

# The Optimal Prediction Model for Successful External Cephalic Version

**Running head:** Comparison of Models for Successful Version

Rahul Sai Yerrabelli<sup>1,2\*</sup>, Peggy K. Palsgaard<sup>1,3</sup>, Priya Shankarappa, MD<sup>1,4</sup>, Valerie Jennings, MD, MS<sup>1,5</sup>

- <sup>1</sup> Carle Illinois College of Medicine, The University of Illinois at Urbana-Champaign, Champaign, Illinois, United States of America
- <sup>2</sup> Department of Obstetrics & Gynecology, Reading Hospital, Reading, Pennsylvania, United States of America
- <sup>3</sup> Department of Obstetrics & Gynecology, Western Michigan University, Kalamazoo, Michigan, United States of America
- <sup>4</sup> Department of Internal Medicine, Brown University, Providence, Rhode Island, United States of America
- <sup>5</sup> Department of Obstetrics & Gynecology, Carle Foundation Hospital, Urbana, Illinois, United States of America

**\* Corresponding Author:**

RSY Email: [rahul.yerrabelli@towerhealth.org](mailto:rahul.yerrabelli@towerhealth.org) ORCID: [0000-0002-7670-9601](https://orcid.org/0000-0002-7670-9601)

**Other author contacts:**

PKP Email: [peggy.palsgaard@wmed.edu](mailto:peggy.palsgaard@wmed.edu) ORCID: [0000-0002-3797-498X](https://orcid.org/0000-0002-3797-498X)  
PS Email: [pshankarappa@lifespan.org](mailto:pshankarappa@lifespan.org)  
VJ Email: [valerie.jennings@carle.com](mailto:valerie.jennings@carle.com)

## Abstract

1 **Objective:** The majority of breech fetuses are delivered by Cesarean birth as few physicians are  
2 trained in vaginal breech birth. An external cephalic version (ECV) can prevent Cesarean  
3 delivery and the associated morbidity in these patients. Current guidelines recommend all  
4 patients with breech presentation be offered an ECV attempt. Not all attempts are successful, and  
5 an attempt does carry some risks so shared decision-making is necessary. To aid in patient  
6 counseling, over a dozen prediction models to predict ECV success have been proposed in the  
7 last few years. However, very few models have been externally validated, and thus none have  
8 been adopted into clinical practice. This study aims to use data from a United States hospital to  
9 provide further data on ECV prediction models.

10 **Study Design:** This study retrospectively gathered data from Carle Foundation Hospital and  
11 used it to test six models previously proposed to predict ECV success. These models were Dahl  
12 2021, Bilgory 2023, López Pérez 2020, Kok 2011, Burgos 2010, and Tasnim 2012 (GNK-PIMS  
13 score).

14 **Results:** 125 patients undergoing 132 ECV attempts were included. 69 attempts were successful  
15 (52.2%). Dahl 2021 had the greatest predictive value (AUC 0.779), while Tasnim 2012  
16 performed the worst (AUC 0.626). The remaining models had similar predictive values as each  
17 other (AUC 0.68-0.71). Bootstrapping confirmed that all models except Tasnim 2012 had  
18 confidence intervals not including 0.5. The bootstrapped 95% AUC confidence interval for Dahl  
19 2021 was 0.71-0.84. In terms of calibration, Dahl 2021 was well calibrated with predicted  
20 probabilities matching observed probabilities. Bilgory 2023 and López Pérez were poorly  
21 calibrated.

22 **Conclusion:** Multiple prediction tools have now been externally validated for ECV success.

23 Dahl 2021 is the most promising prediction tool.

24

25 **Keywords:** breech presentation; cesarean section; clinical prediction rule; logistic regression;

26 machine learning; nomogram; obstetrics; prognostic score

27

28 **Key Points**

- 29 • Prediction models can be powerful tools for patient counseling
- 30 • The odds of ECV success can estimated based on patient factors and clinical findings
- 31 • Of the 6 tested models, only Dahl 2021 appears to have good predictive value and
- 32 calibration

33

## 34 **Introduction**

35           Breech presentation complicates approximately 3-4% of term pregnancies, and transverse  
36 lie complicates an additional 0.1%-0.2%<sup>1-3</sup>. The three main options for malpresentation  
37 presentation include vaginal breech birth (VBB), Cesarean delivery, or external cephalic version  
38 (ECV) followed by cephalic vaginal delivery<sup>1</sup>. Vaginal breech births are extremely rare in the  
39 United States and only offered in a limited number of centers. External cephalic version is a  
40 physical maneuver where physicians press on the mother's abdomen in the third trimester to try  
41 to rotate to a cephalic ("head down") position. If successful, the ECV allows for a cephalic  
42 vaginal delivery and foregoes the need for a Cesarean delivery or VBB. As such, ECVs are cost  
43 effective and noninvasive procedures that decrease Cesarean rates and avoid the associated  
44 morbidity for patients<sup>2</sup>. Current American and British guidelines recommend that patients who  
45 are near term with breech presentation be offered an ECV if there are no contraindications<sup>2,4,5</sup>.

46           Wide variation in ECV success rates have been reported, although most estimates are  
47 usually around 50-60%<sup>2,5-7</sup>. Furthermore, not every patient has the same likelihood of procedure  
48 success. Some risk factors for ECV failure include low amniotic fluid volume, nulliparity,  
49 obesity, descent of the breech into the pelvis, and later gestational age<sup>8-11</sup>. Regardless of whether  
50 the procedure is successful, the procedure can rarely result in complications, the most severe  
51 include placental abruption (0.18%), umbilical cord prolapse (0.18%), fetomaternal hemorrhage  
52 (0.9%), rupture of membranes (0.22%), and fetal death (0.19%)<sup>7</sup>. As such, it is recommended  
53 that an ECV only be performed in an institution where an emergency Cesarean delivery can be  
54 performed and that shared decision making be employed to determine whether to proceed with  
55 an ECV vs a scheduled Cesarean birth.. Given the overall risks of the procedure, there has been a

56 significant move in the field to publish models which predict an individual patient's success, to  
57 aide patients and practitioners in deciding for or against this intervention.

