TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Prognostic model for predicting survival in very preterm infants

Citation for published version (APA):
EPI-DAF study group, & Andriessen, P. (2022). Prognostic model for predicting survival in very preterm infants: an external validation study. *BJOG : An International Journal of Obstetrics and Gynaecology*, *129*(4), 529-538. https://doi.org/10.1111/1471-0528.17010

Document license:
TAVERNE

DOI:
10.1111/1471-0528.17010

Document status and date:
Published: 01/03/2022

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 16. Nov. 2023

DOI: 10.1111/1471-0528.17010
www.bjog.org

Research Article
Epidemiology

Prognostic model for predicting survival in very preterm infants: an external validation study

PE van Beek,[a] F Groenendaal,[b] W Onland,[c] S Koole,[d] PH Dijk,[e] KP Dijkman,[a] FAM van den Dungen,[c] AFJ van Heijst,[f] RF Kornelisse,[g] FABA Schuerman,[h] E van Westering-Kroon,[i] RSGM Witlox,[j] P Andriessen,[a,k] E Schuit,[l] on behalf of EPI-DAF study group*

[a] Department of Neonatology, Máxima Medical Centre, Veldhoven, The Netherlands [b] Department of Neonatology, Wilhelmina Children's Hospital, University Medical Centre Utrecht and Utrecht University, Utrecht, The Netherlands [c] Department of Neonatology, Emma Children's Hospital, Amsterdam University Medical Centres, VU University Medical Centre, University of Amsterdam, Amsterdam, The Netherlands [d] The Netherlands Perinatal Registry, Utrecht, The Netherlands [e] Department of Neonatology, Beatrix Children's Hospital, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands [f] Department of Neonatology, Amalia Children's Hospital, Radboud University Medical Centre, Nijmegen, The Netherlands [g] Department of Paediatrics, Division of Neonatology, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands [h] Department of Neonatology, Isala Clinics, Zwolle, The Netherlands [i] Department of Neonatology, Maastricht University Medical Centre, Maastricht, The Netherlands [j] Department of Neonatology, Willem-Alexander Children's Hospital, Leiden University Medical Centre, Leiden, The Netherlands [k] Department of Applied Physics, School of Medical Physics and Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands [l] Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Utrecht, The Netherlands
Correspondence: PE van Beek, Department of Neonatology, Máxima Medical Centre, Veldhoven, The Netherlands. De Run 4600, Post Office Box 7777, 5500 MB Veldhoven, The Netherlands. Email: pauline.van.beek@mmc.nl

Accepted 5 November 2021. Published Online 30 November 2021.

Objective To perform a temporal and geographical validation of a prognostic model, considered of highest methodological quality in a recently published systematic review, for predicting survival in very preterm infants admitted to the neonatal intensive care unit. The original model was developed in the UK and included gestational age, birthweight and gender.

Design External validation study in a population-based cohort.

Setting Dutch neonatal wards.

Population or sample All admitted white, singleton infants born between 23^{+0} and 32^{+6} weeks of gestation between 1 January 2015 and 31 December 2019. Additionally, the model's performance was assessed in four populations of admitted infants born between 24^{+0} and 31^{+6} weeks of gestation: white singletons, non-white singletons, all singletons and all multiples.

Methods The original model was applied in all five validation sets. Model performance was assessed in terms of calibration and discrimination and, if indicated, it was updated.

Main outcome measures Calibration (calibration-in-the-large and calibration slope) and discrimination (*c* statistic).

Results Out of 6092 infants, 5659 (92.9%) survived. The model showed good external validity as indicated by good discrimination (*c* statistic 0.82, 95% CI 0.79–0.84) and calibration (calibration-in-the-large 0.003, calibration slope 0.92, 95% CI 0.84–1.00). The model also showed good external validity in the other singleton populations, but required a small intercept update in the multiples population.

Conclusions A high-quality prognostic model predicting survival in very preterm infants had good external validity in an independent, nationwide cohort. The accurate performance of the model indicates that after impact assessment, implementation of the model in clinical practice in the neonatal intensive care unit could be considered.

Keywords External validation, mortality, prediction model, very preterm infants.

Tweetable abstract A high-quality model predicting survival in very preterm infants is externally valid in an independent cohort.

Linked article This article is commented on by EM McClure & RL Goldenberg pp. 539 in this issue. To view this minicommentary visit https://doi.org/10.1111/1471-0528.17014.

Please cite this paper as: van Beek PE. , Groenendaal F. , Onland W. , Koole S. , Dijk PH. , Dijkman KP. , van den Dungen FAM. , van Heijst AFJ. , Kornelisse RF. , Schuerman FABA. , van Westering-Kroon E. , Witlox RSGM. , Andriessen P. , Schuit E. ; on behalf of EPI-DAF study group. Prognostic model for predicting survival in very preterm infants: an external validation study. BJOG 2022;129:529–538.

*A complete list of study group members appears in the 32Acknowledgements.

© 2021 John Wiley & Sons Ltd.

