



HAL
open science

Prediction of cesarean delivery in class III obese nulliparous women: an externally validated model using machine learning

Massimo Lodi, Audrey Poterie, Georgios Exarchakis, Camille Brien, Pierre Lafaye de Micheaux, Philippe Deruelle, Benoit Gallix

► **To cite this version:**

Massimo Lodi, Audrey Poterie, Georgios Exarchakis, Camille Brien, Pierre Lafaye de Micheaux, et al.. Prediction of cesarean delivery in class III obese nulliparous women: an externally validated model using machine learning. *Journal of Gynecology Obstetrics and Human Reproduction*, 2023, 52 (7), pp.102624. 10.1016/j.jogoh.2023.102624 . hal-04820076

HAL Id: hal-04820076

<https://hal.science/hal-04820076v1>

Submitted on 5 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Prediction of cesarean delivery in class III obese nulliparous women: an externally validated model using machine learning

Dr. Massimo LODI M.D. ^{† 1,2}, Dr. Audrey POTERIE Ph.D., ^{† 3,4},
Dr. Georgios EXARCHAKIS Ph.D. ^{3,5}, Dr. Camille BRIEN M.D. ¹,
Pr. Pierre LAFAYE DE MICHEAUX Ph.D. ^{6,7,8}, Pr. Philippe
DERUELLE M.D. Ph.D. ¹, and Pr. Benoît GALLIX M.D. Ph.D. ^{3,5}

¹Obstetrics and Gynaecology Department, Strasbourg University Hospitals, 1 Avenue Molière, 67000 Strasbourg, France

²Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS, UMR7104 INSERM U964, Université de Strasbourg, France

³IHU Strasbourg, France

⁴Laboratoire de Mathématiques de Bretagne Atlantique (LMBA) - UMR 6205, France

⁵ICube, CNRS, University of Strasbourg, France

⁶AMIS, Université Paul Valéry Montpellier 3, France

⁷Desbrest Institute of Epidemiology and Public Health, Université de Montpellier, France

⁸PREMEDICAL - Médecine de précision par intégration de données et inférence causale, CRISAM - Inria Sophia Antipolis - Méditerranée, France

Running title: Cesarean prediction in nulliparous obese women

Statements

Funding Statement

This work was supported by French state funds managed within the “Plan Investissements d’Avenir” and by the Agence Nationale de la Recherche (ANR) (reference ANR-10-IAHU-02).

Competing interests

The authors have no competing interests to declare.

Author roles

Philippe Deruelle and Benoît Gallix conceived the study. Benoît Gallix designed the study. Massimo Lodi and Camille Brien performed the bibliographic selection and acquired the relevant data. Audrey Poterie and Georgios Exarchakis developed the model. Massimo Lodi, Audrey Poterie, and Georgios Exarchakis wrote the initial draft. Pierre Lafaye De Micheaux, Philippe Deruelle and Benoit Gallix contributed to critical discussion and manuscript finalization.

Data Availability Statement

The model will be openly available online at the following url: [xxx.xxx](#) (to be updated).

Patient data are not available.

[†] Electronic address: massimo.lodi@chru-strasbourg.fr; Corresponding author [†]Contributed equally.

Abstract

Background: class III obese women, are at a higher risk of cesarean section during labor, and cesarean section is responsible for increased maternal and neonatal morbidity in this population. *Objective:* the objective of this project was to develop a method with which to quantify cesarean section risk before labor. *Methods:* this is a multicentric retrospective cohort study conducted on 410 nulliparous class III obese pregnant women who attempted vaginal delivery in two French university hospitals. We developed two predictive algorithms (a logistic regression and a random forest models) and assessed performance levels and compared them. *Results:* the logistic regression model found that only initial weight and labor induction were significant in the prediction of unplanned cesarean section. The probability forest was able to predict cesarean section probability using only two pre-labor characteristics: initial weight and labor induction. Its performances were higher and were calculated for a cut-point of 49.5% risk and the results were (with 95% confidence intervals): area under the curve 0.70 (0.62,0.78), accuracy 0.66 (0.58, 0.73), specificity 0.87 (0.77, 0.93), and sensitivity 0.44 (0.32, 0.55). *Conclusions:* this is an innovative and effective approach to predicting unplanned CS risk in this population and could play a role in the choice of a trial of labor versus planned cesarean section. Further studies are needed, especially a prospective clinical trial. *Funding:* French state funds "Plan Investissements d'Avenir" and Agence Nationale de la Recherche.

Keywords: obesity, cesarean delivery, personalized medicine, random forests, machine learning, predictive model, predictor selection.