58 Many statistical models have been published in the last decade which predict the success  
59 of an ECV depending on a multitude of variables such as: parity, placental location, amniotic  
60 fluid volume, fetal sex, and descent of the presenting part<sup>12</sup>. We recently published a systematic  
61 review of the 25 models that have been published so far, 14 of which in the last 5 years. The vast  
62 majority of models had no external validation and thus were not being used in clinical practice.  
63 Because of various statistical biases, it is common for a prediction model to be overfit and appear  
64 to perform well in the original study, but be poorly predictive in reality. Thus, it is critical that  
65 rigorous external validation be performed before these models are used clinically as acting on  
66 incorrect predictions is worse than have no prediction at all. Therefore, in our systematic review,  
67 we determined that rather than new models to be developed, there is a need for validation of the  
68 existing models to determine their suitability for clinical use<sup>13</sup>. This study aims to fill that gap in  
69 the current literature and uses data from a hospital in the United States to externally validate six  
70 different existing prediction models for ECV success.

71

## 72 ***Methods***

### 73 ***Study design***

74 We performed a retrospective cohort study at Carle Foundation Hospital in Urbana,  
75 Illinois, USA. Approval from the institutional review board (IRB) was obtained prior to data  
76 collection and the study was deemed exempt. Inclusion criteria were pregnant individuals in  
77 whom one or more ECVs were attempted or performed. ECVs in patients with a prior Cesarean

78 section (TOLAC) and in non-breech (e.g., transverse and oblique lie) presentations were  
79 included. Twin pregnancies were excluded. Patients were excluded if the ECV or the delivery  
80 did not happen at our institution or if the patient did not have any formal ultrasounds available in  
81 our EMR.

82 Data was retrieved from maternal medical records including demographic characteristics,  
83 pregnancy information, ultrasound information, the success of the ECV attempt, and delivery  
84 information. All charts were manually reviewed by the authors to confirm ECV success and final  
85 method of delivery. ECV success was defined as cephalic presentation when the physician left  
86 the room after the procedure. If the fetus returned to non-cephalic presentation after, then a  
87 second attempt was considered a separate ECV altogether. Fetal presentation was obtained from  
88 the patient's most recent formal ultrasound before the ECV; if the patient's last ultrasound  
89 indicated cephalic presentation, then presentation information was obtained from physician notes  
90 at the time of the ECV. REDCap was used to store the data until it was deidentified.

91 Demographic data was collected and tabulated. Pre-gravid BMI (BMI prior to pregnancy  
92 or at first prenatal visit) was able to be collected for almost all patients, but BMI before ECV and  
93 BMI at delivery was only available for a subset of patients. Many of the models included BMI,  
94 but only some specified which BMI was used. The appropriate BMI was used when possible;  
95 otherwise, the last available BMI was used. Some patients had other missing data (most  
96 commonly for fetal presentation); thus, these patients were only used for evaluating prediction  
97 models that did not require that variable.

98 All models except Dahl 2021<sup>14</sup> used amniotic fluid volume; four models used amniotic  
99 fluid index (AFI) while Burgos 2010<sup>15,16</sup> used qualitative terms (e.g. "scarce", "normal", or  
100 "abundant"). AFI was collected from patient records. For calculation of the Burgos 2010 model,

101 qualitative criteria were converted to oligohydramnios (AFI<5), normal, and polyhydramnios  
102 (AFI≥24)<sup>17,18</sup>. Five patients did not have an AFI reported, but did have a single deepest pocket  
103 (SDP); this was approximated to an AFI value by multiplying by 3.

104 Some models required estimated fetal weight (EFW). The last growth ultrasound was  
105 used to obtain the EFW percentile. Then, using the World Health Organization Fetal Growth  
106 Charts, the EFW at the time of ECV was calculated using the EFW percentile, gestational age at  
107 ECV, and the fetal sex<sup>19,20</sup> (see <https://github.com/jcarvalho45/whoFetalGrowth> for code and  
108 dataset).

### 109 *Statistical analysis*

110 Six prediction models were chosen from the prior systematic review: Dahl 2021<sup>14</sup>,  
111 Bilgory 2023<sup>21,22</sup>, López Pérez 2020<sup>12</sup>, Kok 2011<sup>23</sup>, Burgos 2010<sup>15,16</sup>, and Tasnim 2012 (GNK-  
112 PIMS score)<sup>24</sup>. These models were chosen because they did not require any additional  
113 measurements (e.g., forebag size, cervical exam before ECV) beyond what is routinely already  
114 performed at our institution. The score or predicted probability was calculated for each model.  
115 The receiver operating characteristic (ROC) curves were calculated for each model by  
116 determining the sensitivity and specificity values at each threshold. The area under the curve  
117 (AUC) was calculated and used as a predictor of discriminative power, and this was compared  
118 between the models.

119 Because the ROC does not naturally have a method to determine confidence intervals,  
120 bootstrapping statistical techniques were used to simulate 1000 new datasets from our original  
121 dataset by sampling with replacement. This was done for each model and the simulations were  
122 used to create confidence intervals for the AUC as well as the ROC curves themselves.

123 Finally, calibration plots were calculated to give a measure of how well the calculated  
124 probabilities compared to the observed outcomes<sup>25</sup>. This was performed by calculating the odds  
125 of success at each decile of predicted probability (e.g., in cases with a predicted probability of  
126 70-79%, what percent of the time was the ECV actually successful?). Quality of calibration was  
127 assessed visually.

128 Calculations were performed with Python version 3.10 using numpy, pandas, and sklearn  
129 (scikit-learn) libraries and in Jupyter notebooks hosted by Google colab. Figures were created  
130 using plotly 5.15. The code is hosted on GitHub and will be published online after publication of  
131 this study. The underlying data was deidentified and cleaned for publication with this manuscript  
132 for reproduction or for future analyses.

133

## 134 **Results**

### 135 *Population demographics*

136 We ultimately obtained data from 125 patients who underwent a total of 132 ECV  
137 attempts, of which 69 were successful (52.2%) as seen in **Table 1** and **Table 2**. 53 of the patients  
138 ultimately had a vaginal delivery including 50 of those with successful ECVs and 3 of those with  
139 failed ECVs who spontaneously verted afterwards. No patients had vaginal breech births. 107  
140 patients were breech, 15 patients were transverse, and 3 were oblique before the first ECV  
141 attempt. Our sample had a largely white population, which is consistent with the demographics  
142 of our geographical region. No patients had oligohydramnios and six patients had  
143 polyhydramnios (AFI  $\geq 24$ ), of which five had mild polyhydramnios (AFI  $< 30$ ) and five had  
144 successful ECVs (all except the patient with moderate polyhydramnios).