Introduction

Very preterm birth before 32 weeks of gestation occurs in 1.3% of all live births in developed regions and is associated with increased perinatal mortality and neonatal morbidity.[1,2] Despite this low prevalence, complications associated with preterm birth are responsible for 35% of the world's annual neonatal deaths.[3] Accurate risk assessment of perinatal death in very preterm infants can help caregivers and parents to decide whether and when to intervene in a pregnancy or to adjust postnatal care.[4] Prognostic prediction models can be a helpful tool in performing such risk assessment.[5,6]

A recently published systematic review showed that there is an abundance of mortality risk prediction models for very preterm infants, identifying 142 models from 35 studies reporting on model development.[7] Unfortunately, many of the models are of unknown value for daily practice because of a lack of external validation, meaning it is unclear to what extent these models can be used in new patients. Instead of developing yet another model, emphasis should be on external validation and if necessary, adaption of existing models.[8,9]

A logical next step was to externally validate the highest methodological quality model identified in the systematic review. Extensive risk of bias assessment showed that a model published by Manktelow et al. had the highest quality among all 142 models (according to PROBAST), with low risk of bias for all items except one using the PROBAST tool for risk of bias assessment.[10–12] This model was therefore considered as having high potential for use in clinical practice. Because local and timely variation in attitudes towards the care of very preterm infants can influence the validity of the model, a temporal and geographical external validation and, if needed, updating were performed. The aim of this study was to evaluate the extent to which the UK model of Manktelow et al. can be applied to very preterm born infants from another population with different healthcare settings and infrastructure, using a nationwide Dutch perinatal registry.

Methods

This validation study is reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement.[13] There was no patient or public involvement.

Prognostic model

In this study, the prognostic model developed by Manktelow et al. was externally validated. Manktelow et al. included all white singleton infants born at 23^{+0} to 32^{+6} weeks of gestation between 1 January 2008 and 31 December 2010.[12] Infants were excluded if they had lethal congenital anomalies, missing data on gender or indeterminate gender, or implausible birthweight for gestational age (>3 standard deviations from the median for their gestational age and gender). A logistic regression model was developed to predict the probability of survival before discharge. The final model included gestational age, defined as the infant's gestational age in completed weeks and days as a decimal; birthweight, defined as the infant's birthweight in kilograms; gender, with 0 for males and 1 for female; and an indicator variable named *gest23* for birth before 24^{+0} weeks of gestation, with 1 if the gestation at birth was less than 24^{+0} weeks.[12] The regression formula to predict survival was as follows:

$$\mathrm{P(Survival)} = \frac{1}{1 + \exp(-\mathrm{LP(Survival)})},$$

with

$$\mathrm{LP(Survival)} = -10.28 + 0.46 \cdot \mathrm{gestation} + 0.47 \cdot \frac{1}{\mathrm{birthweight}^2} + 0.45 \cdot \mathrm{gender} + -0.28 \cdot \mathit{gest23}.$$

Patient population in the validation study

To validate the model, the prognostic model was applied to five different study populations. First, the model was externally validated in a population exactly the same as the population used for model development by Manktelow et al., i.e. all white, singleton infants born between 23^{+0} and 32^{+6} weeks of gestation between 1 January 2015 and 31 December 2019, who were admitted to the neonatal ward.

To assess whether the model was also externally valid in a more general very preterm infant population including non-white infants and multiple births, the model was subsequently applied to the following four populations, thereby taking into account that only infants born from 24^{+0} to 31^{+6} weeks are routinely admitted to one of the Dutch neonatal intensive care units (NICU): white singletons, non-white singletons, all singletons and all multiples. Similar to the study by Manktelow et al., infants were excluded if they had lethal congenital anomalies, which included anencephaly, encephalocele, bilateral renal agenesis, hydrops fetalis, trisomy 13 and trisomy 18.

Data collection

For this study, data from the Netherlands Perinatal Registry (Perined) were used. This registry contains linked population-based information regarding pregnancy, delivery, (re)admissions and pregnancy outcomes, as registered by midwives, obstetricians and paediatricians/neonatologists

© 2021 John Wiley & Sons Ltd.

and covers approximately 99% of all births from 16^{+0} weeks of gestation in the Netherlands. Variables used from the registry for this study included information on birthweight, sex, gestational age, ethnicity, multiplicity, 5-minute Apgar score, mode of delivery and survival.

Outcome

The predicted outcome was defined as live discharge home in infants admitted to the neonatal ward, excluding delivery room deaths.

Sample size calculation

In general, at least 100 events, i.e. very preterm born infants that died or are alive, whichever group is smallest, are needed for a reliable assessment of a model's external validity, otherwise the risk for biased estimates of model performance becomes more likely.[14–16] With an incidence of mortality of approximately 7%, this means that we needed at least 1500 infants in our validation population.