1 Introduction

Obesity, defined as a body mass index (BMI) ≥ 30 kg/m², is a major public health hazard. [1] It is the most common medical condition in women of childbearing age (with a prevalence of 39.7% in the United States) [2] and can have consequences for both the mother and the child during pregnancy. [3] Moreover, the prevalence of obesity among pregnant women is increasing worldwide. [4, 5] In 2009, obesity rates for pregnant women were estimated to be 14.4% in France in 2021 [Cinelli, 2022 #3822] and 16.1% in Canada. [7] In the United States, 34.9% of women are obese, and it was estimated that, in 2014, there were 1.1 million obese pregnant women. [4] Thus, obesity in pregnancy will become more challenging with time.

It has been well established that, during pregnancy, obesity is associated with increased maternal and fetal morbidity, such as gestational hypertension, gestational diabetes, and large-for-gestational-age fetuses. [5, 8-11] Moreover, obesity has an impact on delivery because it is an independent risk factor for cesarean section (CS). [9, 12-15] Indeed, obese women are at higher risk of CS delivery as compared to non-obese women. [9, 16] In addition to higher CS rates, the scientific literature shows that CS in obese women is more likely to cause maternal and neonatal morbidity. Indeed, obese women undergoing CS are at greater risk of postoperative infection and thrombosis as compared to non-obese women. [9, 17, 18] Infants of obese women more often require intensive care and have higher rates of fetal compromise and meconium-stained liquid. [9]

Obesity is defined in terms of three classes according to the World Health Organization: class I as a BMI between 30.0 and 34.9 kg/m², class II as a BMI between 35.0 and 39.9 kg/m², and class III as a BMI \geq 40 kg/m². [19] Studies suggest that obesity-related risks during pregnancy increase with BMI and complications are highest among class III obese women. Indeed, it has been shown [14] that CS risk increases proportionally to BMI (between 2 and 5% for each 1 kg/m² of BMI) and rates are > 50% in laboring women with BMI \geq 40 kg/m². Also, the odds of CS increase 3.5-fold for each BMI increase of 10 kg/m². [20]

Finally, CS-related complications seem to be higher in cases of emergency or unplanned CS as compared to elective or planned CS. [21, 22] These data indicate that unplanned CS-related complication rates could be lowered if planned CS was performed. Some studies have found that advanced maternal age, multiparity, and unfavorable cervical dilation are independent risk factors in addition to BMI for unplanned CS among class III obese pregnant women. [20] Despite the fact that the risk factors are well known, there is insufficient evidence in the scientific literature to allow clinicians to stratify the unplanned CS risk for obese women during labor and, consequently, individualize those who are at a high risk and could thus be counseled to choose planned CS before labor. Our objective was therefore to develop a method with which to quantify CS risk among a population of class III obese pregnant women before labor.

2 Methods

2.1 Study Design

This is a multicentric, retrospective cohort study conducted on 410 women in two French university hospitals (Strasbourg and Lille). It has been conducted following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [23] and the Developmental and Exploratory Clinical Investigations of DEcision support systems driven by Artificial Intelligence (DECIDE-AI). [24]

2.1.1 Eligibility Criteria

Eligibility criteria were as follows: 1) nulliparous women with BMI ≥ 40 kg/m² at the beginning of the pregnancy, 2) singleton pregnancies, 3) delivery in Strasbourg or Lille University Hospitals after 22 weeks of gestation, and 4) attempted vaginal delivery. Non-eligibility criteria were as follows: 1) stillbirth and medical interruption of pregnancy and 2) planned cesarean sections and unplanned cesarean sections before labor.

2.1.2 Data collection

In this retrospective study, data from two cohorts were used to evaluate the cesarean risk for obese nulliparous women. The first cohort includes $n_S = 247$ women who delivered at Strasbourg University Hospital between the 1st January 2009 and the 31st December 2019. The second data set contains $n_L = 163$ women who gave birth at Lille University Hospital between the 1st January 1997 and the 31st

December 2014 (published in a previous study). [16] For each patient, maternal characteristics (age, height, initial and final weight, gestational weight gain, smoking, diabetes mellitus/gestational, and high blood pressure) and delivery characteristics (labor onset, gestational age at delivery, epidural analgesia, mode of delivery) were recorded. Epidural analgesia and height were subsequently excluded from the analysis because they were not pertinent to the purposes of the model or were redundant given other information (BMI).

2.1.3 Study endpoints

The primary endpoint was the quantification of unplanned CS risk. The secondary endpoints were the performances of the developed predictive algorithm: sensitivity, specificity, and positive and negative predictive values.

2.1.4 Ethics

This work was conducted according to the ethical standards of the French Government Agency “Commission Nationale de l’Informatique et des Libertés (CNIL)” and registered with Strasbourg University Hospital’s ethics committee (21-025). It was also authorized for Lille University Hospital by the National Society of Obstetrics and Gynecology’s “Collège National des Gynécologues Obstétriciens Français (CNGOF)” Research Ethics Committee (CEROG OBS 2014-04-04).

