Title
New labor curves of dilation and station to improve the accuracy of predicting labor progress.

Journal
American Journal of Obstetrics & Gynecology, 231(1)

Authors
Hamilton, Emily
Zhoroev, Tilekbek
Warrick, Philip
et al.

Publication Date
2024-07-01

DOI
10.1016/j.ajog.2024.02.289

Peer reviewed

Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Am J Obstet Gynecol. Author manuscript; available in PMC 2024 July 30.

Published in final edited form as:
Am J Obstet Gynecol. 2024 July ; 231(1): 1–18. doi:10.1016/j.ajog.2024.02.289.

New labor curves of dilation and station to improve the accuracy of predicting labor progress

Emily F. Hamilton, MD,
Tilekbek Zhoroev, MSc [PhD Candidate],
Philip A. Warrick, PhD,
Adi L. Tarca, PhD,
Thomas J. Garite, MD,
Aaron B. Caughey, MD, PhD,
Jason Melillo, MD,
Mona Prasad, DO, MPH,
Duncan Neilson, MD,
Peter Singson, MD,
Kimberlee McKay, MD,
Roberto Romero, MD, DMedSc

Faculty of Medicine and Health Sciences, Department of Obstetrics and Gynecology, McGill University, Montreal, Quebec, Canada (Dr Hamilton); PeriGen, Inc, Cary, NC (Dr Hamilton, Mr Zhoroev, Drs Warrick, and McKay); Faculty of Science, Department of Applied Mathematics, North Carolina State University, Raleigh, NC (Mr Zhoroev); Faculty of Medicine and Health Sciences, Department of Biomedical Engineering, McGill University, Montreal, Quebec, Canada (Dr Warrick); Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI (Dr Tarca); Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, MI (Dr Tarca); Department of Computer Science, Wayne State University College of Engineering, Detroit, MI (Dr Tarca); Department of Obstetrics and Gynecology, University of California, Irvine, Irvine, CA (Dr Garite); Sera Prognostics, The Pregnancy Company, Salt Lake City, UT (Dr Garite); Department of Obstetrics and Gynecology, Oregon Health & Science University School of Medicine, Portland, OR (Dr Caughey); Department of Obstetrics and Gynecology, OhioHealth, Columbus, OH (Dr Melillo); Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, OhioHealth, Columbus, OH (Dr Prasad); Women's Health Services, Legacy Health, Portland, OR (Drs Neilson and Singson); Sanford School of Medicine at the University of South Dakota, Vermillion, SD (Dr McKay); Perinatal Quality and Obstetrics and Gynecology Service Line, Avera Health, Sioux Falls, SD (Dr McKay); Pregnancy Research Branch, Division of Obstetrics and Maternal-Fetal Medicine, Division of

Corresponding authors: Emily F. Hamilton, MD.emilyhamilton2525@outlook.com; Roberto Romero, MD, DMedSci.romeror@mail.nih.gov.

The authors report no conflicts of interest. E.F.H., P.A.W., T.Z., A.L.T., and K.M. report being employed by or are a paid consultant to PeriGen, Inc. A.L.T. was also supported by the Wayne State University Perinatal Initiative in Maternal, Perinatal, and Child Health. R.R. has contributed to this work as part of his official duties as an employee of the United States federal government.

PeriGen, Inc, a Halma company, provided the financial support for the infrastructure required to conduct this research.

Intramural Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, United States Department of Health and Human Services, Bethesda, MD (Dr Romero); Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI (Dr Romero); and Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI (Dr Romero).

Abstract

BACKGROUND: The diagnosis of failure to progress, the most common indication for intrapartum cesarean delivery, is based on the assessment of cervical dilation and station over time. Labor curves serve as references for expected changes in dilation and fetal descent. The labor curves of Friedman, Zhang et al, and others are based on time alone and derived from mothers with spontaneous labor onset. However, labor induction is now common, and clinicians also consider other factors when assessing labor progress. Labor curves that consider the use of labor induction and other factors that influence labor progress have the potential to be more accurate and closer to clinical decision-making.

OBJECTIVE: This study aimed to compare the prediction errors of labor curves based on a single factor (time) or multiple clinically relevant factors using two modeling methods: mixed-effects regression, a standard statistical method, and Gaussian processes, a machine learning method.

STUDY DESIGN: This was a longitudinal cohort study of changes in dilation and station based on data from 8022 births in nulliparous women with a live, singleton, vertex-presenting fetus ≥35 weeks of gestation with a vaginal delivery. New labor curves of dilation and station were generated with 10-fold cross-validation. External validation was performed using a geographically independent group. Model variables included time from the first examination in the 20 hours before delivery; dilation, effacement, and station recorded at the previous examination; cumulative contraction counts; and use of epidural anesthesia and labor induction. To assess model accuracy, differences between each model's predicted value and its corresponding observed value were calculated. These prediction errors were summarized using mean absolute error and root mean squared error statistics.

RESULTS: Dilation curves based on multiple parameters were more accurate than those derived from time alone. The mean absolute error of the multifactor methods was better (lower) than those of the single-factor methods (0.826 cm [95% confidence interval, 0.820–0.832] for the multifactor machine learning and 0.893 cm [95% confidence interval, 0.885–0.901] for the multifactor mixed-effects method and 2.122 cm [95% confidence interval, 2.108–2.136] for the single-factor methods; $P<.0001$ for both comparisons). The root mean squared errors of the multifactor methods were also better (lower) than those of the single-factor methods (1.126 cm [95% confidence interval, 1.118–1.133] for the machine learning [$P<.0001$] and 1.172 cm [95% confidence interval, 1.164–1.181] for the mixed-effects methods and 2.504 cm [95% confidence interval, 2.487–2.521] for the single-factor [$P<.0001$ for both comparisons]). The multifactor machine learning dilation models showed small but statistically significant improvements in accuracy compared to the mixed-effects regression models ($P<.0001$). The multifactor machine learning method produced a curve of descent with a mean absolute error of 0.512 cm (95% confidence interval, 0.509–0.515) and a root mean squared error of 0.660 cm (95% confidence interval, 0.655–0.666). External validation using independent data produced similar findings.

CONCLUSION: Cervical dilation models based on multiple clinically relevant parameters showed improved (lower) prediction errors compared to models based on time alone. The mean prediction errors were reduced by more than 50%. A more accurate assessment of departure from expected dilation and station may help clinicians optimize intrapartum management.

Keywords

artificial intelligence; cervical dilation; dystocia; epidural anesthesia; failure to progress in labor; fetal descent; multifactor; multivariable; labor disorders; labor progression; machine learning; mixed-effects; partogram; prediction error; rupture of membranes; station; Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD)

Introduction

The diagnosis of failure to progress in labor is the most common indication for intrapartum cesarean delivery (CD) and is based on the assessment of changes in cervical dilation and station over time.[1–3] Labor curves serve as references for expected changes in dilation and fetal descent, and they have influenced the recommendations and guidelines of professional societies. The labor curves described by Friedman,[4–7] Zhang et al,[8–10] and others are based on time alone and were derived from mothers with spontaneous onset of labor.[4–16] Given that labor induction is now common in contemporary obstetrical practice and that it influences labor progress, its use should be considered when modeling labor progress.[17–22] Similarly, other factors, such as effacement, membrane status, and epidural anesthesia, have an effect on labor progress and have not been taken into account in the creation of these labor curves.[4–16,23,24]

The accuracy and precision of these models are important and are demonstrated by a hypothetical example. Suppose that, at a point in time, a model gives an average expected dilation of 7 cm with a range between 4 cm and 10 cm (95% prediction interval). Such a model would have limited clinical use for obstetricians and midwives because expected dilation covers nearly all possible dilations. Moreover, if the assessment of normal progress is difficult with this model, the diagnosis of abnormal labor progress can be expected to be much more problematic.

This study aimed to develop and assess the accuracy of labor curves based on either a single factor (time) or multiple factors using two kinds of modeling techniques: mixed-effects regression, a standard statistical method for longitudinal data collected from the same subjects, and Gaussian processes modeling, a machine learning method.[25] Both approaches are data-driven modeling techniques, by contrast with hypothesis-driven techniques based on assumptions about the shape of the labor curve.