145 1 patient did not have any recorded BMI values. 116 patients did not have a recorded  
146 EFW percentile. Only 40 patients had information of breech type (e.g. frank, complete,  
147 incomplete).

148 The underlying data from this study is given in **Supplement 1**.

149

### 150 *Available models – discrimination ability*

151 The following 6 models were studied: Dahl 2021<sup>14</sup>, Bilgory 2023<sup>21,22</sup>, López Pérez  
152 2020<sup>12</sup>, Kok 2011<sup>23</sup>, Burgos 2010<sup>15,16</sup>, and Tasnim 2012 (GNK-PIMS score)<sup>24</sup>. **Table 3** outlines  
153 the varying characteristics of these models including the AUC. In our population, all models had  
154 at least some predictive value (AUC >0.5), with Dahl 2021 having the greatest AUC at 0.78  
155 (**Figure 1**). Bilgory 2023<sup>21,22</sup>, López Pérez 2020<sup>12</sup>, Kok 2011<sup>23</sup>, and Burgos 2010<sup>15,16</sup> all had  
156 similar predictive values with AUC 0.68-0.71. Tasnim 2012<sup>24</sup> had the lowest predictive value in  
157 our study with an AUC of 0.63. When bootstrapping was performed to simulate confidence  
158 intervals for the AUC, Tasnim 2012<sup>24</sup> was not statistically significant with 0.5 being inside the  
159 confidence interval (**Figure 2**). However, the confidence interval was wide for Tasnim 2012  
160 because the majority of our data did not have fetal presentation (e.g., frank vs complete vs  
161 incomplete breech) recorded and thus was not included in the analysis of Tasnim 2012. The  
162 confidence intervals for the AUC overlapped between the majority of the models so confirmation  
163 of which model had the best discriminative ability was not able to be done. A bootstrapped ROC  
164 curve is shown for Dahl 2021 in **Figure 3** and a table of bootstrapped AUC value for all the  
165 models is shown in **Supplement 2**.

## 166 *Available models – calibration*

167 Dahl 2021<sup>14</sup> was relatively well calibrated in our population (**Figure 4**). Bilgory  
168 2023<sup>21,22</sup> had poor calibration without a specific tendency to over or under calibrate. López  
169 Pérez 2020 was consistently under-calibrated. That is, it would consistently underestimate the  
170 probability of success in our population by a relatively fixed amount. The remaining three  
171 models return score values instead of predicted probabilities. Thus, calibration curves cannot be  
172 properly assessed. A probability of success by score value is shown for Kok 2011, Burgos 2010,  
173 and Tasnim 2012 in **Figure 5**. The probability of success poorly correlated with score for Kok  
174 2011 and Tasnim 2012, but correlated well for Burgos 2010.

175

## 176 **Discussion**

177 The prediction of ECV success has the potential to be an incredibly powerful tool in  
178 counseling patients as they weigh the risks and benefits of the procedure. Prediction models have  
179 the opportunity to give specific, personalized information to patients. However, they have the  
180 ability to cause harm if they are inaccurate or poorly calibrated. In reality, confidently providing  
181 a patient an inaccurate odds of success is worse than providing no prediction at all.

182 Unfortunately, there has not been a clinical consensus on the optimal tool despite dozens of  
183 models having been proposed. Our prior systematic review identified lack of external validation  
184 studies as a major gap in the literature, prohibiting clinical use of published models.

185 In this article, we set out to validate the 6 models that could be used in our population as  
186 they used variables which are already collected as part of routine care at our institution. These  
187 models are Dahl 2021<sup>14</sup>, Bilgory 2023<sup>21,22</sup>, López Pérez 2020<sup>12</sup>, Kok 2011<sup>23</sup>, Burgos 2010<sup>15,16</sup>,

188 and Tasnim 2012 (GNK-PIMS score)<sup>24</sup>. Our data demonstrates that all of these models have at  
189 least some predictive value, with the exception of Tasnim 2012 (which was not statistically  
190 significant by bootstrapping, likely because of the small number of patients with sufficient  
191 recorded data to calculate this model). Of all the models, Dahl 2021 had the greatest AUC,  
192 although the bootstrapped confidence interval still overlapped with the other models. Of the three  
193 models that output a specific probability, only Dahl 2021 appeared overall well calibrated. Based  
194 off the available evidence from this study, Dahl 2021 appears to overall be the best model for  
195 clinical use.

196 This is the first external validation study of Dahl 2021. While it is generally expected that  
197 external validation studies will show worse predictive values than the original study because of  
198 some degree of unavoidable statistical bias, the opposite was true here. Dahl 2021's AUC in our  
199 study was 0.779 (confidence interval of 0.712-0.844), which is higher than the original study,  
200 which showed an AUC of 0.667 (confidence interval of 0.634-0.701). Overall, the results are  
201 very promising for Dahl 2021.

202 There are some limitations with our study. Due to the retrospective nature of our study,  
203 we only evaluated models that used variables routinely measured in our institution. Our  
204 institution does not routinely measure uterine tone<sup>26,27</sup>, cervical exam/fetal station<sup>28-30</sup> and  
205 forebag size<sup>31</sup>, which are components of other models. One of these models may in fact have  
206 greater predictive value than the ones we evaluated in our study. On the other hand, these models  
207 would also be less practical to implement in clinical practice because they would require  
208 measuring additional variables. Additionally, these variables would not be known until close to  
209 ECV; thus, it would be hard to counsel patients in advance using one of these models. Since

210 many models required cervical exam or fetal station, we suggest future research study be a  
211 prospective trial where cervical exam is performed.

212

## 213 **Declarations**

### 214 *Ethics approval and consent to participate*

215 No new unpublished data from humans or other animals was included in this study.

### 216 *Consent for publication*

217 Not applicable.

### 218 *Funding*

219 The authors declare that no funds, grants, or other support were received specifically for  
220 the preparation of this manuscript. The authors have no conflicts of interest to declare.

### 221 *Availability of data and materials*

222 The data is published in the supplemental data. The code is hosted in a GitHub  
223 repository, which will be publicly available after publication of this manuscript.

### 224 *Acknowledgments*

225 We thank artificial intelligence engineer Benjamin M. Pikus for his help in deciding how  
226 to evaluate prediction model performance.