Statistical analysis

No missingness occurred among variables used as predictors in the model. Baseline characteristics of the validation and development population were presented as mean ± standard deviation for continuous variables and as number and percentages of the whole population for categorical variables. The model from Manktelow et al. was externally validated by application of the aforementioned regression formula to each individual in the Dutch validation population to calculate a chance of survival. The validity of the prognostic model was assessed in terms of calibration and discrimination using the rms and pROC R packages. Calibration refers to agreement between the predicted chance of survival and the observed proportion of infants who survived in the validation population over the entire possible range of predicted chances. Calibration was assessed graphically with a calibration plot, and numerically with the calibration slope (b) and calibration-in-the-large (a), given that the calibration slope is set to 1 ($a/b = 1$, calibration-in-the-large), as proposed by Cox.[17] Discrimination refers to the ability of the model to distinguish between infants who died and infants who survived, by assigning the highest chance in a random pair of a deceased and alive infant to the infant that survived. Discrimination was assessed with the concordance (c) statistic, which for a binary outcome is identical to the area under the receiver operating characteristic curve. The need for adjustment of the model to the validation population was assessed by assessment of calibration and discrimination and by using a closed testing procedure, a method that subsequently assesses different levels of model adjustment, i.e. none, recalibration-in-the-large (updating the model intercept), recalibration (updating both the model intercept and overall calibration slope) or complete model revision (fitting the regression coefficients of the original model anew).[18] In this closed testing procedure, a likelihood ratio test statistic was used to test if an updated model provided a statistically significant ($P < 0.05$) better fit compared with the original model. All analyses were performed in R version 3.5.2.

Funding

Drs. van Beek was supported by an unrestricted grant from *Stichting Tiny & Anny van Doorne Fonds*. The funding source had no role in the design, conduct, analyses, or reporting of the study or in the decision to submit the manuscript for publication. The other authors received no external funding.

Results

Patient population

Between 1 January 2015 and 31 December 2019, 6092 infants were included in this study when using the similar study population compared with Manktelow et al. Of these infants, 5659 (92.9%) infants survived. The characteristics of the Dutch validation population are shown in Table 1 (third column) next to the infants of the study from Manktelow (second column) that were used to develop the original model, which were comparable, except for the number of 23-week gestation infants. The characteristics of the other NICU validation cohorts (fourth to seventh columns) were comparable to the characteristics of the development cohort, except for gestational age, which was lower because infants were admitted up to 31^{+6} weeks of gestation compared with 32^{+6} weeks of gestation in the development population.

Performance in the validation population similar to the development population

When applying the original model to the validation set, the predicted chance of survival to discharge was in agreement with the observed proportion of survival to discharge for infants admitted (Figure 1A). The mean predicted chance and overall observed proportion for survival were similar (92.6% and 92.9%, respectively), as also indicated by a calibration-in-the-large of 0.003. The calibration slope b was 0.92 (95% CI 0.84–1.00). The discriminative ability of the model was good, with a c statistic of 0.82 (95% CI 0.79–0.84). A difference of 35.4% between the means of the highest (99.3%) and lowest (63.9%) predicted chance deciles was observed, with a range in predicted chances varying from 9.6 to 99.5%. Based on the good performance of the original model and indicated by the closed testing procedure (chi-square = 10.3, df = 5, $P = 0.067$ for testing the model after recalibration-in-the-large against the original model), keeping the original model was favoured. The estimated predicted probabilities on survival according to

© 2021 John Wiley & Sons Ltd.