2.2 Statistical analysis and model development

The Strasbourg cohort was used as the training set to elaborate the predictive models (as it was the one with most patient and more recent data), while the Lille cohort was used to assess the performance of the predictive models. Two predictive models were built. The first model was constructed by using classical logistic regression, and the second model was built by using the probability forest algorithm. [25] All the statistical analyses were performed in R statistical software (Version 4.1.3). All the steps of the statistical analysis are described below. Details on the development of the models are detailed in **Supplementary Materials**.

3 Results

3.1 Cohort description

In total, 410 obese women were included in both centers. The study population is described in Table 1. Among all women, 164 had unplanned CS (40%), while 246 had vaginal delivery (60%). The mean BMI was 43.6 kg/m² (range 40.0–59.6 kg/m²). In the univariate analysis, higher maternal height (p-value = 0.001) and labor induction (p-value < 0.001) were associated with unplanned CS delivery.

Before building the predictive models, we investigated whether there were differences between centers (see Table 2) and discovered that maternal age and diabetes prevalence were greater in Strasbourg (p-value = 0.004 and < 0.001, respectively). Conversely, initial weight, BMI, and high blood pressure prevalence were greater in Lille (p-value = 0.003, < 0.001, and 0.036, respectively). The delivery mode was also different: we observed a higher rate of unplanned CS in Lille as compared to Strasbourg (49.1% versus 34.0%; p-value = 0.002).

Predictive model training and testing were performed on $n = 393$ women because the following statistical analyses were conducted only on complete cases (the incomplete cases were as follows: sixteen missing values for gestational weight gain and one for smoking status). The complete cases cohort is described in Tables S1 and S2, and there were no significant differences as compared to the total cohort. The training phase was performed by using the complete cases in the

Strasbourg cohort ($n = 233$), while the performance of the predictive model was assessed based on the complete cases in the Lille cohort ($n = 160$). For convenience, in the following, the Strasbourg and Lille cohorts will always be used to refer to the complete cases in both cohorts, respectively.

3.2 Classical Predictive Model for Delivery Prediction

We first built a predictive model by following a classical approach based on a logistic regression with a stepwise variable-selection procedure. The logistic model that we ultimately obtained consisted of four predictors, including three maternal characteristics (initial weight, diabetes and age) and one delivery characteristic (labor induction). The fitted model is described in Table 3. Note that, in the logistic model, only the initial weight and labor induction seem to statistically significantly influence the risk of unplanned CS. According to this model, the unplanned CS risk decreased when the initial weight increased (OR[initial weight] = 0.97; 95%CI = 0.95 - 1.0; p-value = 0.031), whereas it increased greatly in the case of labor induction (OR[labor induction] = 3.06; 95%CI = 1.66-5.84; p-value = < 0.001). Conversely, maternal age and diabetes did not seem to impact delivery.

3.3 Probability Forest Model for Accurate Delivery Prediction

The probability forest obtained is based on only two predictors: one maternal characteristic, initial weight, and one delivery characteristic, labor induction. These two predictors represent the best subset of predictors selected by using VSURF,

[26] and they are the two significant predictors included in the logistic model. Our probability forest is built using the following tuned hyper-parameters: 100,000 trees; a minimum node size of 23; and a size of the subset of features used in the split equal to 1. Note that the probability forest error remains stable starting from 20,000 trees. Nonetheless, because we were not limited by time, because the construction of the probability forest model with the selected subset of predictors and the tuned hyper-parameters was not time consuming (execution time:<5s), we chose to use more trees so that the convergence of the probability forest error was guaranteed. In this model, the most important predictor is labor induction, followed by initial weight. According to this model, for any given value of initial weight, the risk of unplanned CS delivery seems greater in the case of labor induction (see Figure 1). The predicted CS risk does not seem to be linearly correlated with labor induction and initial weight. For women with an initial weight inferior to 113kg or superior to 130kg, the predicted CS risk seems to increase when the initial weight increases, and this augmentation seems more important in the case of labor induction. In contrast, for women with an initial weight between 113kg and 130kg, the CS risk seems to decrease when the initial weight increases, regardless of labor induction status. Note that the regression curve showing the predicted CS risk for women with no labor induction must be interpreted with caution for women with an initial weight greater than 130kg because the estimated local polynomial regression curve depends on only a single observation.

3.4 Predictive-model comparison

Table 4 and Figure 2 outline the performance of both the logistic model and the probability random forest model. The two predictive models were built to estimate and stratify individual unplanned CS risk among class III-obese pregnant women in order to counsel them during a delivery planning consultation. Each model was developed to predict the probability of unplanned CS, ranging from 0 to 1. Then, the cut-point, which represents the threshold risk at which the CS risk is considered high (and thus the model predicts a CS delivery), was determined using the Strasbourg cohort. The cut-point for each model was selected so as to maximize the sensitivity (*i.e.*, the probability of correctly predicting CS delivery among obese pregnant women who underwent CS delivery) while maintaining a specificity of at least 80% (*i.e.*, the probability of wrongly predicting CS delivery among women who had vaginal delivery is less than 20%). Thus, the optimal cut-point in the probability forest was calculated as 0.495. With this cut-point, in the probability forest, the specificity for the Lille cohort was 87%, which means that false-positive rate was 13% (*i.e.*, the risk of wrongly predicting CS delivery for women who had vaginal delivery). Conversely, the sensitivity for the Lille cohort was 44%. The optimal cut-point in the logistic model was 0.452. Based on this cut-point, the specificity of the logistic model was quite similar to the one observed with the probabilistic forest model, whereas the sensitivity for the Lille cohort was, in comparison, significantly lower (0.18). Overall, the probability forest outperformed the logistic model in terms of predictive performance levels (accuracy = 0.66 versus 0.55 ; AUC = 0.70,