Material and Methods

Study design and clinical data

This was a longitudinal cohort study of changes in cervical dilation and station based on data from births in nulliparous women with a live, singleton, vertex-presenting fetus at ≥35 weeks of gestation with a vaginal delivery. Data for modeling were collected from all births between June 1, 2017, and June 30, 2021, with a live singleton fetus in vertex presentation and a gestational age of ≥35 weeks in a mix of 10 community or teaching US hospitals in Ohio. Additional inclusion criteria for modeling included nulli-parity, vaginal delivery, a 5-minute Apgar score of ≥7, use of electronic fetal monitoring with uterine activity recording, and more than one cervical examination recorded in the first stage of labor. We excluded data from births with maternal or neonatal admission to an intensive care unit, shoulder dystocia, or descending dilation. We used a standard technique called 10-fold cross-validation to develop and measure model accuracy. With this technique, accuracy statistics were based on validation data that were independent of the data used for model development. Recently, another dataset was collected for external independent validation to determine how a model created in one geographic region would perform when applied to data from a different region. These data came from consecutive deliveries between January 1, 2021, and September 15, 2021, in a mix of six community or teaching hospitals in Oregon and Washington using the same selection or exclusion criteria.

All data were deidentified. The authors obtained permission to use the data for this project from the hospital systems that provided the data. This study was considered exempt by the Institutional Review Boards at Ohio-Health (state of Ohio), Wayne State University (Detroit, MI), and Legacy Health (states of Washington and Oregon). We have followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines when describing model development and evaluation.[26]

Variables used for modeling

Time—All data were deidentified, and all times were converted to relative time. Relative time was measured in two ways: “negative-time” and “forward-time.” All labors had a recorded time of birth. Time can be measured backward (negative-time) from this point. The curves of Zhang et al[8–10] and others used this approach where $Time_0$ was the time when dilation reached 10 cm and the time of a cervical examination used for modeling was calculated backward (eg, 60 minutes before the complete dilation was −60 minutes).[8–16] After the labor curve models had been computed, the x-axis (time) was reverted to a positive value, that is, instead of being $-12 \rightarrow 0$ hours, it became $0 \rightarrow 12$ hours on the labor curve graphs. Although this approach can produce an idealized labor curve, it is not feasible to assess its accuracy prospectively because one does not know when full dilation will occur during labor. Notwithstanding this limitation, we created a basic model of dilation using negative-time to reproduce the approach used in the development of contemporary labor curves.

We chose to examine the 20 hours before delivery. Most published labor curves span 6 to 14 hours, but induced nulliparous labors can be longer.[4–16,22] Although some labors were longer than 20 hours, the numbers of women and cervical examinations at times more remote from delivery were too sparse for reliable modeling. As labor curves are meant to be used prospectively and clinicians do not know how many hours will be required to reach full dilation or birth, we used forward-time, where Time 0 was set to the time of the first recorded cervical examination within that 20-hour window. All subsequent pelvic examination times were positive (ie, the number of hours moving forward from that first examination).

Other variables

Other variables included factors known to affect the rate of dilation, namely, contractions and the cervical dilation, effacement, and fetal station reached on the previous examination.[2,3,23,24]

Contractions were identified and counted, and their duration was measured using automated analysis of the digital electronic fetal monitoring records (Peri-Watch Cues; PeriGen, Cary, NC). We did not estimate contraction strength because we did not have access to the type of sensor in place for contraction monitoring, and the use of intrauterine pressure sensors was not routine. Clinical data were extracted from the electronic medical records (EMR) (Epic; Verona, WI). The relative times of each cervical examination, membrane rupture, epidural administration, and presence or absence of labor induction were included. Missing dilation, effacement, or station values were imputed using interpolation only when there was a valid value recorded in both the preceding and following examinations.

Modeling techniques

The cervical dilation labor curves described by Zhang et al[8–10] and many others used regression modeling for repeated measures where the sole variable was negative-time.[8–14,16] Within this approach, they chose a high-order polynomial function of time to describe the course of dilation.[8–14,16] This polynomial method provided the potential for their curves to have transitions (bends) delimiting labor phases, like the Friedman curves. However, no sharp bends delimiting labor phases were found. We used the same polynomial function of time (eighth order) to generate a curve of dilation based on time alone.[9,10]

The application of machine learning is becoming more common in clinical medicine.[27,28] Neural networks, a machine learning approach in widespread use across many domains, can approximate any mathematical function to characterize complex relationships between predictors and responses. We have chosen a specific type of machine learning called Gaussian processes for several reasons.[29] This method provides a confidence interval (CI) at each prediction point that reflects model uncertainty at that state of dilation. Mixed-effects modeling produces a single and fixed CI across all dilations. A CI around the expected dilation at each examination point gives clinicians the capacity to take under consideration the range of reasonable expectations tailored to that point in labor. This characteristic seemed useful given that the time to transit 1 cm of dilation varies widely at low dilations and varies less at high dilations. In addition, this technique can model multiple processes that share common predictors that, in our case, allow us to develop separate curves for cervical dilation and fetal descent. Contractions cause both cervical dilation and fetal descent, and clinicians do consider both because high station can be a harbinger of an arrest disorder.[30–34]

The mixed-effects models with multiple variables were constructed using the Python package Statsmodel.[25] The Gaussian processes models were created and tested using the GPFlow Python package. The optimization of the models was performed on central processing units and graphics processing units using Python's SciPy optimization package and TensorFlow's Adams optimization algorithm.[25] The details of the Gaussian processes modeling used in this study have been previously described.[35] Supplementary Table S2 contains the TRIPOD Checklist.

Each model was applied repeatedly in a single-step fashion. At each examination point in time, the model predicted the current dilation and station. This approach matches clinical behavior where clinicians make decisions based on available information and not on projections far into the future.

Assessing model performance

We assessed the model accuracy by measuring prediction errors, that is, the difference between every value produced by the model and the corresponding value observed and recorded by the clinician. Smaller prediction errors indicate better accuracy. We plotted the range of prediction errors across time to see how accuracy changed over the course of labor. In addition, we summarized the accuracy for each model overall using two standard statistics: mean absolute error (MAE) and root mean squared error (RMSE). Later, the final models were tested on a geographically independent dataset to assess external validation.

Results

Clinical characteristics

A total of 49,694 women delivered a live, singleton, vertex-presenting fetus at ≥35 weeks of gestation in the parent group. From this group, 8022 births that met the selection and exclusion criteria were used for modeling and cross-validation testing. Their clinical characteristics are shown in Table 1.

The clinical characteristics of the 5528 women who delivered a live, singleton, vertex-presenting fetus at ≥35 weeks of gestation in the geographically independent dataset are presented in Table 2. Based on the study selection and exclusion criteria, 527 births were available for the external validation test.

Individual trajectories of dilation and station

Individual observed trajectories of cervical dilation and fetal station over time for a random sample of patients are displayed in Figure 1. The variability in these trajectories underscores the unpredictable nature of labor and the challenge of modeling this process. To facilitate the comparison of these figures and the subsequent figures with published average labor curves, we transformed the horizontal time axis to be relative to the time of delivery.[4–15] These trajectories illustrate the variability of labor progress, from a consistent and smooth progression to instances of abrupt change in dilation or fetal station and different labor durations and initial values of dilation and station. Although it is always possible to create a mathematical average curve of dilation or descent, the average would not represent the labor curve of many patients because there is no prominent clustering of trajectories close to it.

Cervical dilation models based on a single factor

Results based on modeling with negative-time only are displayed in the Supplementary Materials section.

Cervical dilation models based on a single factor using forward-time

Results based on modeling with forward-time as the only variable are shown in Figure 2. For display purposes only, and after the labor curve models were computed, the x-axis (forward-time) was reverted to a negative value (ie, instead of being 0 → 20 hours, it became 20 → 0 hours). The graphs in Figure 2 are drawn with the patients' predictions realigned on time of birth so that the curves with forward-time can be viewed and compared to curves published by other authors.[4–16]

The top graphs show the median expected dilation curves (black lines) for the polynomial regression and machine learning methods. The pink band shows the range from the 5th percentile to the 95th percentile of the predicted dilations. The 75th and 25th percentiles are the dashed blue and green lines. The two methods produced very similar curves.