Table 1. Baseline characteristics

Population	UK Development population[12]	Dutch Validation populations				
		Population similar to the development population	Other populations			
	White, singleton, 23^{+0}–32^{+6} weeks	White, singleton, 23^{+0}–32^{+6} weeks	White, singleton, 24^{+0}–31^{+6} weeks	Non-white, singleton, 24^{+0}–31^{+6} weeks	All singletons, 24^{+0}–31^{+6} weeks	All multiples, 24^{+0}–31^{+6} weeks
Period and sample size	2008–2010 ($n = 2995$)	2015–2019 ($n = 6092$)	2015–2019 ($n = 4661$)	2015–2019 ($n = 655$)	2015–2019 ($n = 5494$)	2015–2019 ($n = 2025$)
Ethnicity, *n* (%)						
White	2995 (100)	6092 (100)	4661 (100)	NA	4661 (84.8)	1782 (88.0)
Non-white	NA	NA	NA	655 (100)	655 (11.9)	42 (2.1)
Missing	NA	NA	NA	NA	178 (3.2)	201 (9.9)
Died, *n* (%)	244 (8.2)	433 (7.1)	403 (8.6)	75 (11.5)	490 (8.9)	152 (7.5)
Male, *n* (%)	1627 (54.3)	3486 (57.2)	2635 (56.6)	381 (58.2)	3106 (56.5)	1055 (52.1)
Gestational age, *n* (%)						
23^{+0}–23^{+6} weeks	37 (1.2)	8 (0.1)	NA	NA	NA	NA
24^{+0}–27^{+6} weeks	620 (20.7)	1311 (21.5)	1311 (28.1)	211 (32.2)	1578 (28.7)	585 (28.9)
28^{+0}–31 weeks	1575 (52.6)	3350 (55.0)	3350 (71.8)	444 (67.8)	3916 (71.3)	1440 (71.1)
32^{+0}–32^{+6} weeks	763 (25.5)	1423 (23.4)	NA	NA	NA	NA
Mean ± SD	29.5 ± 2.4	29.9 ± 2.3	29.1 ± 2.1	28.9 ± 2.1	29.1 ± 2.1	29.1 ± 2.2
Birthweight, *n* (%)						
0–499 g	8 (0.3)	31 (0.5)	30 (0.6)	4 (0.6)	34 (0.6)	8 (0.4)
500–999 g	626 (20.9)	1358 (22.2)	1343 (28.8)	212 (32.4)	1616 (29.4)	559 (27.6)
1000–1499 g	1092 (36.5)	2159 (35.4)	1913 (41.0)	285 (43.5)	2265 (41.2)	888 (43.9)
1500–1999 g	1040 (34.7)	1934 (31.7)	1235 (26.5)	142 (21.7)	1417 (25.8)	553 (27.3)
2000–2499 g	219 (7.3)	563 (9.2)	132 (2.8)	11 (1.7)	150 (2.7)	16 (0.8)
2500 + g	10 (0.3)	47 (0.8)	8 (0.2)	1 (0.2)	12 (0.2)	1 (0.05)
Mean ± SD	1389 ± 434	1391 ± 456	1258 ± 395	1189 ± 381	1250 ± 397	1253 ± 355
Apgar at 5 min						
Median (IQR)	9 (8–9)	8 (7–9)	8 (7–9)	8 (6–9)	8 (7–9)	8 (7–9)
Missing*	203 (6.8)	26 (0.4)	23 (0.5)	3 (0.4)	35 (0.6)	7 (0.3)
Caesarean section, *n* (%)						
Yes	1566 (52.3)	2740 (45.0)	2128 (45.7)	289 (44.1)	2445 (44.5)	884 (43.7)
No	1417 (47.3)	2875 (47.2)	2128 (45.7)	309 (47.2)	2473 (45.0)	1015 (50.1)
Missing*	12 (0.4)	477 (7.8)	405 (8.7)	57 (8.7)	576 (10.5)	126 (6.2)
Maternal age (years), mean ± SD	NA	30.4 ± 5.0	30.4 ± 5.1	31.0 ± 6.1	30.5 ± 5.2	30.9 ± 4.8
Primiparity, *n*	NA	3784	2895	283	3287	1343

IQR, inter quartile range; NA, not applicable; SD, standard deviation.
*Apgar score and caesarean section were not predictors in the model, thereby missing status did not influence the performance of the model. Information on maternal age and parity was not available in the development population.

gestational age and birthweight are presented separately for male and female infants in Figure 2.

Performance in other populations

The original model showed also good external validity with good calibration and discrimination in singleton populations (Table 2, Figure 1B–1D). Based on the good model performance of the original model keeping the original model was favoured in all singleton populations. However, in multiples, an update of the intercept was indicated (change of −10.28 to −9.98). The model performance of both the original and the intercept-adjusted model in the multiple population is presented in Table 2 and Figure 1(E i and E ii).

© 2021 John Wiley & Sons Ltd.