i.e., 0.60). The ROC curves drawn based on the training set are included in the Appendix (see Figure S1).

4 Discussion

4.1 Principal findings

In this study, we wanted to identify a population of obese women at a high risk of unplanned CS during labor, specifically this risk being high enough to suggest a planned CS instead of a trial of labor. Classical statistical analysis failed to identify specific maternal or pregnancy characteristics important enough to determine the obstetric management of these women and accurately predict unplanned CS. Consequently, we developed a machine learning-based algorithm to stratify the risk of unplanned CS and tested it for both internal and external validation. This innovative methodology allowed us not only to stratify and predict individual unplanned CS risk for obese women during labor but also to gain insights into this risk. *Indeed, unlike the use of simple known risk factors (such as a given BMI threshold or labor induction), the use of this type of approach makes it possible to establish a quantified individual risk.* This model is easy to use in clinical practice because it only requires two parameters: initial weight and labor induction. *Therefore, it could be used via a simple interface (such as a web application) both during a consultation and at the admission in the delivery room.*

4.2 Cesarean section in obese women

In this study, we found that labor induction significantly increased the risk of cesarean section in a nulliparous patient with a BMI ≥ 40 kg/m², which is consistent with previously reported data. [13, 27-29] Interestingly, the predictive models (both logistic and random forest) found that initial weight was inversely proportional to CS risk. This can be explained by the fact that, among women with BMI ≥ 40 kg/m², those with extremely high BMI had higher rates of planned or unplanned CS before a trial of labor, and were therefore excluded. Furthermore, the predictive model also allowed us to individually quantify the unplanned CS risk with only pre-labor data, which, to our knowledge, has never been accomplished before. Nonetheless, the parameter that could limit clinical applications of such a model is false positives, i.e., women predicted to have a high risk of unplanned CS who would otherwise have had a vaginal delivery. This parameter is represented by the specificity: high specificity means a low false positive rate. Indeed, false-negative women would undergo a labor trial as standard clinical care, but false-positive women could undergo unnecessary planned CS, which could be iatrogenic.

Based on actual scientific literature, class III obese women are at higher risk of CS [37]. In a large retrospective cohort of 64,272 infants born to obese pregnant women, published in 2012, the authors showed that, for nulliparous class III obese women, scheduled CS accounted for 21.9% of deliveries, and emergency CS for 24.6%. [30] Moreover, planned and unplanned CS rates increased with BMI ($p < 0.0001$). [30] Within the perspective of reducing the rate of unplanned CS during

labor without increasing the overall CS rate in a significant way, we opted for a specificity threshold of 80%, which is proportional to the rate of planned CS in this population. Based on this parameter, we attempted to maximize sensitivity, with the goal of screening as many women as possible who will have an unplanned CS during labor to reduce morbidity related to emergency CS.

It has been shown that CS is associated with more morbidity than vaginal delivery in obese women. [17, 18] Conversely, without the specification of maternal obesity, a recent meta-analysis showed that planned CS was associated with significantly lower maternal and neonatal morbidity as compared to unplanned CS, such as infection (relative risk [RR] = 0.44), postoperative fever (RR = 0.29), urinary tract infection (RR = 0.31), wound dehiscence (RR = 0.67), disseminated intravascular coagulation (RR = 0.34), reoperation (RR = 0.44), and infant mortality (RR = 0.16). [22] Another study found similar outcomes in terms of post-operative wound infection, post-partum hemorrhage and the necessity of blood transfusion, urinary tract infection, fever, and maternal intensive care unit admission. Neonatal outcomes such as birth asphyxia, meconium-stained liquid, and the need for neonatal intensive care unit admission were also significantly higher when unplanned CS was performed as compared to planned CS. [21] From these data, we can extrapolate that similar maternal complication trends could be observed in obese women, and even worse infant outcomes, such as the decision-incision time, could be increased because of more problematic transfer. Also, incision-birth time could also be lengthened because of surgical difficulties. [31]