The middle graphs shown in Figure 2 illustrate the accuracy of the models over the course of the 20 hours of labor. The vertical axis is the prediction error in centimeters. The median prediction error over 20 hours is shown by the black line. The pink band shows the range of prediction errors from the 5th percentile to the 95th percentile. In other words, 90% of all prediction errors fell within this range. The 75th and 25th percentiles are shown by the dashed blue and green lines. Both modeling methods showed similar ranges of prediction errors across time, which were notably wide. Both methods overestimated expected dilation in early labor and underestimated it in late labor. With each method, the RMSE was 2.504 cm (95% CI, 2.487–2.521), and the MAE was 2.122 cm (95% CI, 2.108–2.136). In short, when forward-time was the only variable, there was no advantage to using machine learning, and both methods showed very poor accuracy throughout labor.

Cervical dilation models based on multiple factors: 7 variables

We explored many combinations of variables and several parameters of contractility, such as the cumulative count of contractions from $Time_{0,}$ and the number and frequency of contractions in the most recent inter-examination interval. The cumulative count of contractions provided the best results. It is the contraction measure in all the multifactor models.

Figure 3 shows the results with a multifactor approach using seven explanatory variables: (1) forward-time at each cervical examination, (2) the cumulative count of contractions, (3) the presence or absence of epidural anesthesia, (4) the presence or absence of labor induction, and the recorded values at the previous examination for (5) dilation, (6) effacement, and (7) station. The median curve for the station and its associated measures of accuracy are shown in Figure 4.

The multifactor 7-variable approach gave several notable results. The mixed-effects and machine learning methods produced similar labor curves as shown in the top graphs of Figure 3. Both methods produced models that were substantially more accurate than the single-factor models based on forward-time (P<.0001) as demonstrated by the median error that is closer to 0 in Figure 3 than in those in Figure 2. Moreover, the range of prediction errors was smaller, relatively stable, and centered around 0 across time. Compared to the single-factor forward-time models, the average prediction errors were improved (lowered) by more than half. This comparison is summarized numerically in Table 3. In summary, the multifactor approaches were much more accurate than the single-factor approaches, and their improved accuracy was more stable across time.

Using the same group of seven variables, the machine learning method showed small but significant improvements compared to the mixed-effects regression method. It exhibited a lower (better) RMSE of 1.126 cm (95% CI, 1.118–1.133) compared to 1.172 cm (95% CI, 1.164–1.181) for the mixed-effects approach (P<.0001). In addition, the machine learning method produced a lower (better) MAE of 0.826 cm (95% CI, 0.820–0.832) compared to 0.893 cm (95% CI, 0.885–0.901) for the mixed-effects method (P<.0001).

Variations on the multifactor 7-variable model using machine learning

The Supplementary Materials section includes the results of additional experiments on the effect of using different combinations of predictor variables. Adding time of membrane rupture to the machine learning model (multifactor 7 variables) did not result in a significant improvement in the average prediction error. Removing the examination time variable from the model did not result in a significant change in the average prediction error.

We compared the performance of the machine learning model (multifactor 7 variables) in patients who underwent labor induction and in those with spontaneous onset of labor. The non-induced group had a slightly lower RMSE and a lower MAE, as shown in Table 4. However, the absolute size of the difference was <1 mm, which is clinically imperceptible.

External validation

The developed models (single factor with forward-time, multifactor 7 variables, and multifactor without time) were tested on the external validation dataset. Despite somewhat different background rates of CD, labor induction, and epidural use (Table 2), the results were generally very consistent with those seen in the dataset used for modeling. The RMSE and MAE results from the external validation are shown in Figure 5. The single-factor models had the highest prediction errors. The multifactor models from either the mixed-effects or machine learning method had prediction errors that were about one-half of those from the single-factor models.

When applying the multifactor 7-variable methods to these data, the machine learning method again showed a small but statistically significant improvement in the RMSE compared to the mixed-effects method (1.352 [95% CI, 1.351–1.353] vs 1.378 [95% CI, 1.377–1.379]; P<.0001). The MAEs were better (lower) in the machine learning method than in the mixed-effects method (0.946 [95% CI, 0.942–0.946] vs 1.020 [95% CI, 1.020–1.021]; P<.0001).

The effect of removing time was slightly different from that observed in the cross-validation dataset. There was a small but statistically significant deterioration in RMSE and MAE when time was removed for both the mixed-effects and the machine learning methods.

A comparison of single factor vs multifactor models in examples of normal and abnormal labor

To place the findings in a clinical context, we illustrate the application of models in a patient who had a vaginal delivery in Figure 6 and in another patient who had a CD for arrest of dilation in Figure 7.

The observed dilations and stations are listed in the tables and indicated by dots in the graphical displays. The shaded bands are the prediction intervals (5th–95th percentiles) based on the model in use. Dilation or station is predicted to be within that range 90% of the time. The solid line within the shaded band is the 50th percentile.

The prediction interval (shown by the pink band) is wide for the model solely based on time (A), whereas the prediction interval for the multifactor 7-variable (B) model is much narrower. Observed dilations that fall below this band are among the slowest 5% of expected dilations after accounting for this individual's condition concerning the seven factors in the model. In addition, the difference between an observed dilation and an expected dilation can be expressed as a percentile as shown in the table above the graphs.

Graph (C) shows the expected station based on the multifactor 7-variable model. The vertical axis is shown in the typical reversed order so that observed stations that are below the band represent better-than-expected descent.

These examples demonstrate how deviation from expected dilation can be illustrated graphically and quantified numerically and why a model with low accuracy and precision is a blunt assessment tool.

Comment

Principal findings of the study

First, cervical dilation models that incorporated seven factors demonstrated improved (lower) prediction errors compared to dilation models that were based on time alone. The mean prediction errors measured using the RMSE and the MAE were reduced by more than one-half when multiple factors were considered compared to the single-factor models based on forward-time only. Second, the machine learning method showed small improvements compared to the mixed-effects regression method. Third, testing with an external independent dataset validated the findings.

Results in the context of what is known

Reports from three continents have demonstrated remarkably similar average dilation curves for nulliparous women with spontaneous onset of labor using negative-time.[8–16] We have confirmed this general relationship using multiple variables and forward-time. On average, the rate of dilation increases gradually over time with no sharp transitions.

This study reported on machine learning techniques using multiple predictor variables, including forward-time to create labor curves of both cervical dilation and fetal descent for nulliparous labors and to measure their prediction errors. Furthermore, this study reported on labor curves that consider the use of labor induction and that span 20 hours. Models of dilation based on forward-time alone have low accuracy. Models that include multiple explanatory variables have substantially improved accuracy.

Clinical implications

A better method to assess labor progress is urgently needed. We now have information from two cluster randomized clinical trials on labor management guidelines based on the Zhang curves vs standard approaches in regions with different background CD rates ranging from 6% to 24%. Both studies reported no effect on CD rates.[36,37]

In addition, clinician compliance with these guidelines based on the Zhang curves is strikingly low. One of the prospective cluster randomized studies on labor curves included an audit of compliance with intervention guidelines.[37] The hospitals using the Zhang curve-based guidelines received pretrial educational sessions, refresher sessions, and periodic feedback during the trial. Despite these interventions, only 31% of the CDs performed at dilation of ≥6 cm for dystocia-related reasons adhered to the guidelines.[37] This number underestimates overall compliance because it does not include CDs performed before 6 cm for dystocia, which, by definition, are non-adherent to these guidelines. In addition, retrospective studies report low rates of compliance with the current American College of Obstetricians and Gynecologists guidelines in CDs performed solely for a labor progress disorder.[38–41] Low compliance suggests that clinicians consider other factors that affect dilation and descent that can lead to overriding the guidelines.

Past studies provide evidence that the arrest of dilation at ≥6 cm dilation shows almost no discrimination for other conditions, such as obstetrical hemorrhage or neonatal depression, and minimal discrimination for CD for fetal heart rate concerns.[38,42] In short, it would seem that clinicians do not find the existing labor curves germane. Guidelines based on these curves do not lower CD rates or detect labors prone to complications.[36–39] By contrast, percentile rankings from multifactor models similar to the ones presented here showed much better discrimination measures.[38]

The new approach allows the clinician to determine whether an individual patient’s course of labor progress conforms with or does not conform with the course of 95% of labors of women who delivered vaginally, taking many factors into account rather than considering only the passage of time. By considering more factors, these models are more accurate and likely to be more relevant for clinicians. Departures below the normal range of expected dilation and station shown by the multifactor models mean that the slow progress is unlikely to be explained by the patient’s state concerning those factors. We propose that this feature helps to discriminate slow progress because of dystocia from slow progress because of unfavorable conditions regarding the variables in the model, such as poor effacement or low contraction frequency.