227

228

## 229 **Tables**

230

231 **Table 1. Demographic information and other baseline characteristics**

	Count	%
<b><i>ECVs Performed</i></b>		
1	121	97%
2	2	2%
3	1	1%
4	1	1%
<b><i>Parity</i></b>		
0	54	43%
1	32	26%
2	24	19%
3	10	8%
≥4	5	4%
<b><i>Race</i></b>		
White	85	69%
Black or African American	21	16%
Asian	11	9%
American Indian / Alaskan Native	1	1%
Patient refused	1	1%
All other/none of the above	14	11%
<b><i>Ethnicity</i></b>		
Not Hispanic or Latino	111	89%
Hispanic or Latino	13	10%
Unknown/patient refused	1	1%
<b><i>Fetal Sex</i></b>		
Female	68	54%
Male	57	46%
<b><i>Delivery</i></b>		
<b>Cesarean - Total</b>	<b>82</b>	<b>64%</b>
Low transverse, primary	69	53%
Low transverse, repeat	7	5%
Low transverse, unspecified	5	4%
Vertical, primary	1	1%
<b>Vaginal - Total</b>	<b>47</b>	<b>36%</b>
Vacuum-assisted	3	2.4%
Spontaneous	44	34%
<b><i>Presentation / Lie</i></b>		
<b>Breech - Total</b>	<b>107</b>	<b>86%</b>
Frank	25	20%
Footling	7	6%
Complete	7	6%
Incomplete	1	1%
Other/unknown breech	67	54%
<b>Transverse</b>	<b>15</b>	<b>12%</b>
<b>Oblique</b>	<b>3</b>	<b>2%</b>

232

233 **Table 2. Baseline characteristics continued**

Variable	Mean ± SD	Min - Max
Maternal age (yr)	30.2 ± 4.9	18 - 40
Initial BMI (kg/m <sup>2</sup> )	28.2 ± 6.7	17.6 - 46.3
Last BMI (kg/m <sup>2</sup> )	31.1 ± 6.3	19.9 - 47.3
GA at ECV (wk)	37.9 ± 0.92	36w 4d - 40w 6d
GA at delivery (wk)	39.3 ± 0.86	37w 1d - 41w 3d
ECV to delivery time (days)	9.6 ± 7.34	0 - 27
AFI	14.7 ± 4.96	5.73 - 34.97
EFW at last US (percentile), n=112	46.8 ± 19.9	9 - 95
EFW, calculated at ECV (g), n=112	3121 ± 298	2597 - 4277

234

235 **Table 3. Background on prediction models**

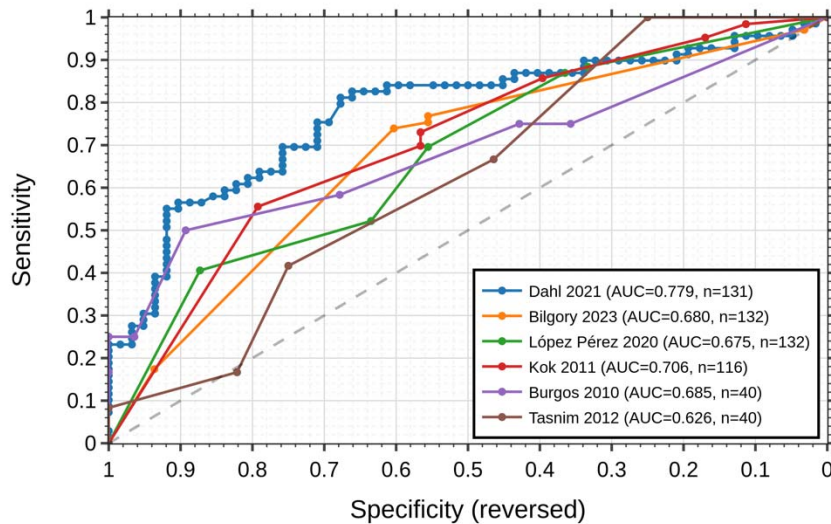
Model	Output	Predictive factors	Prior validation statistics	Current study AUC
<b>Dahl 2021</b> (Chicago, USA)	Probability (logistic)	Parity, BMI, placenta location, fetal lie	<b>Internal validation</b> AUC: 0.667 [0.634-0.701]	AUC: 0.779 [0.712-0.844]
<b>Bilgory 2023</b> (Bnei Brak, Israel)	Probability	Parity, AFV	<b>Training data</b> AUC: 0.71 [0.672– 0.748]	AUC: 0.680 [0.611-0.751]
<b>López Pérez 2020</b> (Murcia, Spain)	Probability	Parity, AFV, placenta location	<b>Training data</b> Accuracy: 74.0% Sensitivity: 98.8%, Specificity: 12.3%	AUC: 0.675 [0.597-0.748]
<b>Kok 2011</b> (Amsterdam, The Netherlands)	Score (9-22)	Parity, AFV, placenta location, EFW	<b>Internal validation</b> AUC: 0.738 [0.695–0.782] <b>External validation</b> AUC: 0.669 [0.622– 0.716]	AUC: 0.706 [0.628-0.777]
<b>Burgos 2010</b> (Basque Country, Spain)	Score (4-14)	Parity, AFV, placenta location, breech type	<b>Training data</b> AUC: 0.738 [0.695– 0.782] <b>External validation</b> AUC: 0.669 [0.622– 0.716]	AUC: 0.685 [0.512-0.845]
<b>GNK-PIMS / Tasnim 2012</b> (Islamabad, Pakistan)	Score (0-10)	Parity, AFV, BMI, GA, breech type	None	AUC: 0.626 [0.471-0.766]

236 All models are logistic regression models or are score charts derived from them. AFV, amniotic  
 237 fluid volume; EFW, estimated fetal weight; AUC, area under the ROC curve. Brackets represent  
 238 95% confidence intervals.