4.3 Clinical Implications

Based on the algorithm's performance, the use of this model in routine practice could decrease the rate of unplanned CS at the cost of a modest increase in the overall CS rate. However, according to the scientific evidence, it seems that, in obese patients, the benefits of avoiding one unplanned CS out of two outweigh the cost of performing one unnecessary planned CS out of five women, considering both maternal and neonatal outcomes. Nonetheless, it should be taken into account when discussing obstetrical management in any individual situation. This model has two important advantages in clinical practice. First, it is simple to use because it requires only two parameters: initial weight and labor induction. Therefore, when discussing the route of delivery with a pregnant woman with class III obesity, an estimation of risk can be quickly obtained during the consultation. In addition, when there is an indication of the need to perform labor induction at a specific time, for example, because of overdue delivery, the risk can be estimated for each situation (in the case of spontaneous labor, before the planned induction date or in case of planned induction). This could allow better advice to be given to women at the end of their pregnancy by proposing personalized management according to the term and mode of onset of labor. For instance, a pregnant woman with a high risk of cesarean section with the induction of labor but a low risk of cesarean section for spontaneous labor could be advised to perform a trial of labor if she goes into spontaneous labor and, if she arrives at the expected induction date without going into spontaneous labor, opt for a scheduled cesarean section instead.

Recently, different predictive models have been published in obstetrics that allow delivery planning and enhance information quality during the consultation. These models can predict delivery outcomes based on machine learning algorithms, and they also employed random forest algorithms but with different clinical objectives. [2, 32, 33] The performance levels of this kind of model are remarkable, and some of them can already be used in clinical practice. Nonetheless, this innovative approach must also meet several quality criteria, with some of them being applicable to all clinical studies but others being specific to artificial-intelligence-based models. For these reasons, this retrospective, population-based, observational study was conducted according to a rigorous methodology, following the STROBE checklist but also the CONSORT-AI guidelines [34] published in September 2020, which specify the quality criteria to be applied to machine-learning models developed within the framework of randomized trials, in particular the performance of a test on a different cohort, allowing the internal validation to be completed via an external validation. To date, no specific guidelines are available on machine-learning models developed from retrospective cohort studies.

In order to meet important quality criteria, we included women from two different centers and found that both cohorts were different in some regards, such as age, initial weight, BMI, diabetes and high blood pressure prevalence, but also regarding clinical practices, such as unplanned CS rates. These differences can be explained by a center effect with different populations: Lille's cohort had higher BMI and diabetes, while Strasbourg's cohort had more women with high blood pressure and higher maternal age. Moreover, the center effect can also explain differences in

clinical practices because unplanned CS indications may differ not only because of a different population but also because of different internal protocols and obstetric practices. Finally, these differences can also be explained by a time effect due to the difference in years between the Strasbourg and Lille recruitment efforts, and the evolution of practices during the study periods (particularly on labor induction indications and techniques, still there was no evolution in CS rates in both cohorts, results not shown). Within the context of this study, heterogeneity between training and testing cohorts reinforces the value of the model in terms of external validation and makes it more applicable to other centers. In contrast, the fact that the two cohorts came from two different centers and time periods limited the model because the pre-labor characteristics of the women in the Lille cohort were not all available for the analysis. We therefore had to restrict the training variables of the algorithm because we had only a small number of common variables between the two cohorts. Some criteria of the Strasbourg patients that could be involved in the risk of cesarean section could therefore not be tested, thus reducing the power of the model (for example, uterine height, Bishop score at admission, and ultrasound estimation of fetal weight).

4.4 Research Implications

Further investigations are needed within a prospective clinical trial to investigate its performances in other cohorts. As this model was trained on two French cohorts, supplementary data is needed to assess its validity in other countries, given the fact that population and practices may differ. Moreover, other

pre-labor variables may be implicated in delivery prediction among this population and may improve the prediction quality. Consequently, it is important to continue these investigations to improve the model.

4.5 Limitations of the model

It must be noted that this model was developed based on pre-labor characteristics because we wanted to discuss the unplanned CS risk before labor and, eventually, suggest planned CS. It was therefore not possible to obtain a very high accuracy for this predictive model because several parameters will only have an effect at the time of labor and are not predictable: for example, an abnormality in terms of fetal heart rate or a mechanical dystocia. The use of pre-labor data makes model performance lower but more relevant in case of a choice between a trial of labor and planned CS.

Moreover, the model was established only for nulliparous patients with class III obesity ($\text{BMI} \geq 40 \text{ kg/m}^2$). Therefore, it cannot be used for patients with a BMI between 30 and 40 kg/m^2 or for patients with a history of cesarean section, although they are also among the patients most at risk of cesarean section during labor. It must also be noted that, unlike Anglo-Saxon populations, the number of women with extreme BMIs ($> 50 \text{ kg/m}^2$) is lower in France. Finally, the results of this model should be interpreted with caution for women with initial weights $> 130 \text{ kg}$ and no labor induction because the estimated regression curve depends only on a single observation.