Practical implications

The computational capacity required to apply any of the multifactor models is minimal. Calculations at each examination take <1 second using a standard laptop. Acquiring the required information automatically is feasible. All major Electronic Medical Record (EMR) platforms can export the clinical data needed. In addition, contraction signals from the fetal monitors can be obtained from existing systems and networks that relay monitoring information to central stations. We have demonstrated that it is possible to run models, such as those described herein using software that receives data from and works alongside EMRs and fetal monitoring systems. Results can be sent back to the EMR or elsewhere, such as smartphones, central nursing stations, or remote centers. These technical aspects are operational today using a previously developed and Food and Drug Administration-cleared multifactor model.

Research implications

The research described herein represents an attempt to improve the assessment of labor progress by the development of more accurate labor curves, which use clinically relevant factors and state-of-the-art modeling techniques. The findings suggest that it is possible to reduce the magnitude of the prediction error of the models. Our findings have been validated internally through cross-validation and externally in an independent group of patients. Although our results seem biologically plausible—multiple factors improved the prediction error over that derived only from time—replication of our observations is desirable. However, our findings are consistent with previous observations reported by our group, indicating that models with multiple parameters are valuable in describing the expected progress of labor.[38]

Our study has not addressed the question of discrimination of labor that progresses toward a spontaneous vaginal delivery without complications vs those that are associated with adverse perinatal outcomes. Previous attempts to address this question suggest that there is great variability in the course of normal labor and that the change in cervical dilation over time by itself does not distinguish patients with normal outcomes and those with adverse outcomes.[39]

The current definitions of labor abnormalities (protracted active phase, arrest of dilatation, and descent) were derived by selecting thresholds of time using simple statistical methods, such as identifying slow labor progress as 2 standard deviations below the mean or as the 5th percentile. However, at this point, there is no evidence that such a cutoff alone represents the optimal criteria to identify the patients who may benefit from intervention.[39]

Strengths and limitations

The strengths of this study include the use of a large contemporary clinical dataset from 16 hospitals, automated measurement of uterine activity, and access to the extensive computing infrastructure required to train complex machine learning models. Another strength is related to the measurement of time. The labor curves reported herein span 20 hours, which is longer than most labor curves reported to date, and are relevant to contemporary practice where labor induction is increasingly common and the labor of induced nulliparous women generally exceeds the 6 hours depicted in the dilation curves that have been used in the guidelines of professional guidelines.[2,3,10] Importantly, we have used forward-time rather than negative-time to build our labor curves. This is because clinicians do not know how many hours it will take for a given patient to reach full dilation. The use of forward-time and a long time span imposes substantial challenges to the modeling of labor progress. However, these issues are clinically relevant because obstetricians and midwives face these challenges in clinical practice.

The limitations of this study and all other studies that produce a labor curve relate to the challenges of data selection and measurement of the variables. All studies reporting labor curves are based on data from individuals who had a vaginal birth. In particular, they are based on a select group of patients. Given the widely different CD rates, a group of patients who deliver vaginally could represent 50% to 90% of women. The applicability of the models reported herein could be different if the anthropometric and medical characteristics of patients in other geographical regions differ. To provide perspective and context on these issues, we have reported the clinical characteristics and CD rates for the study groups. The rates of labor induction in our study groups were near the high end of the reported rates across the United States.[19] Labor induction rates in low-risk mothers can vary 4-fold by state. However, an individual woman either has a labor induction or does not. This option affects labor and underlines the importance of having a model that adjusts well for the presence or absence of labor induction rather than assuming “one size fits all.”

The clinical determination of dilation, effacement, and station is inherently imprecise and is measured irregularly over time. The precise timing when a patient attains a specific dilation, effacement, or station is largely unknown. Despite these limitations, the MAE of the multifactor models was <1 cm, which is similar to that reported in studies of interclinician variation on dilation measurement.[43,44]

We did not examine the role of reported maternal race and ethnicity or maternal factors, such as body mass index (BMI). The amount of data needed for modeling grows steeply with the number of predictor variables under study to find sufficient examples of all factor combinations. The US Census Bureau tracks at least seven different race and ethnicity groups and notes that mixed backgrounds are increasing.[45] BMI is a continuous variable, and even if it were categorized, a much larger dataset would be required to characterize its effect adequately.

The multifactor models required a simple count of contractions, which was obtained by an automated method. Theoretically, contraction counts could be derived from the contraction frequency as written in medical records. We did not verify if this was feasible or as useful.

Conclusions

Future machine learning modeling will be able to account for more explanatory factors, but the requirement for data for model training will be exceedingly large. Clinicians will always have to apply clinical judgment because no labor curve can account for every relevant factor in every labor. For example, an occiput posterior position or coexisting complications, such as chorioamnionitis, will influence labor progress and inform clinical decisions. However, accounting for the 7 variables considered in these models reduces the ad hoc mental adjustments that clinicians must make when assessing labor and helps to standardize the assessment of labor progression. Moreover, the assessment can be expressed in quantitative terms, such as percentile rankings or graphical displays.

This objective quantification of labor progress can be informative during labor as a quality assessment tool or in research on the relationships between labor progress and obstetrical complications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was also supported, in part, by the Pregnancy Research Branch, Division of Obstetrics and Maternal-Fetal Medicine, Division of Intramural Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, United States Department of Health and Human Services (NICHD/NIH/DHHS), and, in part, with federal funds from the NICHD/NIH/DHHS under contract number HHSN275201300006C.

References

1. Zhang J, Troendle J, Reddy UM, et al. Contemporary cesarean delivery practice in the United States. Am J Obstet Gynecol 2010;203: 326.e1–10.
2. American College of Obstetricians and Gynecologists (College); Society for Maternal-Fetal Medicine; Caughey AB, Cahill AG, Guise JM, Rouse DJ. Safe prevention of the primary cesarean delivery. Am J Obstet Gynecol 2014;210: 179–93. [PubMed: 24565430]
3. WHO labour care guide: user's manual. World Health Organization. Available at: https://www.who.int/publications/i/item/9789240017566. Accessed September 25, 2023.
4. Friedman E. The graphic analysis of labor. Am J Obstet Gynecol 1954;68:1568–75. [PubMed: 13207246]
5. Friedman EA. Primigravid labor; a graphicostatistical analysis. Obstet Gynecol 1955;6: 567–89. [PubMed: 13272981]
6. Friedman EA, Sachtleben MR. Station of the fetal presenting part. I. Pattern of descent. Am J Obstet Gynecol 1965;93:522–9. [PubMed: 5891546]
7. Friedman EA. Abnormal labor patterns. In: Friedman EA, ed. Labor: clinical evaluation and management. New York, NY: Appleton Century Crofts; 1978. p. 63–72.
8. Zhang J, Troendle JF, Yancey MK. Reassessing the labor curve in nulliparous women. Am J Obstet Gynecol 2002;187:824–8. [PubMed: 12388957]
9. Zhang J, Troendle J, Mikolajczyk R, Sundaram R, Beaver J, Fraser W. The natural history of the normal first stage of labor. Obstet Gynecol 2010;115:705–10. [PubMed: 20308828]
10. Zhang J, Landy HJ, Ware Branch D, et al. Contemporary patterns of spontaneous labor with normal neonatal outcomes. Obstet Gynecol 2010;116:1281–7. [PubMed: 21099592]
11. Suzuki R, Horiuchi S, Ohtsu H. Evaluation of the labor curve in nulliparous Japanese women. Am J Obstet Gynecol 2010;203: 226.e1–6.
12. Inde Y, Nakai A, Sekiguchi A, Hayashi M, Takeshita T. Cervical dilatation curves of spontaneous deliveries in pregnant Japanese females. Int J Med Sci 2018;15:549–56. [PubMed: 29725244]
13. Shi Q, Tan XQ, Liu XR, Tian XB, Qi HB. Labour patterns in Chinese women in Chongqing. BJOG 2016;123(Suppl3):57–63. [PubMed: 27627599]

