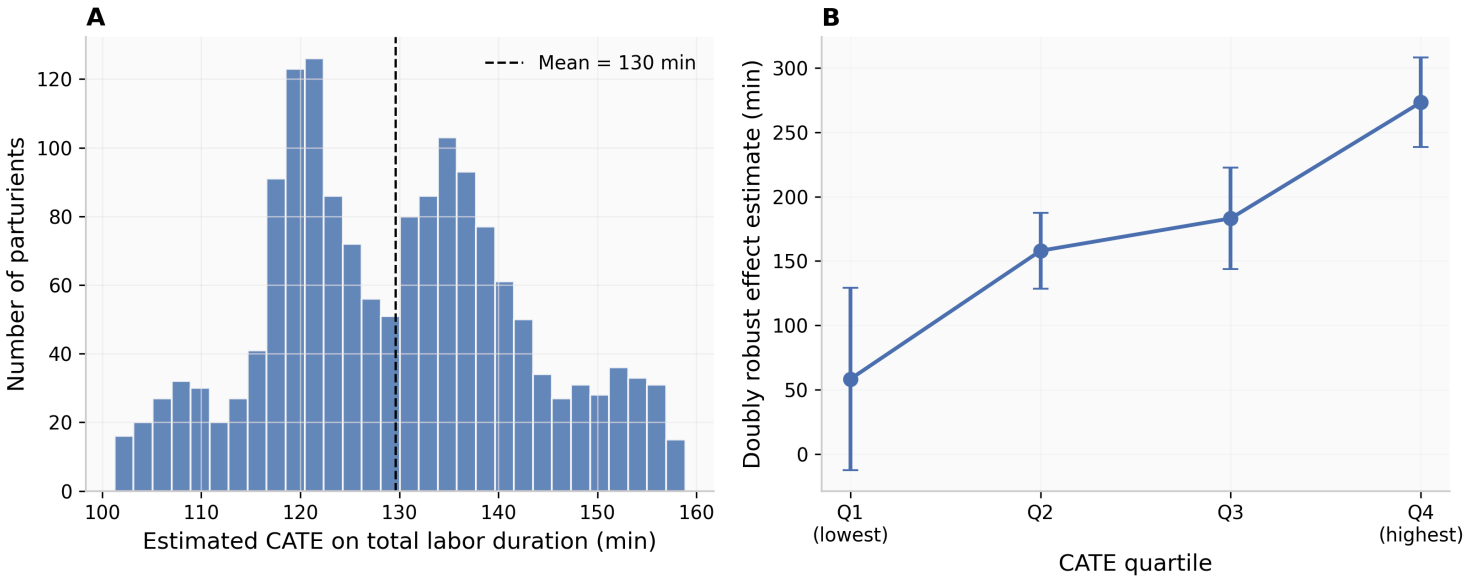


Figure 2

(A) Distribution of estimated CATEs of epidural analgesia on total labor duration (primary contrast, epidural vs Lamaze); (B) quartile calibration: doubly robust effect estimates within CATE quartiles with 95% confidence intervals.