4.6 Conclusions

Delivery planning in obese women is an everyday clinical challenge in obstetrics, and each choice has its specific advantages and risks. With its innovative approach, this model is effective in predicting unplanned CS risk in nulliparous class III obese women undergoing a trial of labor and could improve maternal and neonatal outcomes due to unplanned CS-related morbidity in this population. Still, this model is not reliable enough to become the sole element of obstetric decisions, though it could be an additional argument for explaining unplanned CS risk and participate in the choice of a trial of labor versus planned CS. Further investigations are needed within a prospective clinical trial, because this model could be used to meet the demand for patient information and risk calculation in a personalized fashion.

References

- [1] Bluher M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol* 2019;15(5):288-98.
- [2] Meyer R, Hendin N, Zamir M, Mor N, Levin G, Sivan E, et al. Implementation of machine learning models for the prediction of vaginal birth after cesarean delivery. *J Matern Fetal Neonatal Med* 2022;35(19):3677-83.
- [3] Catalano PM, Shankar K. Obesity and pregnancy: mechanisms of short term and long term adverse consequences for mother and child. *BMJ* 2017;356:j1.
- [4] Chen C, Xu X, Yan Y. Estimated global overweight and obesity burden in pregnant women based on panel data model. *PLoS One* 2018;13(8):e0202183.
- [5] Kuitunen I, Huttunen TT, Ponkilainen VT, Kekki M. Incidence of obese parturients and the outcomes of their pregnancies: A nationwide register study in Finland. *Eur J Obstet Gynecol Reprod Biol* 2022;274:62-7.
- [6] Garabedian C, Servan-Schreiber E, Riviere O, Vendittelli F, Deruelle P. [Maternal obesity and pregnancy: Evolution of prevalence and of place of birth]. *J Gynecol Obstet Biol Reprod (Paris)* 2016;45(4):353-9.
- [7] Fuchs F, Senat MV, Rey E, Balayla J, Chaillet N, Bouyer J, et al. Impact of maternal obesity on the incidence of pregnancy complications in France and Canada. *Sci Rep* 2017;7(1):10859.
- [8] Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health* 2009;9:88.
- [9] Marchi J, Berg M, Dencker A, Olander EK, Begley C. Risks associated with obesity in pregnancy, for the mother and baby: a systematic review of reviews. *Obes Rev* 2015;16(8):621-38.
- [10] Shin D, Song WO. Prepregnancy body mass index is an independent risk factor for gestational hypertension, gestational diabetes, preterm labor, and small- and large-for-gestational-age infants. *J Matern Fetal Neonatal Med* 2015;28(14):1679-86.
- [11] Weiss JL, Malone FD, Emig D, Ball RH, Nyberg DA, Comstock CH, et al. Obesity, obstetric complications and cesarean delivery rate--a population-based screening study. *Am J Obstet Gynecol* 2004;190(4):1091-7.
- [12] Chu SY, Kim SY, Schmid CH, Dietz PM, Callaghan WM, Lau J, et al. Maternal obesity and risk of cesarean delivery: a meta-analysis. *Obes Rev* 2007;8(5):385-94.
- [13] Heslehurst N, Simpson H, Ells LJ, Rankin J, Wilkinson J, Lang R, et al. The impact of maternal BMI status on pregnancy outcomes with immediate short-term obstetric resource implications: a meta-analysis. *Obes Rev* 2008;9(6):635-83.
- [14] Kominiarek MA, Vanveldhuisen P, Hibbard J, Landy H, Haberman S, Learman L, et al. The maternal body mass index: a strong association with delivery route. *Am J Obstet Gynecol* 2010;203(3):264 e1-7.