14. Shindo R, Aoki S, Misumi T, et al. Spontaneous labor curve based on a retrospective multi-center study in Japan. J Obstet Gynaecol Res 2021;47:4263–9. [PubMed: 34622514]
15. Oladapo OT, Souza JP, Fawole B, et al. . Progression of the first stage of spontaneous labour: a prospective cohort study in two sub-Saharan African countries. PLoS Med 2018;15:e1002492.
16. Hamilton EF, Romero R, Tarca AL, Warrick PA. The evolution of the labor curve and its implications for clinical practice: the relationship between cervical dilation, station, and time during labor. Am J Obstet Gynecol 2023;228: S1050–62. [PubMed: 37164488]
17. Osterman MJK, Hamilton BE, Martin JA, Driscoll AK, Valenzuela CP. Births: final data for 2021. Natl Vital Stat Rep 2023;72: 1–53.
18. Simpson KR, Atterbury J. Trends and issues in labor induction in the United States: implications for clinical practice. J Obstet Gynecol Neonatal Nurs 2003;32:767–79.
19. Wood R, Freret TS, Clapp MA, Little SE. Geographic variation in induction of labor at 39-weeks in low-risk nulliparas across the United States. Am J Obstet Gynecol 2023;228: S721–2.
20. Grobman WA. The role of labor induction in modern obstetrics. Am J Obstet Gynecol 2024;230(3S):S662–8. [PubMed: 38299461]
21. Sanchez-Ramos L, Levine LD, Sciscione AC, et al. Methods for the induction of labor: efficacy and safety. Am J Obstet Gynecol 2024;230(3S):S669–95. [PubMed: 38462252]
22. Harper LM, Caughey AB, Odibo AO, Roehl KA, Zhao Q, Cahill AG. Normal progress of induced labor. Obstet Gynecol 2012;119: 1113–8. [PubMed: 22569121]
23. Langen ES, Weiner SJ, Bloom SL, et al. Association of cervical effacement with the rate of cervical change in labor among nulliparous women. Obstet Gynecol 2016;127:489–95. [PubMed: 26855099]
24. Quincy MM, Weng C, Shafer SL, Smiley RM, Flood PD, Mirza FG. Impact of cervical effacement and fetal station on progress during the first stage of labor: a biexponential model. Am J Perinatol 2014;31:745–51. [PubMed: 24338118]
25. Python packaging user guide. Available at: https://packaging.python.org/en/latest/. Accessed September 25, 2023.
26. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594. [PubMed: 25569120]
27. Tarca AL, Carey VJ, Chen XW, Romero R, Drăghici S. Machine learning and its applications to biology. PLoS Comput Biol 2007;3:e116. [PubMed: 17604446]
28. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. United States Food and Drug Administration. Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. Accessed December 14, 2023.
29. Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Cambridge, MA: MIT Press; 2008:1–246.
30. Murphy K, Shah L, Cohen WR. Labor and delivery in nulliparous women who present with an unengaged fetal head. J Perinatol 1998;18: 122–5. [PubMed: 9605302]
31. Oboro VO, Tabowei TO, Bosah JO. Fetal station at the time of labour arrest and risk of caesarean delivery. J Obstet Gynaecol 2005;25: 20–2. [PubMed: 16147687]
32. Debby A, Rotmensch S, Girtler O, Sadan O, Golan A, Glezerman M. Clinical significance of the floating fetal head in nulliparous women in labor. J Reprod Med 2003;48:37–40. [PubMed: 12611093]
33. Holt RO, Diehl SJ, Wright JW. Station and cervical dilation at epidural placement in predicting cesarean risk. Obstet Gynecol 1999;93:281–4. [PubMed: 9932570]
34. Roshanfekr D, Blakemore KJ, Lee J, Hueppchen NA, Witter FR. Station at onset of active labor in nulliparous patients and risk of cesarean delivery. Obstet Gynecol 1999;93: 329–31. [PubMed: 10074972]
35. Zhoroev T, Hamilton EF, Warrick PA. Data-driven insights into labor progression with Gaussian processes. Bioengineering (Basel) 2024;11:73.

36. Bernitz S, Dalbye R, Zhang J, et al. The frequency of intrapartum caesarean section use with the WHO partograph versus Zhang's guideline in the Labour Progression Study (LaPS): a multicentre, cluster-randomised controlled trial. Lancet 2019;393:340–8. [PubMed: 30581039]
37. Wood S, Skiffington J, Brant R, et al. The REDUCED trial: a cluster randomized trial for REDucing the utilization of Cesarean delivery for dystocia. Am J Obstet Gynecol 2023;228: S1095–103. [PubMed: 37164490]
38. Hamilton EF, Warrick PA, Collins K, Smith S, Garite TJ. Assessing first-stage labor progression and its relationship to complications. Am J Obstet Gynecol 2016;214:358. e1–8.
39. Souza JP, Oladapo OT, Fawole B, et al. Cervical dilatation over time is a poor predictor of severe adverse birth outcomes: a diagnostic accuracy study. BJOG 2018;125: 991–1000. [PubMed: 29498187]
40. Alrais MA, Ankumah NE, Amro FH, et al. Adherence to consensus guidelines for the management of labor arrest disorders in a Single Academic Tertiary Care Medical Center. Am J Perinatol 2019;36:911–7. [PubMed: 30396221]
41. Escobar CM, Grünebaum A, Nam EY, et al. Non-adherence to labor guidelines in cesarean sections done for failed induction and arrest of dilation. J Perinat Med 2020;49:17–22. [PubMed: 33555148]
42. Hamilton EF, Bujold E, McNamara H, Gauthier R, Platt RW. Dystocia among women with symptomatic uterine rupture. Am J Obstet Gynecol 2001;184:620–4. [PubMed: 11262462]
43. Tuffnell DJ, Bryce F, Johnson N, Lilford RJ. Simulation of cervical changes in labour: reproducibility of expert assessment. Lancet 1989;2: 1089–90. [PubMed: 2478849]
44. Buchmann EJ, Libhaber E. Accuracy of cervical assessment in the active phase of labour. BJOG 2007;114:833–7. [PubMed: 17567418]
45. U.S. decennial census measurement of race and ethnicity across the decades: 1790e2020. United States Census Bureau. Available at: https://www.census.gov/library/visualizations/interactive/decennial-census-measurement-of-race-and-ethnicity-across-the-decades-1790-2020.html. Accessed December 14, 2023.

AJOG at a Glance

Why was this study conducted?

The standard labor curves of dilation and station are solely based on the passage of time and were derived from births with spontaneous labor onset. These curves have limitations in characterizing expected labor progress because of the highly ariable course of normal labor. We propose that labor curves using multiple clinically relevant parameters and new modeling techniques (machine learning and mixed-effects regression) could improve the prediction of labor progress.

Key findings

Cervical dilation models based on multiple clinically relevant parameters showed improved (lower) prediction errors. Their mean prediction errors were reduced by more than 50% compared to models solely based on time. The dilation curves derived with the machine learning method had better prediction errors than those from the mixed-effects regression method. A new model of descent was presented along with its prediction errors.

What does this add to what is known?

Our study reports on dilation and descent curves based on multiple variables rather than time alone, using machine learning and mixed-effects regression methods. Moreover, our study provides a method to quantify departures from expected dilation and descent. The new curves allow for individualized assessment of labor progress.

