		Model development			Model performance		
Model		**Type of model and presentation**	**Predictive factors used for final model**	**Validation techniques**	**Model calibration**	**Model discrimination (AUC of RUC aka c-statistic)**	**Clinical usefulness measures (e.g. sensitivity, specificity, shown ROC curve)**
Newman-Peacock model (Newman 1993, Tasnim 2012, Silva 2018)		**Univariate analysis:** t-test, U Mann-Whitney test, ANOVA (continuous) and χ2 test, Fisher tests (categorical) **Multivariate analysis**: Stepwise linear regression analysis **Presentation**: Score chart (Newman-Peacock score)	parity, EFW, cervical dilation, placental location and fetal station	**External validation #1 (Original article)** Temporal validation (prospective group) **External validation #2 (Tasnim 2012)** New nation and investigational team **External validation #3 (Silva 2018)** New nation and investigational team	**External validation #1**: Other: Graph of observed chance by score **External validation #2**: Table: observed chance by score **External validation #3**: Table: observed chance by score	**External validation #3:** 0.642 [95% CI 0.583-0.701]	**External validation #3:** ROC curve shown
Lin 2022		**Univariate analysis:** t-test (continuous), Mantel–haenszel test (categorical). **Multivariate analysis:** Logistic regression analysis with stepwise input covariates **Presentation**: Nomogram (cutoffs derived to maximize Youden index)	parity, BMI increase, AFI, anesthesia, placenta location	**External validation #1:** Temporal validation (second prospective group of n=35; occured in same facility, but did not restrict who performed the ECV while all ECVs in training data were performed by the same person)	**Training data:** Hosmer-Lemeshow: χ2=3.05, p=0.931 (good) Calibration plot: shown (good) **External validation #1:** Calibration plot: shown (good)	**Training data:** 0.835 [95% CI 0.771–0.899] f* **External validation #1:** 0.889 [95% CI 0.776–1.000]	**Training data**: ROC curve shown, DCA-curve shown **External validation #1:** ROC curve shown, DCA-curve shown
Wong 2000		**Univariate analysis:** Logistic regression **Multivariate analysis**: None **Presentation**: Score chart	palpable head, breech engagement, fundal height, uterine relaxation	**External validation #1:** Temporal validation (second prospective group)	**Training + validation**: Table: likelihood ratio and observed chance by score	-	**Training + validation:** Likelihood ratio by score: *Score 0-1*: 0; *Score 2*: 1.8; *Score 3*: 30; *Score 4*: 57.7
Burgos post-ECV model (Burgos 2011, Burgos 2012)		**Univariate analysis:** Logistic regression **Multivariate analysis**: Logistic regression (maximum modelling strategy) **Presentation**: Score chart	parity, placenta location, amnionic fluid index, type of breech	**External validation #1:** Temporal validation (second prospective group of n=500 in same center)	**Training + validation**: Table: predicted vs observed chance by score group	**Training data**: 0.738 [95% CI 0.695-0.782] c* **External validation #1**: 0.669 [95% CI 0.622-0.716]	**Training + validation:** Predicted success [95% CI] by score: *Score 4-6*: 30.9% [26.1-36.1] *Score 7-8*: 56.5% [51.8-61.2] *Score 9-14*: 76.8% [70.4-82.5]
Kok model (Kok 2011, DeHundt 2012)		**Univariate analysis:** Logistic regression, others unstated **Multivariate analysis:** Logistic regression with stepwise backward elimination **Presentation**: Equation; Score chart	parity, EFW, placenta location, AFI	**Internal validation:** Bootstrapping (200 times) **External validation #1:** Temporal validation (retrospective data from years before and after)	**Internal validation (implied):** Hosmer-Lemeshow: p=0.66 (good) Calibration plot: shown h* **External validation #1:** Hosmer-Lemeshow: p=0.3 (good) Calibration plot: shown (good) Table: predicted vs observed chance by score	**Internal validation (implied)**: 0.738 [95% CI 0.695–0.782] h* **External validation #1**: 0.669 [95% CI 0.622–0.716]	**Internal validation (implied):** ROC curve shown h* **External validation:** ROC curve shown **All**: Figure of predicted odds (with 95% CI) by score value
Lau 1997	**With uterine tone**	**Univariate analysis:** t-test (continuous), χ2 test (categorical) **Multivariate analysis**: Logistic regression with stepwise forward elimination **Presentation**: With uterine tone: None (except regression coefficients and constants in each data split); With parity: variable combination table;	engagement, how easy it was to feel the fetal head, whether uterus was tense on palpation	**Internal validation:** Random data split (two subgroups to derive models and tested against each other, n=129 and n=114) a*	-	-	**Internal validation:** Accuracy: 79.4% (193/243)** Sensitivity: 89.3% (151/169)** Specificity: 56.8% (42/74)**
	With parity		engagement, how easy it was to feel the fetal head, parity	Model derived without any validation	**Training data:** Other: Data by variable combination	-	-
Velzel 2018		**Univariate analysis:** Logistic regression **Multivariate analysis:** Logistic regression with stepwise backward elimination and Akaike's information criterion **Presentation:** Equation	parity, ethnicity, administration of fenoterol, AFI, fetal position, placental position, engagement of the breech in the pelvis, uterine tone, palpation of the head, GA	**Internal validation:** Bootstrapping (100 times, n=818)	**Internal validation:** Calibration plot: shown with histogram (good)	**Internal validation:** 0.78 [95% CI 0.75-0.81]	-
Isakov 2019 (and erratum)		**Univariate analysis:** Logistic regression and χ2 test (categorical), t-test (continuous, normal), Kruskal-Wallis tests (continuous, non-normal) **Multivariate analysis:** Logistic regression (both models). Cross-validation (10-fold) within training set. **Presentation**: Decision tree	parity, BMI, and size of fore-bag	**Internal validation:** Random data split (75%-25%)	-	**Training data:** 0.928 [95% CI 0.924–0.931] **Internal validation:** 0.933 [95% CI 0.863–1]	**Internal validation**: ROC curve shown
Dahl 2021		**Univariate analysis:** Logistic regression **Multivariate analysis:** Logistic regression with backward elimination. Bootstrapping (1000x, n=2000) within training set and model with highest AUC used. **Presentation**: Equation and website g*	parity, BMI, placenta location, and type of malpresentation	**Internal validation:** Random data split (80%-20%);	**Internal validation (implied):** Calibration plot: shown with histogram (good) h*	**Internal validation (implied):** 0.667 [95% CI 0.634–0.701] h*	**Internal validation (implied)**: ROC curve shown h*