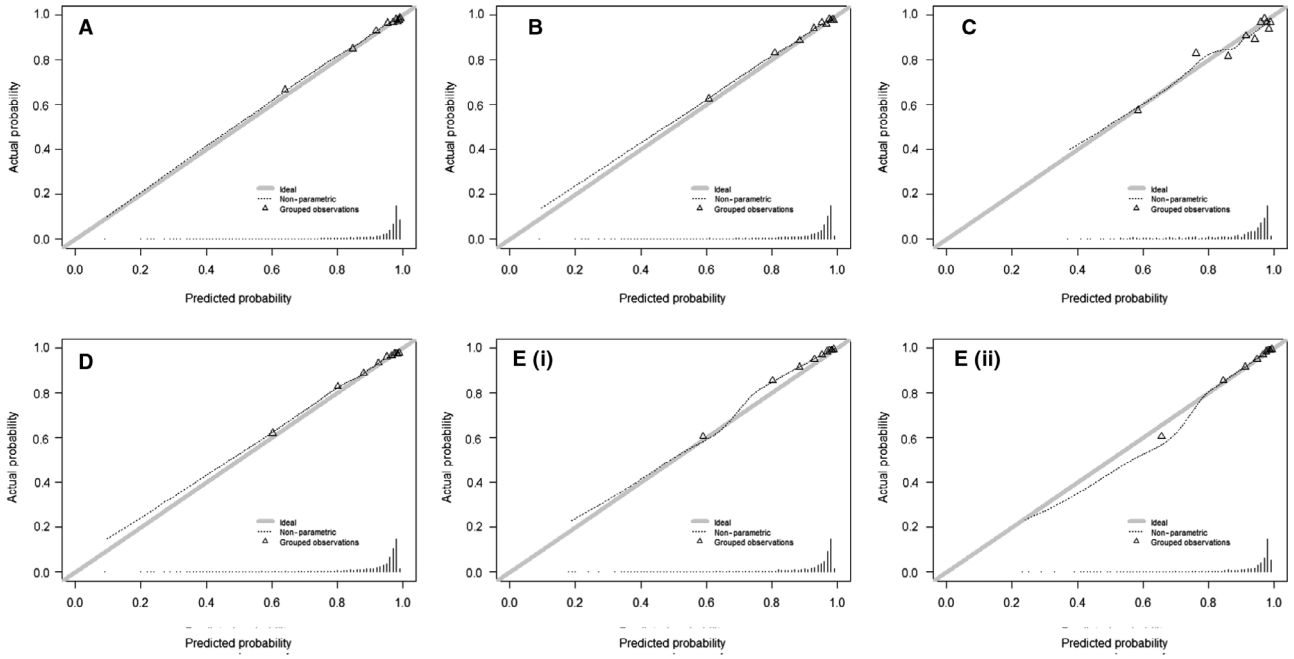

Figure 1. This figure shows the calibration plot as a result of applying the original model to (A) the validation population similar to the UK population used for development, i.e. all white, singleton infants born between 23^{+0} and 32^{+6} weeks of gestation, and to the different validation populations of infants born between 24^{+0} and 31^{+6} weeks of gestation: (B) white singletons, (C) non-white singletons, (D) all singletons, (E, i) all multiples before intercept adjustment, (E, ii) all multiples after intercept adjustment. The triangles indicate deciles of predicted risk.

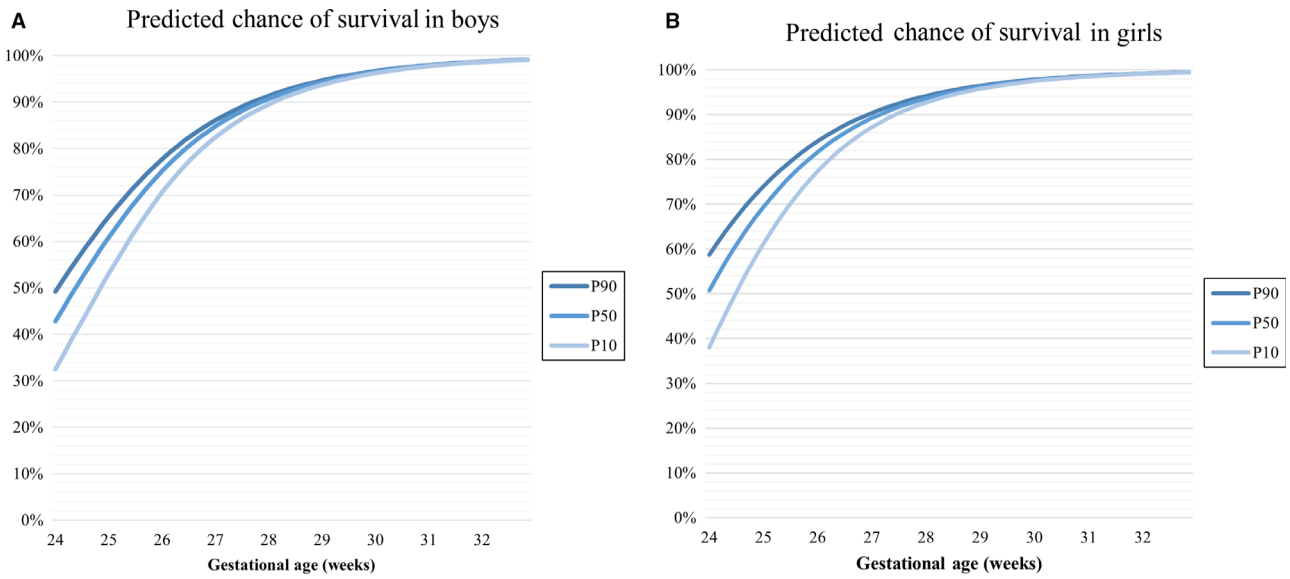

Figure 2. This figure shows predicted chance of survival according to gestational age and birthweight centile, separately for boys (A) and girls (B).

Discussion

Main findings

The aim of this study was to perform an external validation, and if needed updating, of a high-quality prognostic model to make it widely applicable in daily practice. The performance of the prognostic model predicting survival in preterm infants was assessed in an independent, nationwide perinatal registry cohort. The model was previously developed with data from the UK and was found to be valid in a large recent cohort of Dutch infants. When applying the model to the Dutch population with similar inclusion criteria to the development population by Manktelow et al.,[19] the model showed good external validity without the need

© 2021 John Wiley & Sons Ltd.