- [15] Sheiner E, Levy A, Menes TS, Silverberg D, Katz M, Mazor M. Maternal obesity as an independent risk factor for caesarean delivery. *Paediatr Perinat Epidemiol* 2004;18(3):196-201.
- [16] Borghesi Y, Labreuche J, Duhamel A, Pigeyre M, Deruelle P. Risk of cesarean delivery among pregnant women with class III obesity. *Int J Gynaecol Obstet* 2017;136(2):168-74.
- [17] Robinson HE, O'Connell CM, Joseph KS, McLeod NL. Maternal outcomes in pregnancies complicated by obesity. *Obstet Gynecol* 2005;106(6):1357-64.
- [18] Sebire NJ, Jolly M, Harris JP, Wadsworth J, Joffe M, Beard RW, et al. Maternal obesity and pregnancy outcome: a study of 287,213 pregnancies in London. *Int J Obes Relat Metab Disord* 2001;25(8):1175-82.
- [19] World Health Organization (WHO). Obesity and overweight. 2021.
- [20] Gunatilake RP, Smrtka MP, Harris B, Kraus DM, Small MJ, Grotegut CA, et al. Predictors of failed trial of labor among women with an extremely obese body mass index. *Am J Obstet Gynecol* 2013;209(6):562 e1-5.
- [21] Darnal N, Dangal G. Maternal and Fetal Outcome in Emergency versus Elective Caesarean Section. *J Nepal Health Res Counc* 2020;18(2):186-9.
- [22] Yang XJ, Sun SS. Comparison of maternal and fetal complications in elective and emergency cesarean section: a systematic review and meta-analysis. *Arch Gynecol Obstet* 2017;296(3):503-12.
- [23] von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370(9596):1453-7.
- [24] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* 2022;28(5):924-33.
- [25] Malley JD, Kruppa J, Dasgupta A, Malley KG, Ziegler A. Probability machines: consistent probability estimation using nonparametric learning machines. *Methods Inf Med* 2012;51(1):74-81.
- [26] Genuer R, Poggi J-M, Tuleau-Malot C. VSURF: An R package for variable selection using random forests. *The R Journal* 2015;7.
- [27] Brien C, Bel S, Boudier E, Deruelle P. [Caesarean risk factors during labor for a class III obese nulliparous]. *Gynecol Obstet Fertil Senol* 2021;49(6):517-21.
- [28] Gunatilake RP, Perlow JH. Obesity and pregnancy: clinical management of the obese gravida. *Am J Obstet Gynecol* 2011;204(2):106-19.
- [29] Poobalan AS, Aucott LS, Gurung T, Smith WC, Bhattacharya S. Obesity as an independent risk factor for elective and emergency caesarean delivery in nulliparous women--systematic review and meta-analysis of cohort studies. *Obes Rev* 2009;10(1):28-35.
- [30] Marshall NE, Guild C, Cheng YW, Caughey AB, Halloran DR. Maternal superobesity and perinatal outcomes. *Am J Obstet Gynecol* 2012;206(5):417 e1-6.

- [31] Abenhaim HA, Benjamin A. Higher caesarean section rates in women with higher body mass index: are we managing labour differently? *J Obstet Gynaecol Can* 2011;33(5):443-8.
- [32] Lindblad Wollmann C, Hart KD, Liu C, Caughey AB, Stephansson O, Snowden JM. Predicting vaginal birth after previous cesarean: Using machine-learning models and a population-based cohort in Sweden. *Acta Obstet Gynecol Scand* 2021;100(3):513-20.
- [33] Lipschuetz M, Guedalia J, Rottenstreich A, Novoselsky Persky M, Cohen SM, Kabiri D, et al. Prediction of vaginal birth after cesarean deliveries using machine learning. *Am J Obstet Gynecol* 2020;222(6):613 e1- e12.
- [34] Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Spirit AI, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020;26(9):1364-74.
- [35] Hosmer D, Lemeshow S, Sturdivant R. *Applied Logistic Regression*, Third Edition. 2013.
- [36] Breiman L. Random Forests. *Machine Learning* 2001;45(1):5-32.
- [37] Su X, Peña AT, Liu L, Levine RA. Random forests of interaction trees for estimating individualized treatment effects in randomized trials. *Stat Med* 2018;37(17):2547-60.
- [38] Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 2006;7:3.
- [39] Howard J, Bowles M. The two most important algorithms in predictive modeling today. *Strata Conference presentation, February. Vol. 28.* 2012.
- [40] Biau G, Scornet E. A random forest guided tour. *TEST* 2016;25(2):197-227.
- [41] Collett D. *Modelling Binary Data*. CRC Press; 2002.
- [42] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837-45.

Tables

Table 1: Total cohort description according to the delivery mode

Variable	Cesarean section (N=164)	Vaginal delivery (N=246)	Total (N=410)	P- Value
Age(years)				
Mean (SD)	28.8 (5.1)	27.8 (5.1)	28.2 (5.1)	0.063
Range	18 - 42	18 - 43	18 - 43	
Height (cm)				
Mean (SD)	163.6 (5.9)	165.7 (7.0)	164.7 (6.0)	0.001
Range	148 - 186	148 - 183	148 - 186	
Initial Weight (kg)				
Mean (SD)	118.0 (13.9)	119.1 (12.9)	118.7 (13.3)	0.432
Range	89 - 160	90 - 162	89 - 162	
Body Mass Index (kg/m²)				
Mean (SD)	44.1 (4.2)	43.3 (3.4)	43.0 (3.8)	0.0501
Range	40.0 - 59.6	40.0 - 56.3	40.0 - 59.6	
Gestational Weight Gain (kg)				
Mean (SD)	6.4 (8.3)	5.6 (9.0)	5.9 (8.7)	0.389
Range	-19 - 30	-36 - 29	-36 - 30	
Unknown	5	11	16	
Smoking				
No	132 (80.5%)	198 (80.8%)	330 (80.7%)	0.934
Yes	32 (19.5%)	47 (19.2%)	79 (19.3%)	
Unknown	0	1	1	
Diabetes mel litus				
No	131 (79.9%)	196 (79.7%)	327 (79.8%)	0.960
Yes	33 (20.1%)	50 (20.3%)	83 (20.2%)	
High Blood Pressure				
No	137 (83.5%)	203 (82.5%)	340 (82.9%)	0.789
Yes	27 (16.5%)	43 (17.5%)	70 (17.1%)	
Labor induction				
No	48 (29.3%)	126 (51.2%)	174 (42.4%)	<0.001
Yes	116 (70.7%)	120 (48.8%)	236 (57.6%)	
Gestational age at delivery (weeks, days)				
Mean (SD)	39w 6d (1w 5d)	39w 4d (2w 4d)	39w 5d (2w 2d)	0.257
Range	31w 3d - 42w 1d	22w 0d - 42w 2d	22w 0d - 42w 2d	