239

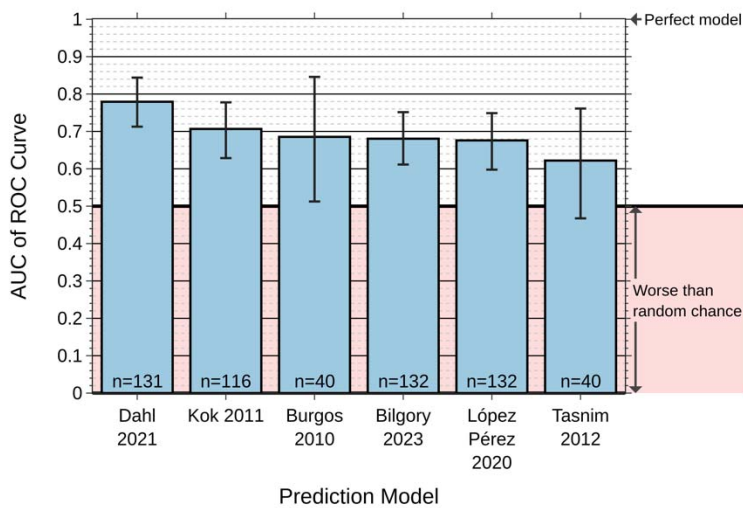
## 240 Figures

241 **Figure 1. ROC curves of all models.** The predictive ability of all the models is displayed using  
242 receiver operating characteristic (ROC) curves and the area under the curve (AUC).



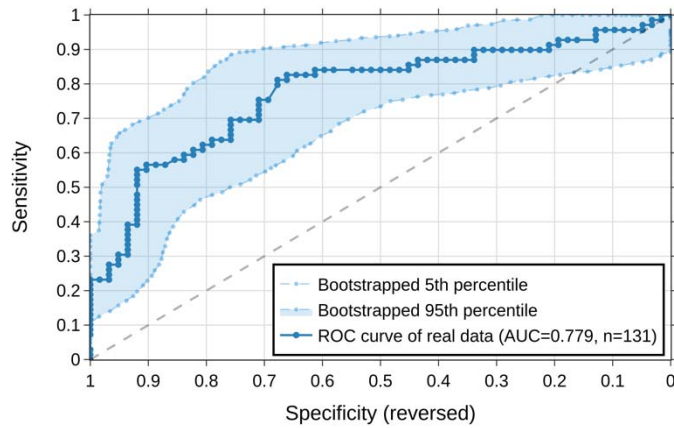
243

244 **Figure 2. AUC of models with bootstrapping.** The AUC of all the models is displayed as a bar  
245 chart for comparison. 95% confidence intervals were estimated using bootstrapping from 1,000  
246 simulated datasets



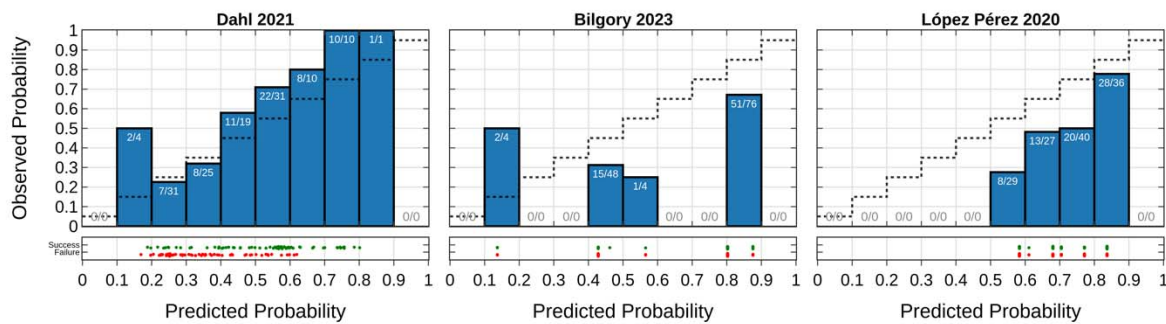
247

248 **Figure 3 ROC curve for Dahl 2021 with bootstrapping**



249

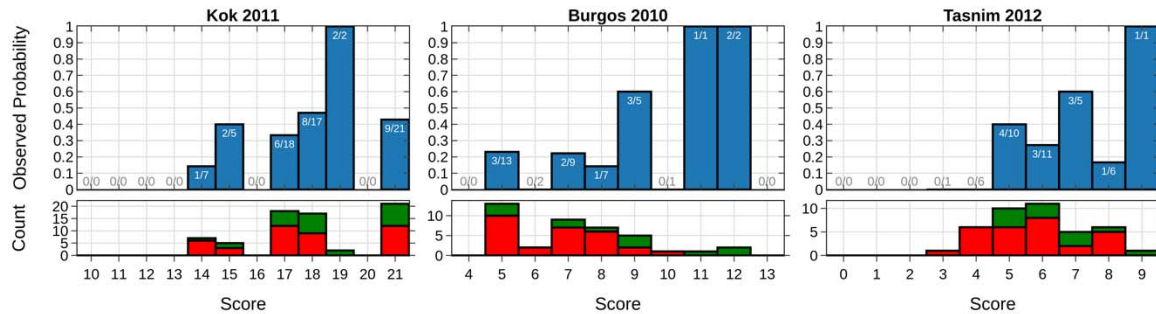
250 **Figure 4. Calibration curves.** The calibration is displayed for the models that output  
 251 probabilities: Dahl 2021, Bilgory 2023, and López Pérez 2020. This figure shows how well the  
 252 predicted probabilities match the observed probability. Only Dahl 2021 has a good match.



253



254 **Figure 5 Observed probability by predicted score.** This is shown as an alternative to the  
 255 calibration curve for the models (Kok 2011, Burgos 2010, Tasnim 2012) that do not output a  
 256 probability score.



257  
 258

259 **Supplement**

260 **Supplement 1 Deidentified patient data to reproduce the analyses.** For yes/no questions, 0:  
 261 no, 1: yes. For baby gender, 1: female, 2: male. For ethnicity, 11: Hispanic, 12: not Hispanic, 13:  
 262 unknown/patient refused. For language, 1: English, 2: French, 3: Spanish, 15: Portuguese. For  
 263 patient race, 1: White, 2: Black, 3: none of the above/all other races, 4: Hispanic, 5: Asian, 6:  
 264 American Indian or Alaskan Native, 7: Pacific Islander or Native Hawaiian, 8: Patient refused.  
 265 EFW at ECV and at delivery were calculated using the percentile from the most recent  
 266 ultrasound as described in the methods. ECV success for the first, second, third, and fourth ECV  
 267 attempts is given by the variables `ecv_successful_1`, `ecv_successful_2`, etc, respectively.  
 268 Maternal age at delivery (`age_delivery`) is given in years. Gestational ages are given in days.  
 269 EFWs are given in grams. Patient weights and total weight gain (`twg`) is given in pounds, while  
 270 BMI is given in kg/m<sup>2</sup>