Table 2. Performance of the models

	Population similar to the development population	Other populations				
	White, singleton, 23^{+0}–32^{+6} weeks	White, singleton, 24^{+0}–31^{+6} weeks	Non-white, singleton, 24^{+0}–31^{+6} weeks	All singletons, 24^{+0}–31^{+6} weeks	All multiples, 24^{+0}–31^{+6} weeks	
					Original model	Adapted model
Infants included	6092	4661	655	5494	2025	2025
Infants survived	5659 (92.9%)	4258 (91.4%)	580 (88.5%)	5004 (91.1%)	1873 (92.5%)	1873 (92.5%)
Infants died	433 (7.1%)	403 (8.6%)	75 (11.5%)	490 (8.9%)	152 (7.5%)	152 (7.5%)
Predicted chance of survival	92.6%	90.7%	89.4%	90.5%	90.6%	92.5%
Calibration-in-the-large	0.003	0.006	–0.008	0.006	0.019	0.000
Calibration slope (95% CI)	0.92 (0.84–1.00)	0.95 (0.86–1.04)	0.82 (0.61–1.03)	0.93 (0.85–1.01)	1.16 (1.00–1.32)	1.16 (1.00–1.32)
c statistic (95% CI)	0.82 (0.79–0.84)	0.81 (0.78–0.83)	0.77 (0.71–0.83)	0.80 (0.78–0.82)	0.86 (0.83–0.89)	0.86 (0.83–0.89)
Lowest–highest decile	63.9–99.3%	60.7–98.7%	58.4–98.8%	60.7–98.8%	59.0–98.9%	65.7–99.2%
Lowest–highest predicted risk	9.6–99.5%	9.6–99.1%	37.7–99.1%	9.6–99.1%	18.8–99.1%	23.9–99.4%
Closed test procedure						
Method chosen	Keep original model	Keep original model	Keep original model	Keep original model	Model with updated intercept	
Testing recalibration-in-the-large against the original model	$\chi^2 = 10.3$, df = 5, $P = 0.067$	$\chi^2 = 5.3$, df = 4, $P = 0.26$	$\chi^2 = 4.2$, df = 4, $P = 0.37$	$\chi^2 = 7.5$, df = 4, $P = 0.11$	$\chi^2 = 19.0$, df = 4, $P < 0.001$	
Testing recalibration against recalibration-in-the-large	NA	NA	NA	NA	$\chi^2 = 7.6$, df = 3, $P = 0.056$	

95% CI, 95% confidence interval; df, degrees of freedom; NA, not applicable.

© 2021 John Wiley & Sons Ltd.

for updating, as indicated by good calibration and the ability to discriminate well between infants that survived and did not survive to discharge. The *c* statistic in the external validation population was slightly lower than the *c* statistic as found by internal validation in the development population (0.82 versus 0.86), which is to be expected when applying an existing model to new patients. The overall observed proportion of survival was similar to the mean predicted chances, indicating that overall, the model estimates the chance of survival accurately. When applying the model more broadly to very preterm infant populations, the model showed good external validity, with only in multiples a need for a minor adjustment (intercept) of the original model.

When externally validating a prediction model, researchers should evaluate and quantify the relatedness between the population of the development and validation samples. Otherwise, inferences on the actual clinical value or transportability of a prediction model may be misleading and cause prediction models to be implemented in incompatible populations.[19] Our study showed similar baseline characteristics and survival rates between the development and the validation populations, which may explain why the model performs well in Dutch infants. Although the model needed an intercept adjustment in the group of multiples, the adjustment was small and external validation indicated that the overall predictor–outcome associations (as indicated by the calibration slope) of the original model were adequate in the multiples population. Adjustment of the model intercept is common to allow for a difference in baseline risk that is not reflected in the predictors in the model.[18] The resemblance of characteristics and mortality rates combined with the accurate performance of the model in all validation populations indicates that implementation of the model in clinical practice could be considered.

Interpretation

Up-to-date external validation studies are scarce. However, it is very important to assess the performance (i.e. generalisability) of clinical prediction models in a population different from the development population. To provide a complete and accurate judgement of the performance of a model, information on calibration, ideally by providing a calibration plot, is needed. Having a model that discriminates well, such as the current model, does not directly imply clinical usefulness.[20] When a model is used to inform patients or physicians in making decisions, evaluation of calibration might be of greater importance. In fact, poor calibration may make an algorithm less clinically useful than a competitor algorithm that has a lower area under the curve but is well calibrated.[21,22] Our calibration slope had a value of 0.92, with a target value of 1. This suggests that estimated chances are slightly too extreme, i.e. too high for patients who are at high chance and too low for patients who are at low chance. However, our calibration slope was close to 1 in combination with a calibration-in-the-large of 0.003 close to 0, which implies that, after assessment of the model's impact on doctor's behaviour and/or clinical outcomes,[5] the model could be considered for application in clinical practice.