Comparison of maternal and labor characteristics according to the delivery mode in the total cohort. Legend: SD = standard deviation

Table 2: Total cohort description according to the center

Variable	Lille (N=163)	Strasbourg (N=247)	Total (N=410)	P-value
Age (years)				
Mean (SD)	27.3 (5.2)	28.8 (5.1)	28.6 (5.1)	0.004
Range	18 - 40	19 - 43	18 - 43	
Height (cm)				
Mean (SD)	165.1 (6.6)	164.8 (6.7)	164.9 (6.6)	0.659
Range	148 - 183	148 - 186	148 - 186	
InitialWeight(kg)				
Mean (SD)	121.0 (13.9)	117.1 (12.7)	118.7 (13.3)	0.003
Range	90 - 162	89 - 156	89 - 162	
BodyMassIndex(kg/m2)				
Mean(SD)	44.4(4.4)	43.1(3.2)	43.0(3.0)	<0.001
Range	40.0-59.6	40.0-54.0	40.0-59.6	
Gestatio nalWeightGain(kg)				
Mean (SD)	5.8 (8.7)	6.0 (8.7)	5.9 (8.7)	
Range	-18 - 27	-36 - 30	-36 - 30	0.790
Unknown	3	13	16	
Smoking				
No	125 (76.7%)	205 (83.3%)	330 (80.7%)	
Yes	38 (23.3%)	41 (16.7%)	79 (19.3%)	0.096
Unknown	0	1	1	
Diabetes				
No	108 (66.3%)	219 (88.7%)	327 (79.8%)	
Yes	55 (33.7%)	28 (11.3%)	83 (20.2%)	< 0.001
Hi; h Blood Pressure				
No	143 (87.7%)	197 (79.8%)	340 (82.9%)	
Yes	20 (12.3%)	50 (20.2%)	70 (17.1%)	0.036
Delivery				
Cesarean	80 (49.1%)	84 (34.0%)	104 (40.0%)	0.002
Vaginal	83 (50.9%)	163 (66.0%)	246 (60.0%)	
Labor induction				
No	75 (46.0%)	99 (40.1%)	174 (42.4%)	0.234
Yes	88 (54.0%)	148 (59.9%)	236 (57.6%)	
Gestational age at delivery (weeks, c ays)				
Mean (SD)	39w 6d (2w 0d)	39w 4d (2w 3d)	39w 5d (2w 2d)	0.277
Range	22w0d-42w1d	22w5d-42w2d	22w0d-42w2d	

Comparison of maternal, labor and delivery characteristics according to the center.

Legend: SD = standard deviation

Table 3: Final logistic model built on Strasbourg cohort

Variable	Estimate (SE)	Odds ratio (95%CI)	P-value
Initial weight	-0.03 (0.01)	0.97(0.95;1.00)	0.031
Labor induction	1.12 (0.32)	3.06(1.66;5.84)	<0.001
Age	0.0484 (0.03)	1.05(0.99;1.11)	0.095
Diabetes	0.711 (0.48)	0.49(0.18;1.22)	0.142

Feature characteristics and weights in the logistic model. Legend: SE = standard error; 95%CI = 95% confidence interval

Table 4: Model performances evaluated on the test cohort.

Model	Prob forest	Logistic model
<i>Cut-point</i>	0.495	0.452
<i>AUC</i>	0.70 (0.62, 0.78)	0.66 (0.58, 0.75)
<i>Specificity</i>	0.87 (0.77, 0.93)	0.90 (0.82, 0.96)
<i>Sensitivity</i>	0.44 (0.32, 0.55)	0.18 (0.10, 0.28)
<i>Accuracy</i>	0.66 (0.58, 0.73)	0.55 (0.47, 0.63)

Comparison of the performances of the two developed models (probability forest and logistic regression). Legend: AUC = area under the curve

Figures titles and legends

Figure 1: Predicted CS risk in the Strasbourg cohort for the probability forest model according to the labor induction status and the initial weight.

Note: the lines represents the estimated local polynomial regression curves with their 95% confidence interval.

Figure 2: Receiver Operating Characteristic curve of the models on the test set

Legend: the ROC curve of the probability forest model (solid line) and of the logistic model (dotted line) drawn on the test set. The black cross on each curve indicates the selected cut-point and its coordinates are the values of 1-specificity and the sensitivity measured in the test set.