271

272 **Supplement 2 Area under the curve data with bootstrapping.** AUC values >0.5 are  
 273 highlighted

	Original Data		1000 Bootstrapped Simulations								
			Center		Percentiles						
	n	AUC	AUC Mean	AUC SD	AUC Min	AUC 1%	AUC 5%	AUC Median	AUC 95%	AUC 99%	AUC Max
<b>Dahl 2021</b>	131	0.7791	0.7794	0.0414	0.6496	0.6797	0.7124	0.7817	0.8437	0.8692	0.8891
<b>Bilgory 2023</b>	132	0.6800	0.6805	0.0433	0.5581	0.5764	0.6108	0.6802	0.7510	0.7723	0.8310
<b>López Pérez 2020</b>	132	0.6754	0.6760	0.0464	0.5161	0.5644	0.5972	0.6775	0.7482	0.7722	0.7960
<b>Kok 2011</b>	116	0.7063	0.7066	0.0456	0.5693	0.5978	0.6282	0.7089	0.7773	0.8018	0.8389
<b>Burgos 2010</b>	40	0.6845	0.6854	0.1021	0.2396	0.4273	0.5116	0.6859	0.8449	0.8899	0.9436
<b>Tasnim 2012</b>	40	0.6265	0.6221	0.0879	0.3177	0.4121	0.4718	0.6225	0.7656	0.8219	0.8860

274

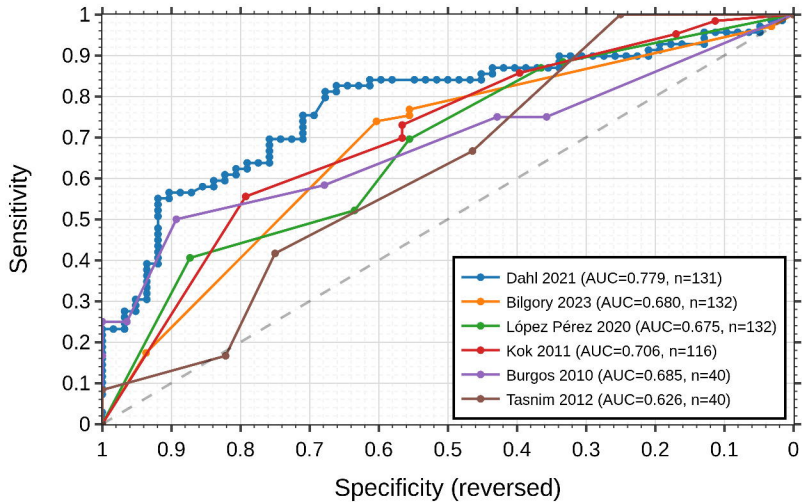
## 275 **References**

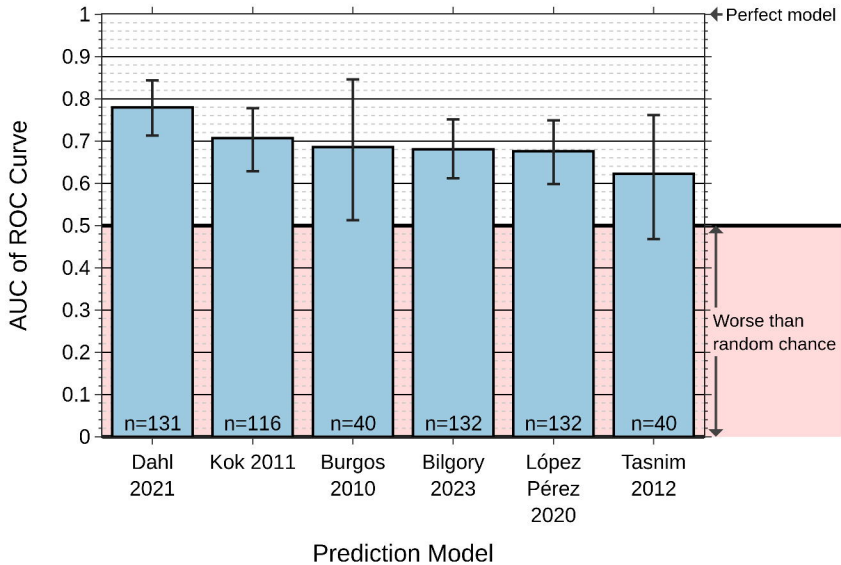
- 276 1. Richmond AK, Ashworth JR. Management of malposition and malpresentation in labour.  
277 *Obstet Gynaecol Reprod Med.* 2023;33(11):325-333. doi:10.1016/j.ogrm.2023.08.004
- 278 2. External Cephalic Version: ACOG Practice Bulletin, Number 221. *Obstet Gynecol.*  
279 2020;135(5):e203-e212. doi:10.1097/AOG.0000000000003837
- 280 3. Hill LM. Prevalence of breech presentation by gestational age. *Am J Perinatol.* 1990;7(1):92-  
281 93. doi:10.1055/s-2007-999455
- 282 4. ACOG Committee Opinion No. 745: Mode of Term Singleton Breech Delivery. *Obstet*  
283 *Gynecol.* 2018;132(2):e60-e63. doi:10.1097/AOG.0000000000002755
- 284 5. External Cephalic Version and Reducing the Incidence of Term Breech Presentation. *BJOG*  
285 *Int J Obstet Gynaecol.* 2017;124(7):e178-e192. doi:10.1111/1471-0528.14466
- 286 6. Nalam RL, Chinnachamy P, Emmanuel P. External Cephalic Version: A Dying Art Worth  
287 Reviving. *J Obstet Gynecol India.* 2018;68(6):493-497. doi:10.1007/s13224-018-1090-z
- 288 7. Grootscholten K, Kok M, Oei SG, Mol BWJ, van der Post JA. External cephalic version-  
289 related risks: a meta-analysis. *Obstet Gynecol.* 2008;112(5):1143-1151.  
290 doi:10.1097/AOG.0b013e31818b4ade
- 291 8. Ben-Meir A, Erez Y, Sela HY, Shveiky D, Tsafrir A, Ezra Y. Prognostic parameters for  
292 successful external cephalic version. *J Matern-Fetal Neonatal Med Off J Eur Assoc Perinat*  
293 *Med Fed Asia Ocean Perinat Soc Int Soc Perinat Obstet.* 2008;21(9):660-662.  
294 doi:10.1080/14767050802244938
- 295 9. Ebner F, Friedl TWP, Leinert E, et al. Predictors for a successful external cephalic version: a  
296 single centre experience. *Arch Gynecol Obstet.* 2016;293(4):749-755. doi:10.1007/s00404-  
297 015-3902-z
- 298 10. Kok M, Cnossen J, Gravendeel L, van der Post J, Opmeer B, Mol BW. Clinical factors to  
299 predict the outcome of external cephalic version: a metaanalysis. *Am J Obstet Gynecol.*  
300 2008;199(6):630.e1-7; discussion e1-5. doi:10.1016/j.ajog.2008.03.008
- 301 11. Chaudhary S, Contag S, Yao R. The impact of maternal body mass index on external  
302 cephalic version success. *J Matern-Fetal Neonatal Med Off J Eur Assoc Perinat Med Fed*  
303 *Asia Ocean Perinat Soc Int Soc Perinat Obstet.* 2019;32(13):2159-2165.  
304 doi:10.1080/14767058.2018.1427721
- 305 12. López-Pérez R, Lorente-Fernández M, Velasco-Martínez M, Martínez-Cendán JP. Prediction  
306 model of success for external cephalic version. Complications and perinatal outcomes after a  
307 successful version. *J Obstet Gynaecol Res.* 2020;46(10):2002-2009. doi:10.1111/jog.14385