The prognostic model includes a small number of readily available characteristics at birth that are independent of healthcare system or local protocol, are familiar to caregivers and are already known to influence the outcome of the newborn. The advantage of using the prognostic model is that the characteristics are combined in a more formal way, allowing for individual mortality risk estimation. This is demonstrated in Figure 2, in which the model provides insight into the mutual relationship of the variables and shows the actual effect of increased gestational age or birthweight on the probability of survival. Nowadays, the dominant variable that predicts survival rate is gestational age, mostly reported as a function in (completed) weeks of gestation, for example in the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) calculator.[23] Our validated model shows that a 25% survival probability difference may exist depending on gender, birthweight and days within a particular week category of gestation. To illustrate this, a male growth-restricted (645 g, P10) preterm infant born at 25^{+0} weeks of gestation has a survival probability of 53%, whereas within the same week category, a female preterm infant born at 25^{+6} weeks of gestation with a normal birthweight (840 g, P50) has a survival probability of 80%. Obviously, this is of clinical importance for parents. Hence, three simple variables available at birth may account for large variation of survival probability within the same week category. Other well-known prediction models have added additional variables, such as antenatal corticosteroids administration (NICHD)[23] or variables related to physiological derangements such as blood pH, oxygen administration and blood pressure (SNAPPE-II, CRIB-II).[24,25] These more complex models might be capable of refining individual survival probability, so a comparative evaluation of these models might be of additional value. A model that is straightforward and allows adequate estimation of survival rates that are clinically useful, without relying on clinical variability, physiological variables or availability of therapies is needed. In our opinion, the high quality of the model by Manktelow et al., i.e. based on a very low risk of bias as found in our recent review, in combination with the small number of readily available characteristics, warrants future applicability and face validity of the model in daily clinical practice in the NICU.

The current prognostic value of the factors included in the model was unknown for the Dutch population.[26] In 2012, a previous UK model by Draper and Manktelow

© 2021 John Wiley & Sons Ltd.

et al. was validated in a Dutch population, born from 2000 to 2007.[4,27] However, survival of preterm infants has improved over the past decades and two guideline changes in active treatment of extremely preterm infants have been implemented in the Netherlands, in 2007 and 2010.[12,28] The current validation study provides up-to-date information and can be of additional value in decision-making in individual cases. It is important to provide parents and clinicians with useful prognostic information enabling them to make informed, well-considered decisions.[29] Although this model was developed for postnatal application, it can be a helpful tool to support decisions when counselling parents prenatally, e.g. using estimated fetal birthweight.

Strengths and limitations

Strengths of this study include our well-described, large, nationwide and heterogeneous cohort of infants from the Netherlands, standardised data collection and validation of a model with a small number of readily available characteristics. Although the model was developed in a population restricted to white, singleton infants, this validation study succeeded in generalising the model to non-white and multiple births. Our study has some limitations as well. We used data from a registry database, meaning that data were not prospectively collected with the aim of external validation of the Manktelow et al. model. Nevertheless, the characteristics included in the model are collected accurately, as indicated by the low number of missing values, and the included variables did not leave a broad scope for free interpretation. Second, the population of non-white singletons was small with respect to the number of infants that died ($n = 75$ instead of the needed 100 events), resulting in increased risk for biased estimates of model performance in this population. The multiple infant population was too small to apply the model separately in white and non-white infants. Third, as the aim of our study was to perform validation of the model of Manktelow et al., we included infants with a similar broad gestational age range. However, clinical relevance may be highest for infants; a more specific study population of infants born below 28 weeks of gestation is needed. Last, the non-white population was analysed as one group, but substantial risk differences might exist among this heterogeneous group including Asian, African and many other ethnicities. Although the model appeared to be applicable in the full group of children, no conclusions can be drawn on specific ethnicities other than white children.

Conclusions

In this external validation study, we showed that a high-quality prognostic model predicting survival in very preterm infants has good external validity in an independent, nationwide cohort. The high quality of the model by Manktelow et al. with very low risk of bias, in combination with the small number of readily available characteristics and good performance at external validation make the future applicability and face validity of the model high for use in daily clinical practice in the NICU. After assessment of the model's impact on clinician's behaviour and/or clinical outcomes, implementation of the model in clinical practice could be considered. This model can be a helpful step towards an individualised, prognosis-based approach in counselling parents on the limits of viability.

Disclosure of interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work. Completed disclosure of interests form available to view online as supporting information.

Contribution of authorship

PB and ES designed the study. PB analysed the data and wrote the first draft of the manuscript. FG, WO, PA and ES provided critical feedback and helped shape the research, analysis and manuscript. FG, WO, SK, PD, KD, FD, AH, RK, FS, EW and RW made substantial contributions to the acquisition of the data. All authors provided critical feedback on previous versions of the manuscript and approved the final version of the submitted manuscript. PA and ES supervised the project. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Details of ethics approval

No ethical approval was required. All data obtained from the registry were extracted from individual medical records and rendered anonymous, taking the European privacy policy into account. Therefore, medical ethical approval and individual informed consent for participation were not necessary.