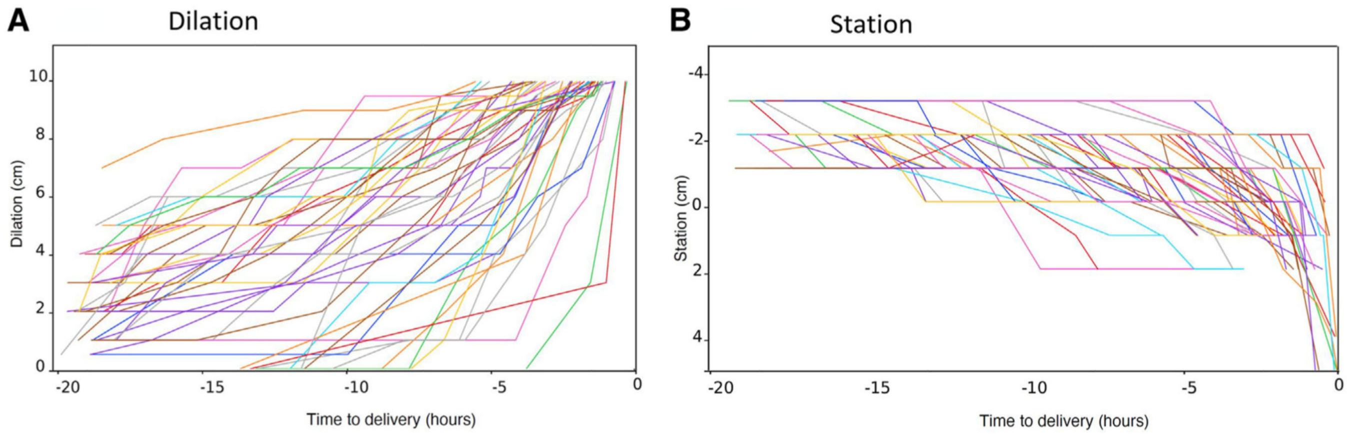

FIGURE 1. Individual trajectories over time for cervical dilation and fetal station

A, Cervical dilation. **B**, Fetal station.

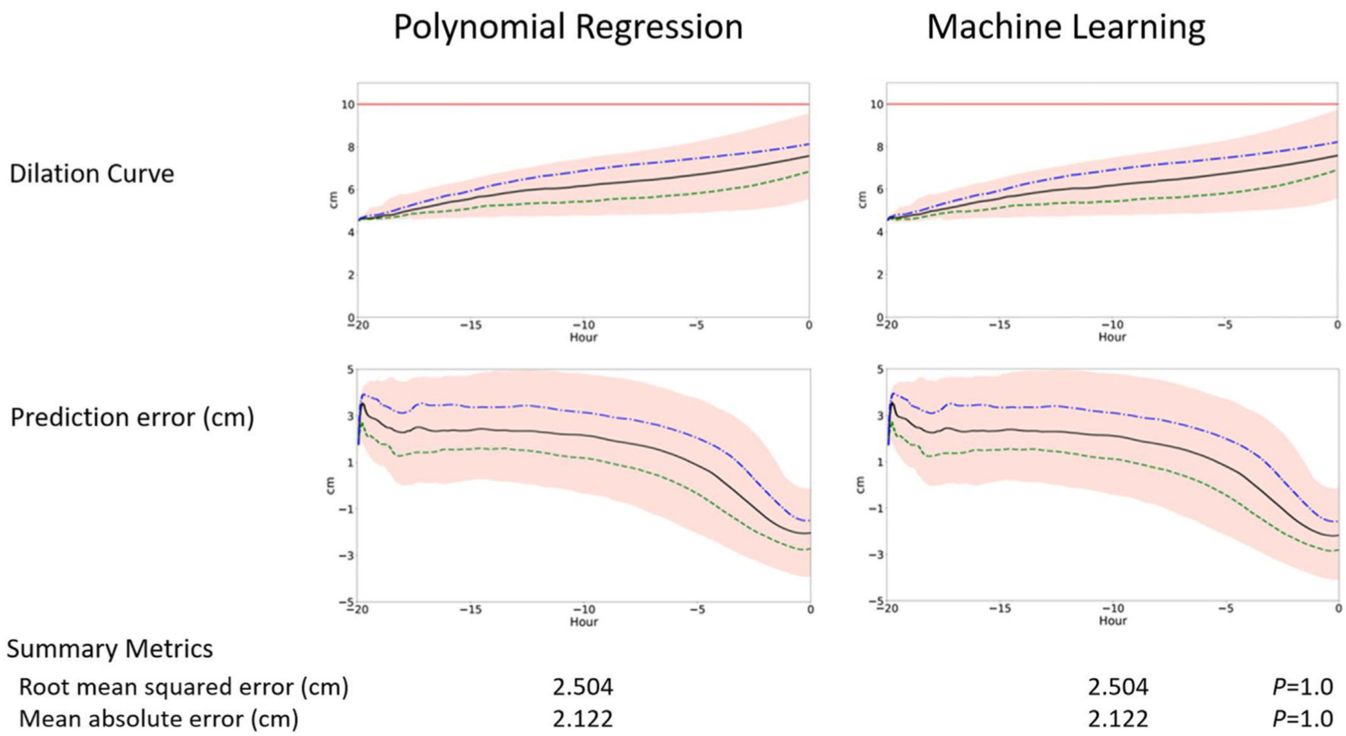

FIGURE 2. Characteristics of the dilation models based on forward-time only

The *black line* indicates the median. The *pink band* indicates the range from the 5th percentile to the 95th percentile. The *dashed blue* and *green lines* indicate the 75th and 25th percentiles.

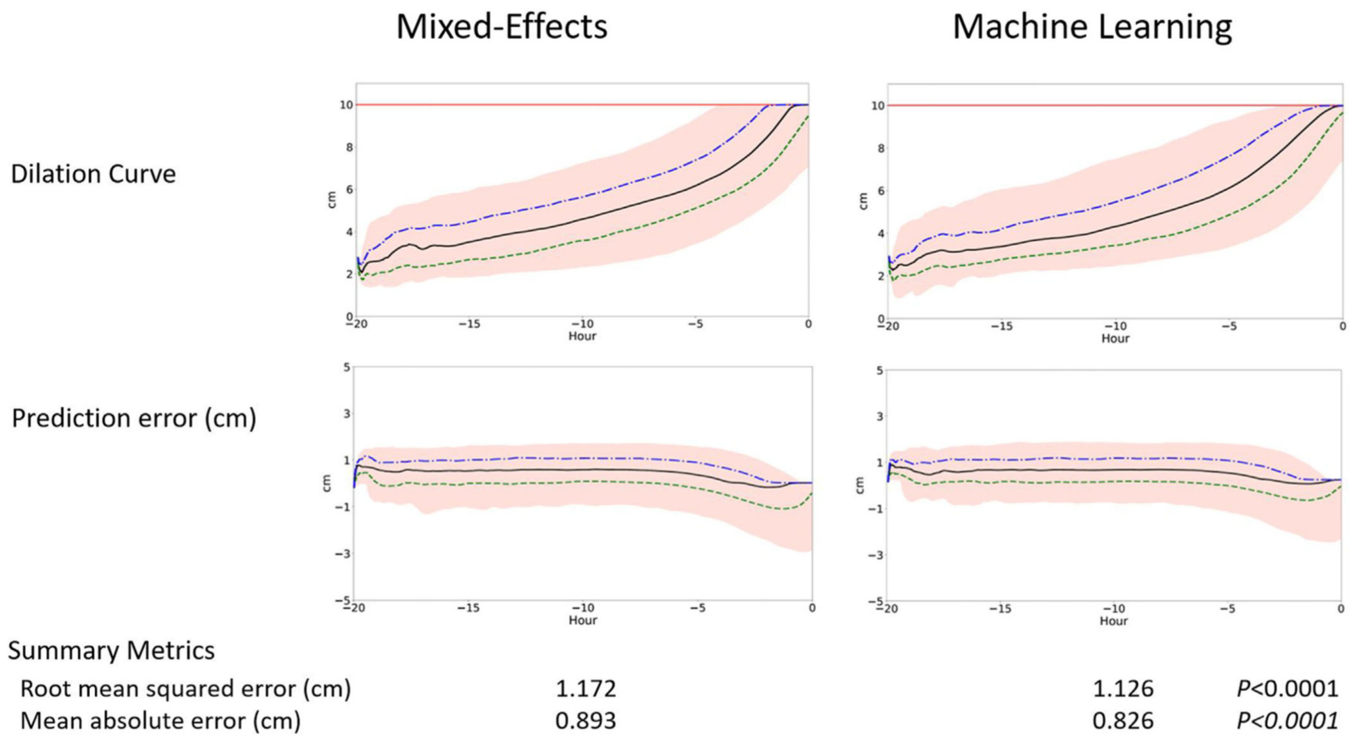

FIGURE 3. Characteristics of the dilation models based on 7 variables, including forward-time

The *black line* indicates the median. The *pink band* indicates the range from the 5th percentile to the 95th percentile. The *dashed blue* and *green lines* indicate the 75th and 25th percentiles.