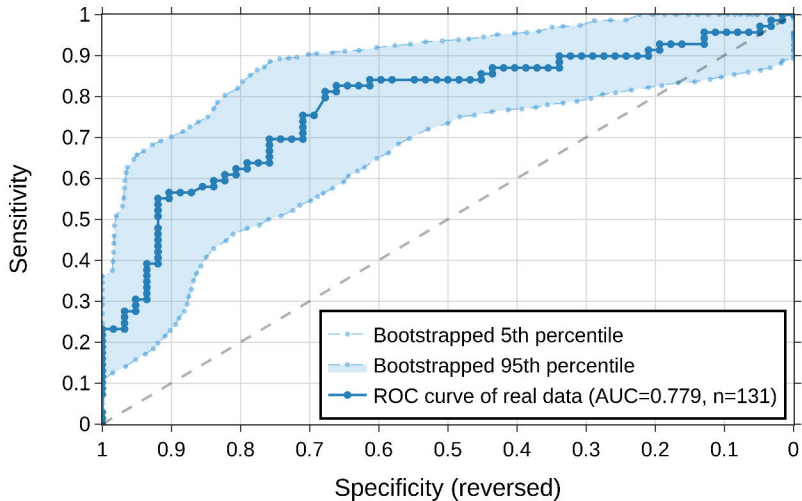
- 308 13. Yerrabelli RS, Lee C, Palsgaard PK, Lauinger AR, Abdelsalam O, Jennings V. Prediction  
309 Models for Successful External Cephalic Version: An Updated Systematic Review. *Am J*  
310 *Perinatol*. Published online January 12, 2024. doi:10.1055/a-2211-4806
- 311 14. Dahl CM, Zhang Y, Ong JX, et al. A Multivariable Predictive Model for Success of External  
312 Cephalic Version. *Obstet Gynecol*. 2021;138(3):426-433.  
313 doi:10.1097/AOG.0000000000004518
- 314 15. Burgos J, Melchor JC, Pijoán JI, Cobos P, Fernández-Llebarez L, Martínez-Astorquiza T. A  
315 prospective study of the factors associated with the success rate of external cephalic version  
316 for breech presentation at term. *Int J Gynecol Obstet*. 2011;112(1):48-51.  
317 doi:10.1016/j.ijgo.2010.07.023
- 318 16. Burgos J, Cobos P, Rodriguez L, et al. Clinical score for the outcome of external cephalic  
319 version: A two-phase prospective study: Clinical score for external cephalic version. *Aust N*  
320 *Z J Obstet Gynaecol*. 2012;52(1):59-61. doi:10.1111/j.1479-828X.2011.01386.x
- 321 17. Reddy UM, Abuhamad AZ, Levine D, Saade GR. Fetal Imaging: Executive Summary of a  
322 Joint Eunice Kennedy Shriver National Institute of Child Health and Human Development,  
323 Society for Maternal-Fetal Medicine, American Institute of Ultrasound in Medicine,  
324 American College of Obstetricians and Gynecologists, American College of Radiology,  
325 Society for Pediatric Radiology, and Society of Radiologists in Ultrasound Fetal Imaging  
326 Workshop. *Obstet Gynecol*. 2014;123(5):1070-1082. doi:10.1097/AOG.0000000000000245
- 327 18. Society for Maternal-Fetal Medicine (SMFM). Electronic address: [pubs@smfm.org](mailto:pubs@smfm.org), Dashe  
328 JS, Pressman EK, Hibbard JU. SMFM Consult Series #46: Evaluation and management of  
329 polyhydramnios. *Am J Obstet Gynecol*. 2018;219(4):B2-B8. doi:10.1016/j.ajog.2018.07.016
- 330 19. Kiserud T, Piaggio G, Carroli G, et al. The World Health Organization Fetal Growth Charts:  
331 A Multinational Longitudinal Study of Ultrasound Biometric Measurements and Estimated  
332 Fetal Weight. *PLoS Med*. 2017;14(1). doi:10.1371/journal.pmed.1002220
- 333 20. Kiserud T, Benachi A, Hecher K, et al. The World Health Organization fetal growth charts:  
334 concept, findings, interpretation, and application. *Am J Obstet Gynecol*. 2018;218(2):S619-  
335 S629. doi:10.1016/j.ajog.2017.12.010
- 336 21. Bilgory A, Minich O, Shvaikovsky M, Gurevich G, Lessing JB, Olteanu I. Predictive Factors  
337 for Successful Vaginal Delivery after a Trial of External Cephalic Version: A Retrospective  
338 Cohort Study of 946 Women. *Am J Perinatol*. 2023;40(15):1679-1686. doi:10.1055/s-0041-  
339 1739505
- 340 22. Bilgory A, Minich O, Shvaikovsky M, Gurevich G, Lessing JB, Olteanu I. Erratum:  
341 Predictive Factors for Successful Vaginal Delivery after a Trial of External Cephalic  
342 Version: A Retrospective Cohort Study of 946 Women. *Am J Perinatol*. Published online  
343 January 4, 2022. doi:10.1055/s-0041-1742110