Funding

Dr van Beek was supported by an unrestricted grant from Stichting Tiny & Anny van Doorne Fonds. The funding source had no role in the design, conduct, analyses or reporting of the study or in the decision to submit the manuscript for publication. The other authors received no external funding.

© 2021 John Wiley & Sons Ltd.

Acknowledgements

We thank Dr S.E. Seaton and Prof. B.N. Manktelow for providing the baseline characteristics of the original development study. Written permission was obtained from both to include their names in this Acknowledgements section. SES and BNM did not receive any compensation for their contributions.

The EPI-DAF study group (project leader, Peter Andriessen) and EPI-DAF PhD student (Pauline van Beek) are indebted to all collaborators, including paediatricians, psychologists and physiotherapists of all perinatal centres in the Netherlands: Amsterdam UMC location AMC (PI: Aleid van Wassenaer-Leemhuis), Amsterdam UMC location VUmc (PI: Céleste Laarman), Erasmus MC (PI: Renate Swarte), Isala Clinics (PI: Susanne Mulder-de Tollenaer), LUMC (PI: Monique Rijken), Maastricht UMC+ (PI: Elke van Westering-Kroon), Máxima MC (PI: Ellen de Kort), Radboud UMC (PI: Katerina Steiner), UMCG (PI: Henk ter Horst) and UMC Utrecht (PI: Corine Koopman-Esseboom). The EPI-DAF study group is also grateful for the support by the Netherlands Perinatal Registry (Perined; Lisa Broeders, Ger de Winter), the Dutch working group on Neonatal Follow-up (LNF; Cornelieke Aarnoudse-Moens, Monique Rijken, Renate Swarte), the Dutch National Neonatal Registry Working Group (LNR; Floris Groenendaal, René Kornelisse) and the Neonatology Network Netherlands (N3; Debbie Nuytemans, Wes Onland).

Data availability statement

Data are not available upon request. Data are obtained from a third party and are not publicly available. All data relevant to the study are included in the article and do not include patient identifiable data. ■

References

1 WHO. Preterm birth. Fact sheet: reviewed. February 2018. [http://www.who . int/news-room/fact-sheets/detail/preterm-birth]. Accessed 20 December 2019.
2 Tucker J, McGuire W. Epidemiology of preterm birth. *BMJ* 2004;329:675–8.
3 Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, Narwal R, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet* 2012;379:2162–72.
4 Schuit E, Hukkelhoven CW, Manktelow BN, Papatsonis DN, de Kleine MJ, Draper ES, et al. Prognostic models for stillbirth and neonatal death in very preterm birth: a validation study. *Pediatrics* 2012;129:120.
5 Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;04: b606.
6 Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, et al. Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009;15:537–52.
7 van Beek PE, Andriessen P, Onland W, Schuit E. Prognostic models predicting mortality in preterm infants: systematic review and meta-analysis. *Pediatrics* 2021;147:e2020020461.
8 Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratinam S, Khan KS, et al. Prognostic models in obstetrics: available, but far from applicable. *Am J Obstet Gynecol* 2016;214:79–90.e36.
9 Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;16: i2416.
10 Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;170:W1–W33.
11 Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51–8.
12 Manktelow BN, Seaton SE, Field DJ, Draper ES. Population-based estimates of in-unit survival for very preterm infants. *Pediatrics* 2013;131:425.
13 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;07: g7594.
14 Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475–83.
15 Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214–26.
16 Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016;1:167–76.
17 Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;45:562.
18 Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. *Stat Med* 2017;36:4529–39.
19 Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;68:279–89.
20 Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
21 Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Making* 2015;35:162–9.
22 Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;17(1):230–7.
23 Tyson JE, Parikh NA, Langer J, Green C, Higgins RD. Intensive care for extreme prematurity — moving beyond gestational age. *N Engl J Med* 2008;358:1672–81.

© 2021 John Wiley & Sons Ltd.

24 Parry G, Tucker J, Tarnow-Mordi W. CRIB II: an update of the clinical risk index for babies score. *Lancet* 2003;361:1789–91.
25 Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. *J Pediatr* 2001;138:92–100.
26 Geurtzen R, Draaisma J, Hermens R, Scheepers H, Woiski M, van Heijst A, et al. Perinatal practice in extreme premature delivery: variation in Dutch physicians' preferences despite guideline. *Eur J Pediatr* 2016;175:1039–46.
27 Draper ES, Manktelow B, Field DJ, James D. Prediction of survival for preterm births by weight and gestational age: retrospective population based study. *BMJ* 1999;319:1093–7.
28 de Laat MW, Wiegerinck MM, Walther FJ, Boluyt N, Mol BW, van der Post JA, et al. Practice guideline 'Perinatal management of extremely preterm delivery'. *Ned Tijdschr Geneeskd* 2010;154:A2701.
29 De Proost L, Ismaili M'hamdi H, Verhagen A, Reiss I, Geurtzen R, Verweij EJ. On the limits of viability: toward an individualized prognosis-based approach. *J Perinatol* 2020;40:1736–8.

© 2021 John Wiley & Sons Ltd.