Dong 2021	**Post-ECV model**	*For each of the two outcomes:* **Univariate analysis:** Logistic regression **Multivariable analysis:** Logistic regression with stepwise backward elimination in each imputation set. Variables were included in the final model if it was selected in ≥5 of the sets. The final model was fitted within each imputation set and regression values were pooled. **Presentation**: Equations	parity, GA, BMI before pregnancy, palpation of the fetal head, breech engagement, AFI, BPD and posterior placenta	**Internal validation:** Bootstrapping (500 times for each of the 10 imputation sets)	**Training data:** Hosmer-Lemeshow: χ2=3.711, p=0.882 (good) Calibration plot: shown with histogram (good)	**Training data:** 0.823 [95% CI 0.764–0.881] **Imputed training data:** 0.815 **Internal validation:** 0.80	**Training data:** ROC curve shown At max Youden (0.512), Sensitivity: 66.7%, Specificity: 84.5%, PPV: 89.5%, NPV: 56.1%
	Delivery model		parity, maternal age, BMI before pregnancy, placenta location, position of the fetal spine	**Internal validation:** Bootstrapping (500 times for each of the 10 imputation sets)	**Training data:** Hosmer-Lemeshow: χ2=5.085, p=0.748 (good) Calibration plot: shown with histogram (unsatisfactory)	**Training data:** 0.786 [95% CI 0.724–0.848] **Imputed training data**: 0.777 **Internal validation:** 0.749	**Training data:** ROC curve shown At max Youden (0.466), Sensitivity: 65.5%, Specificity: 81.1%, PPV: 91.7%, NPV: 42.6%
Anand 2019		**Univariate analysis:** Logistic regression **Multivariable analysis:** Logistic regression with stepwise backward elimination **Presentation**: Equation; approximating score chart	parity, AFI, type of breech, placental location, uterine tone, position of breech, station	**Internal validation:** Bootstrapping (500 times, overfitting corrected to derive final regression equation)	**Training data:** Hosmer-Lemeshow: χ2=5.96, p=0.65 (good) Calibration plot: shown (good), p=0.965	**Training data:** 0.782 [95% CI 0.741–0.822] **Internal validation:** 0.765	**Training data**: ROC curve shown
Aisenbrey 1999		**Univariate analysis:** Fisher exact test (all variables are categorical) **Multivariate analysis**: Logistic regression with backward elimination **Presentation**: None (except regression coefficients and constant)	uterine tone, placental location, and breech location	Model derived without any validation	**Training data:** Hosmer-Lemeshow: p=0.924 (good)	-	**Training data:** When 50% probability cutoff: Accuracy: 93% (119/128) Sensitivity: 96.1% (75/78)** Specificity: 88.0% (44/50)** PPV: 92.6% (75/81)**, NPV: 93.6% (44/47)**
GNK-PIMS score (Tasnim 2012)		**Univariate analysis:** t-test, χ2 test **Multivariate analysis**: unclear/none **Presentation**: Score chart (GNK-PIMS score)	parity, EFW, cervical dilation, placental location and fetal station	Model derived without any validation b*	**Training data:** Table: observed chance by score group	-	-
López-Pérez 2020		**Univariate analysis:** t -est (continuous), χ2 test (categorical) **Multivariate analysis:** Logistic regression with stepwise forward elimination (Wald) **Presentation:** None given (except regression coefficients and constant) d*	parity, placentation and amniotic fluid data	Model derived without any validation	**Training data**: Hosmer-Lemeshow: χ2(7)=10.06, p=0.185 (good)	-	**Training data:** Accuracy: 74.0% (168/227) e* Sensitivity: 98.8% (160/162)** Specificity: 12.3% (8/65)**
Burgos 2015		**Univariate analysis:** Logistic regression **Multivariate analysis:** Logistic regression followed by manual, systematic backward elimination **Presentation:** None given (except odds ratio without constant)	parity, BMI, previous cesarean delivery, induction of labor, and time elapsed between ECV and delivery	Model derived without any validation	**Training data:** Hosmer-Lemeshow: reported as "good" without specific metrics	**Training data:** 0.728 [No CI given]	-
Zheng 2021		**Univariate analysis:** Factor weight by XGBoost algorithm. χ2 test (categorical), t-test (continuous, normal), Mann-whitney U test (continuous, non-normal) **Multivariate analysis:** Logistic regression (elimination method not reported) **Presentation:** Score chart	Sum of fundal height and station, Station of the fetal buttocks, Fetal head location (O'clock), Could the fetal breech be grasped easily?, Could the fetal head be grasped easily?	Model derived without any validation	**Training data:** Table: observed chance by score	**Training data:** 0.885 [95% CI 0.837-0.933]	**Training data:** ROC curve shown
Bilgory 2021 (and erratum)		**Univariate analysis:** Logistic regression, Fisher exact or χ2 test (categorical), t-test (continuous, normal), Mann-Whitney U tests (continuous, non-normal) **Multivariate analysis:** Logistic regression with backward elimination (Wald test) **Presentation**: Equation	parity, AFI	Model derived without any validation	**Training data:** Hosmer-Lemeshow: p=0.836 (good)	**Training data:** 0.71 [95% CI 0.672–0.748]	**Training data:** ROC curve shown
Hutton 2017's Combined parity model		**Univariate analysis:** Wilcoxon-rank sum test (continuous), χ2 test (categorical), Fisher exact test (categorical, any cells with values <7), logistic regression **Multivariate analysis:** Logistic regresion (using Bayesian information criteria) and CART analysis (analyzed multiparous and nulliparous separately) **Presentation**: Decision trees	**In multiparous women:** engagement, GA, placenta location, palpable head **In nulliparous women:** engagment, palpable head, placenta location, BMI, breech presentation	Model derived without any validation	-	-	-
Ebner 2015		**Univariate analysis**: χ2 test (categorical), t-test (continuous, normal), Mann-Whitney U tests (continuous, non-normal per significant Shapiro-Wilk test). All 2-sided. **Multivariate analysis**: Logistic regression with stepwise backward elimination. **Presentation**: None given	parity, AFI, GA	Model derived without any validation	-	-	-
Cobec 2020		**Univariate analysis**: χ2 test or Fisher's exact test (categorical), t-test (continuous), Welch's test (GA) **Multivariate analysis**: Logistic regression (elimination method not reported) **Presentation**: None given (except odds ratio)	parity, maternal age, GA, fetal weight, and AFI	Model derived without any validation	-	-	-
Palepu 2021		**Univariate analysis:** Logistic regression **Multivariable analysis:** Logistic regression (elimination method not reported) **Presentation**: None given (except odds ratio)	maternal age, previous C-section	Model derived without any validation	-	-	-
Svensson 2021	**Model 1**	**Univariate analysis:** Fisher's exact test, χ2 test (categorical) **Multivariate analysis:** Logistic regression (both models) **Presentation:** None given	parity, maternal age, GA	Model derived without any validation	-	-	-
	Model 2		parity, maternal age, GA, placenta position, position of fetal head	Model derived without any validation	-	-	-