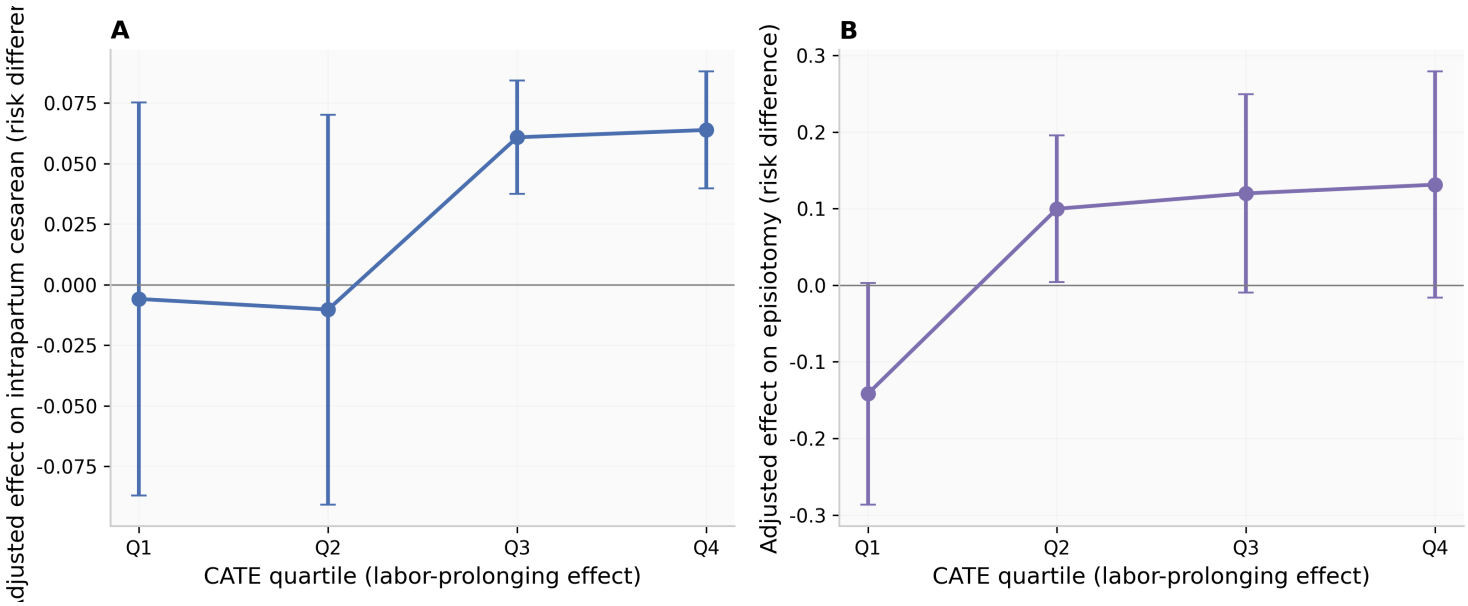


Figure 3

• CATE-stratified doubly robust effects of epidural analgesia on (A) intrapartum cesarean and (B) episiotomy, with 95% confidence intervals.

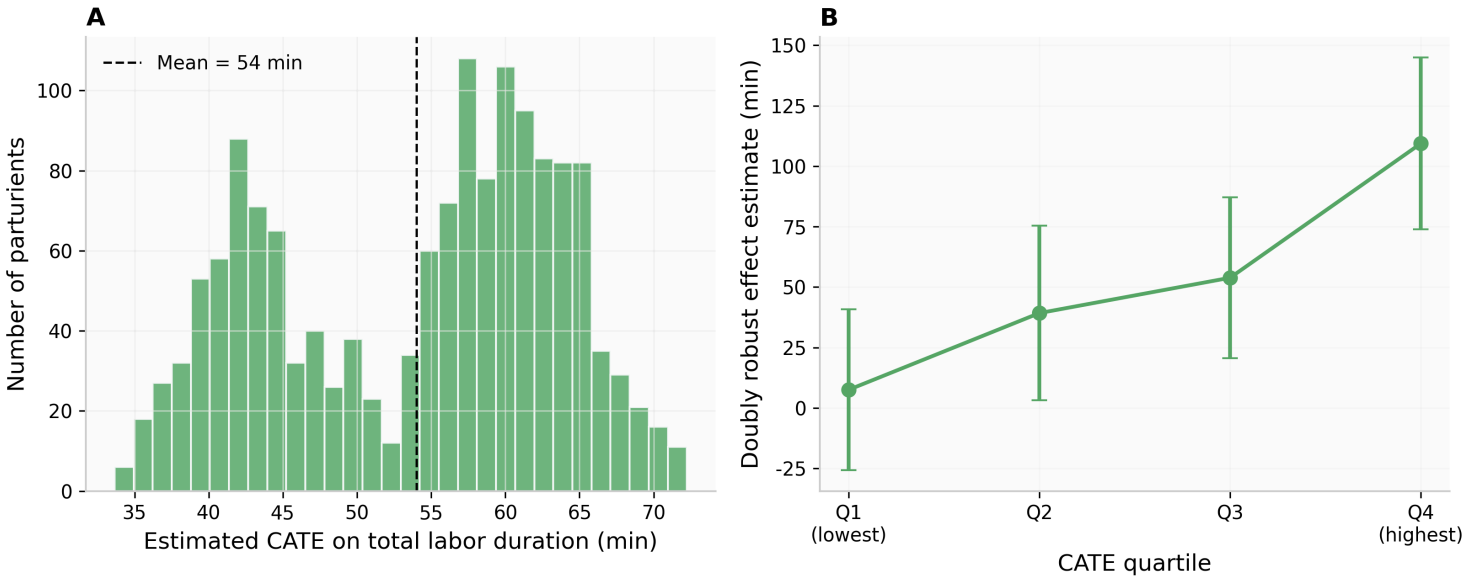


Figure 4

• (A) Distribution of estimated CATEs of intrapartum uterotonic augmentation on total labor duration; (B) quartile calibration with 95% confidence intervals.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1VIstability.png
- SupplementaryFileS3Analysiscode.py
- SupplementaryFileS4VIstability.xlsx
- SupplementaryFileS2Dataset.xlsx