- 344 23. De Hundt M, Vlemmix F, Kok M, et al. External Validation of a Prediction Model for  
345 Successful External Cephalic Version. *Am J Perinatol*. 2012;29(03):231-236. doi:10.1055/s-  
346 0031-1285098
- 347 24. Tasnim N, Mahmud G, Javaid K. GNK-PIMS Score: A Predictive Model for Success of  
348 External Cephalic Version. *J South Asian Fed Obstet Gynaecol*. 2012;4(2):99-102.  
349 doi:10.5005/jp-journals-10006-1184
- 350 25. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A  
351 calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin*  
352 *Epidemiol*. 2016;74:167-176. doi:10.1016/j.jclinepi.2015.12.005
- 353 26. Anand K, Keepanasseril A, Amala R, Nair NS. Development and validation of a clinical  
354 score to predict the probability of successful procedure in women undergoing external  
355 cephalic version. *J Matern-Fetal Neonatal Med Off J Eur Assoc Perinat Med Fed Asia*  
356 *Ocean Perinat Soc Int Soc Perinat Obstet*. 2019;34(18):2925-2931.  
357 doi:10.1080/14767058.2019.1674803
- 358 27. Wong WM, Lao TT, Liu KL. Predicting the success of external cephalic version with a  
359 scoring system. A prospective, two-phase study. *J Reprod Med*. 2000;45(3):201-206.
- 360 28. Zheng LG, Zhang HL, Chen RX, et al. Scoring system to predict the success rate of external  
361 cephalic versions and determine the timing of the procedure. *Eur Rev Med Pharmacol Sci*.  
362 2021;25(1):45-55. doi:10.26355/eurrev\_202101\_24345
- 363 29. Hutton EK, Simioni JC, Thabane L. Predictors of success of external cephalic version and  
364 cephalic presentation at birth among 1253 women with non-cephalic presentation using  
365 logistic regression and classification tree analyses. *Acta Obstet Gynecol Scand*.  
366 2017;96(8):1012-1020. doi:10.1111/aogs.13161
- 367 30. Newman RB, Peacock BS, Peter VanDorsten J, Hunt HH. Predicting success of external  
368 cephalic version. *Am J Obstet Gynecol*. 1993;169(2):245-250. doi:10.1016/0002-  
369 9378(93)90071-P
- 370 31. Isakov O, Reicher L, Lavie A, Yogev Y, Maslovitz S. Prediction of Success in External  
371 Cephalic Version for Breech Presentation at Term. *Obstet Gynecol*. 2019;133(5):857-866.  
372 doi:10.1097/AOG.0000000000003196

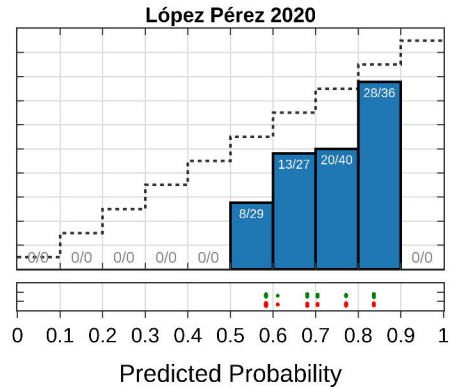
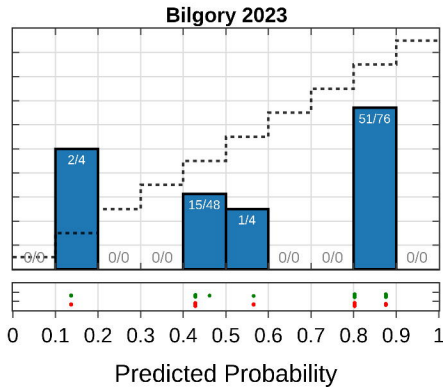
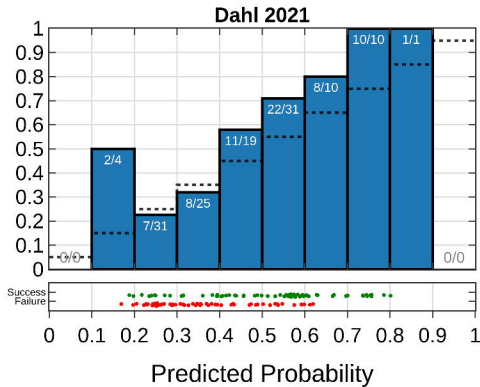
373  
374





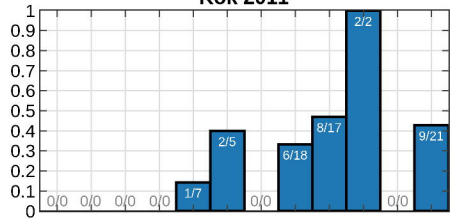


Observed Probability

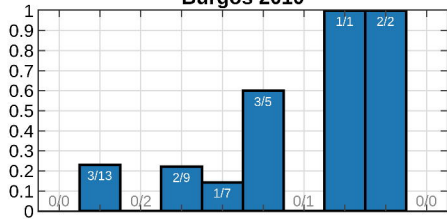




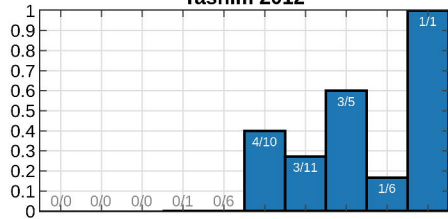
Kok 2011



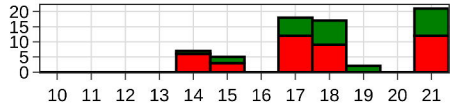
Burgos 2010



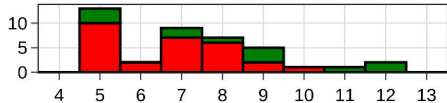
Tasnim 2012



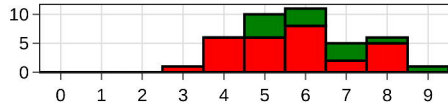
Count



Score



Score



Score