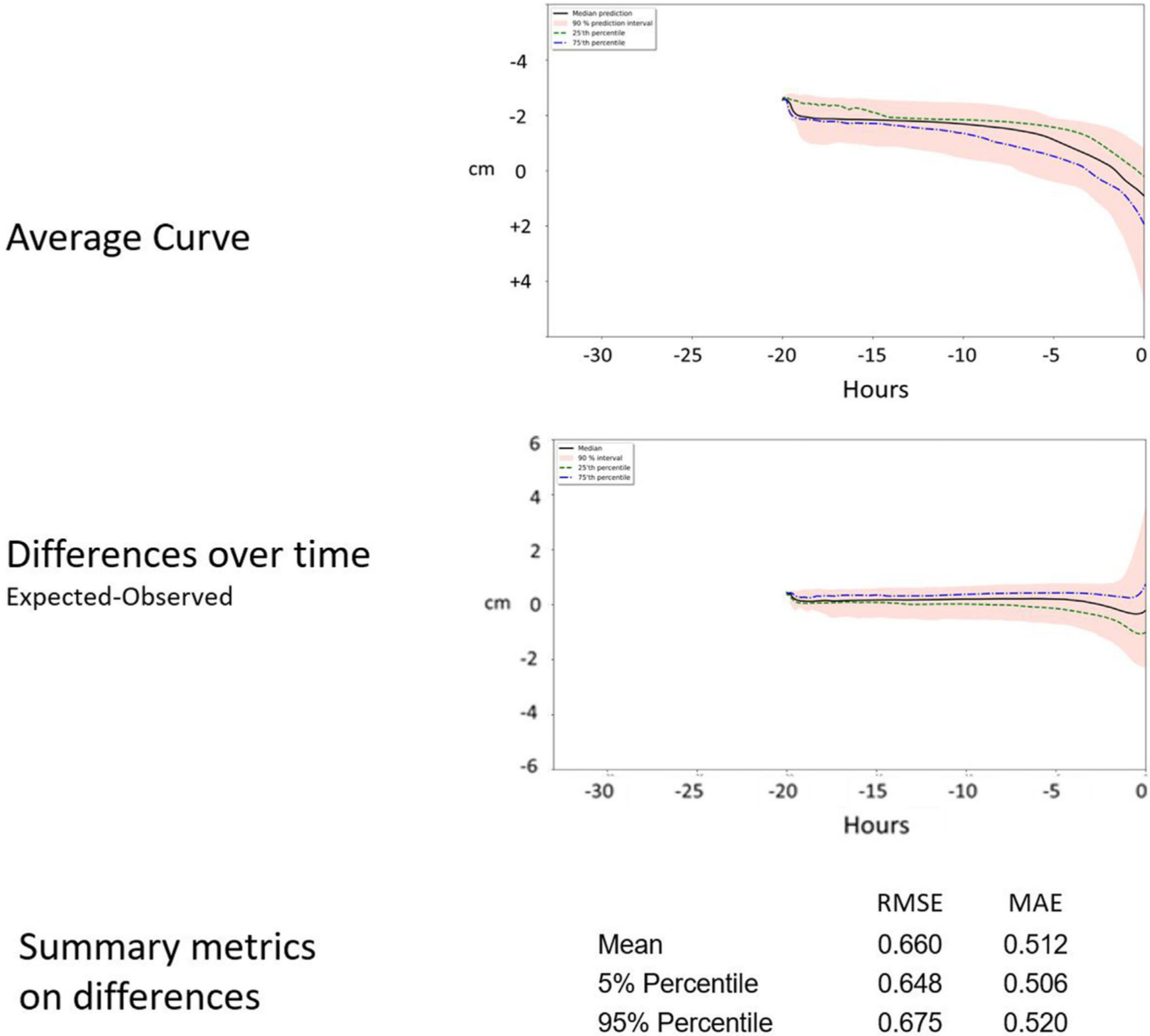

FIGURE 4. Characteristics of the station model based on 7 variables, including forward-time

The *black line* indicates the median. The *pink band* indicates the range from the 5th percentile to the 95th percentile. The *dashed blue* and *green lines* indicate the 75th and 25th percentiles.

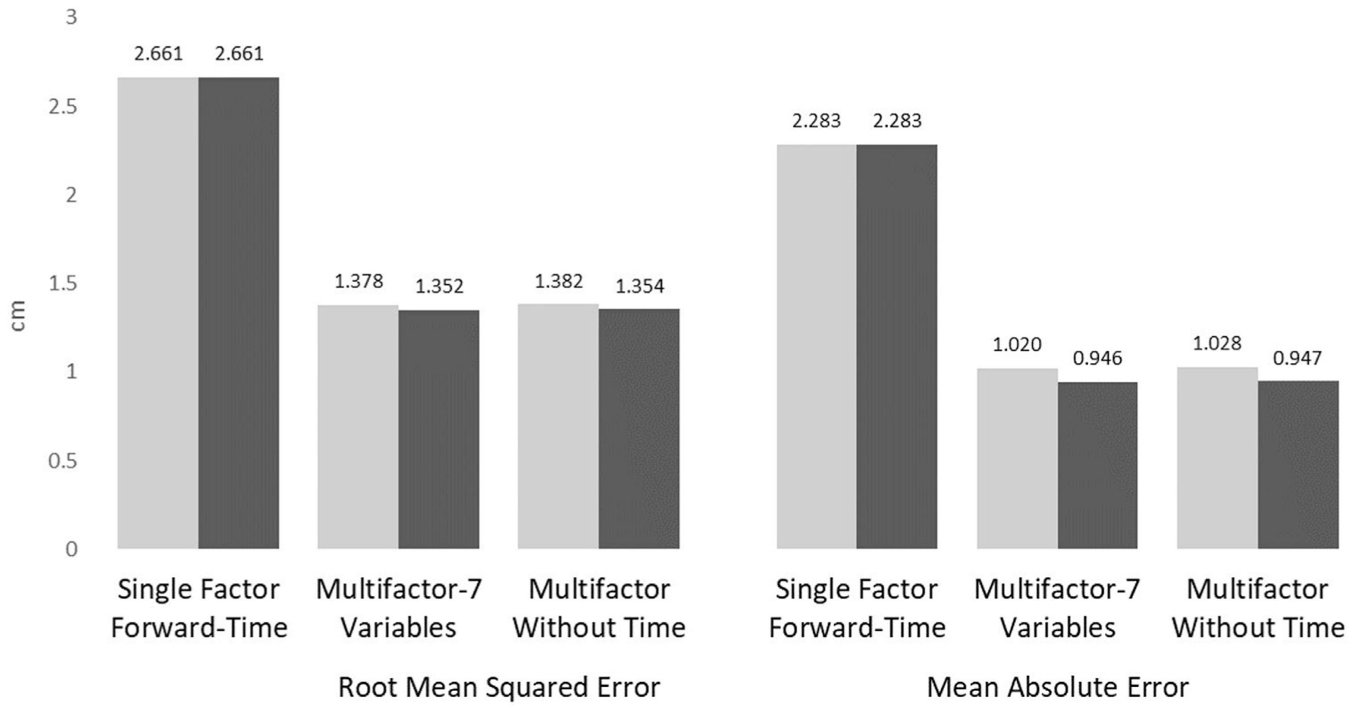

FIGURE 5. Prediction errors with external validation data

The *gray bars* indicate the mixed-effects data, and the *black bars* indicate the machine learning data.