In situations where it was not completely clear on what data section (training, internal validation, external validation) the metrics are reported for, the text "implied" was added in paranthesis.

** Indicates metrics that were not reported directly, but were calculated using the available information.

a* For Lau et al.'s uterine tone model, close examination of the article reveals methods that are functionally equivalent to k-fold cross-validation with k=2. However, Velzel 2015's review referred to it as "split-sample" validation.

b* For the GNK-PIMS score, it is unclear if Tasnim 2012 is the initial study describing model development or if it is an external validation study or a model development study. If the latter, the model development study was not referenced or could be found.

c* In Burgos 2011, an ROC curve is shown for model that only has 3 of the 4 final variables. Burgos 2012 has the #s for the final model, but no graphs.

d* López-Pérez 2020 has a screenshot of a calculator, but there is no way for readers to access a working calculator.

e* For López-Pérez 2020's accuracy and related metrics, the total is only 227 even though the n is 317 for the study. The article does not clarify why.

f* Lin 2022 gave both a C-index (0.841) and AUC value (training: 0.835 [95% CI 0.771–0.899], internal validation: 0.889 [95%CI 0.776–1.000]). It is not clear why these are different numbers as C-index is usually considered a synonym for AUC for binary discrimination problems.

g* Dahl 2021 provides a web link for their calculator (https://www.ecvcalculator.com/). However, the web calculator does not currently work as of 2023. It returns the error "There was something wrong, please contact the admin." after submitting the input.

h* Dahl 2021 implies that the reported AUC numbers are for internal validation and not for the training cohort, but this is not explicitly stated. A similar situation happened in Kok 2011.