Exam Number	Time elapsed from first exam (hours)	Dilation (cm)	Effacement (%)	Station (cm)	Duration of unchanged dilation (hours)	Average number of contractions/10 min since last exam	Percentile ranking of observed cervical dilation based on a model that uses time only (A)	Percentile ranking of observed cervical dilation based on a model that uses 7 factors (B)
1	0.0	1	50	-3	0.0			
2	1.4	1	50	-3	1.4	2.0	5.3%	2.1%
3	5.0	1	50	-3	5.0	0.1	1.5%	2.1%
4	16.2	4	100	0	0.0	0.1	6.4%	99.4%
5	17.3	7	100	1	0.0	0.4	41.4%	82.7%
6	17.8	10	100	2	0.0	2.1	86.5%	95.2%

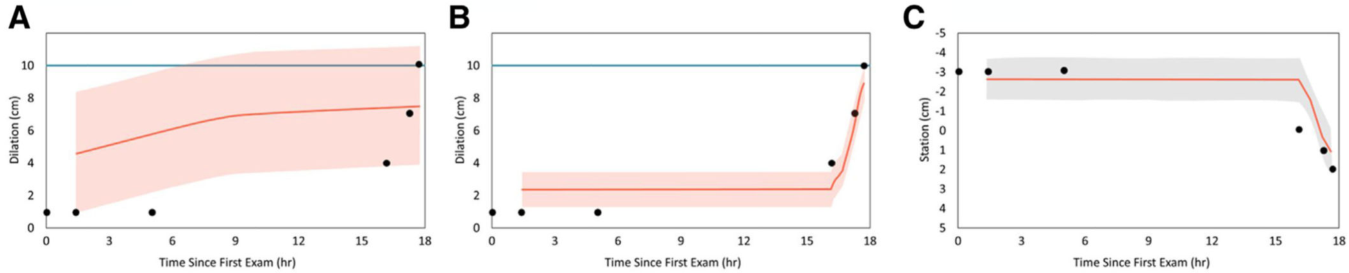

FIGURE 6. Observations from a labor ending with vaginal delivery

The *black circles* indicate the observed values of dilation or station. The *shaded bands* indicate the prediction intervals (5th–95th percentiles) based on the model in use. The *solid line* within the shaded band indicates the 50th percentile. **A,** The model of expected dilation is based on forward-time only. **B,** The model of expected dilation is based on forward-time, contractions, dilation, effacement, and station at the previous examination, presence of epidural anesthesia, and use of labor induction. **C,** The model of expected station is based on forward-time, contractions, dilation, effacement, and station at the previous examination, presence of epidural anesthesia, and use of labor induction.

Exam Number	Time elapsed from first exam (hours)	Dilation (cm)	Effacement (%)	Station (cm)	Duration of unchanged dilation (hours)	Average number of Contractions/10 min since last exam	Percentile ranking of observed cervical dilation based on a model that uses time only (A)	Percentile ranking of observed cervical dilation based on a model that uses 7 factors (B)
1	0.0	0.5		-3				
2	5.8	0.5		-3	5.8	3.3	0.6%	1.9%
3	7.9	1		-3	0.0	1.4	0.5%	7.4%
4	13.5	1		-3	5.6	1.0	0.3%	0.7%
5	17.9	2	80	-2	0.0	1.7	0.7%	16.0%
6	21.2	4	80	-2	0.0	2.7	5.1%	87.8%
7	23.8	4	90	-2	2.6	3.0	4.5%	10.1%
8	27.4	5	100	-2	0.0	2.9	7.6%	12.4%
9	30.1	6	100	-2	0.0	2.8	11.5%	6.8%
10	33.4	6	100	-2	3.3	2.6	7.1%	0.2%
11	34.9	6	100	-2	4.8	1.2	5.8%	0.2%

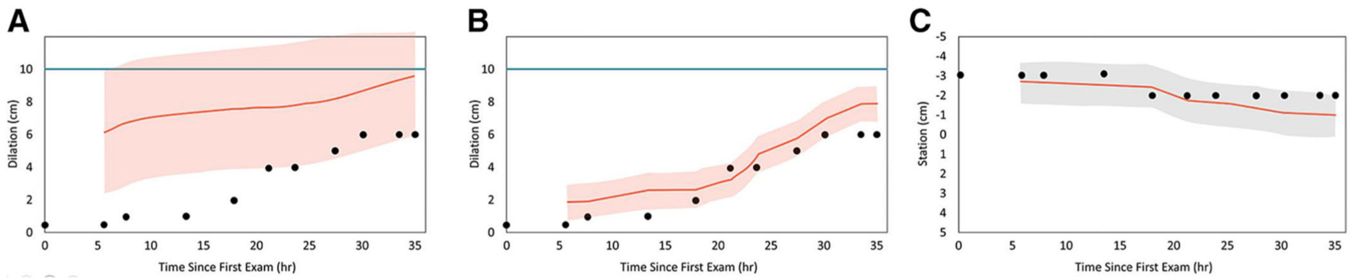

FIGURE 7. Observations from a labor with cesarean delivery with arrest of dilation

The *black circles* indicate the observed values of dilation or station. The *shaded bands* indicate the prediction intervals (5th–95th percentiles) based on the model in use. **A,** The model of expected dilation is based on forward-time only. **B**, The model of expected dilation is based on forward-time, contractions, dilation, effacement, and station at the previous examination, presence of epidural anesthesia, and use of labor induction. **C,** The model of expected station is based on forward-time, contractions, dilation, effacement, and station at the previous examination, presence of epidural anesthesia, and use of labor induction.

TABLE 1

Clinical characteristics of the parent delivery group and the subgroup used for modeling and cross-validation

Characteristic	Parent delivery group			Modeling and cross-validation subgroup		
Total number	49,694			8022		
	n	%		n	%	
Nulliparity	18,974	38.2%		8022	100.0%	By design
Induction	35,103	70.6%		5548	69.2%	
Epidural	43,457	87.4%		7402	92.3%	
Cesarean	12,945	26.0%		0	0.0%	By design
	50th percentile	25th percentile	75th percentile	50th percentile	25th percentile	75th percentile
Birthweight (g)	3348	3039	3659	3284	2996	3566
Gestational age (wk)	39.3	38.6	40.0	39.6	38.7	40.3
Latest gravid BMI (kg/m^2)	31.2	27.6	35.9	30.2	27.1	34.6

BMI, body mass index.

TABLE 2

Clinical characteristics of the other parent delivery group and subgroup used for external validation

Characteristic	Parent delivery group			Validation subgroup		
Total number	5528			527		
	n	%		n	%	
Nulliparity	2190	39.6%		527	100.0%	By design
Induction	3716	67.2%		364	69.1%	
Epidural	2935	53.1%		433	82.2%	
Cesarean	1301	23.5%		0	0.0%	By design
	50th percentile	25th percentile	75th percentile	50th percentile	25th percentile	75th percentile
Birthweight (g)	3400	3090	3720	3312	3010	3620
Gestational age (wk)	39.3	38.4	40.0	39.4	38.7	40.3
Pregravid BMI (kg/m^2)	26.3	22.7	31.7	24.0	21.2	29.2

BMI, body mass index.

TABLE 3

Comparison of prediction error metrics for the single-factor and multifactor 7-variable models for dilation, both using forward-time

Variable	**Statistical method**		**Machine learning method**	
	RMSE (cm)	**MAE (cm)**	**RMSE (cm)**	**MAE (cm)**
	Mean (95% CI)	**Mean (95% CI)**	**Mean (95% CI)**	**Mean (95% CI)**
Single-factor	2.504 (2.487–2.521)[a]	2.122 (2.108—2.136)[a]	2.504 (2.487–2.521)	2.122 (2.108–2.136)
Multifactor 7 variables	1.172 (1.164–1.181)[b]	0.893 (0.885—0.901)[b]	1.126 (1.118–1.133)	0.826 (0.820–0.842)
Percentage improvement	53.2%	57.9%	55.0%	61.1%
P value (single vs multifactor)	<.0001	<.0001	<.0001	<.0001

CI, confidence interval; *MAE*, mean absolute error; *RMSE*, root mean squared error.

[a]Eighth-degree polynomial regression

[b]Mixed-effects regression method.

TABLE 4

Comparison of error metrics in labors with and without labor induction for the multifactor 7-variable machine learning model

Variable	**Labor induction**	**No labor induction**	***P* value**
	Mean (95% CI)	Mean (95% CI)	
Root mean squared error (cm)	1.140 (1.127–1.153)	1.087 (1.070–1.104)	<.0001
Mean absolute error (cm)	0.838 (0.829–0.847)	0.794 (0.783–0.805)	<.0001

CI, confidence interval.