Antenatal prediction models for short- and medium-term outcomes in preterm infants

Takafumi Ushida, MD, PhD[1,2], Yoshinori Moriyama, MD, PhD[1,3], Masahiro Nakatochi, PhD[4], Yumiko Kobayashi, BS[5], Kenji Imai, MD, PhD[1], Tomoko Nakano-Kobayashi, MD, PhD[1], Noriyuki Nakamura, MD[1], Masahiro Hayakawa, MD, PhD[6], Hiroaki Kajiyama, MD, PhD[1], Tomomi Kotani, MD, PhD[1,2], for the Neonatal Research Network of Japan

[1]Department of Obstetrics and Gynecology, Nagoya University Graduate School of Medicine, Nagoya, Japan

[2]Division of Perinatology, Center for Maternal-Neonatal Care, Nagoya University Hospital, Nagoya, Japan

[3]Department of Obstetrics and Gynecology, Fujita Health University School of Medicine, Toyoake, Japan

[4]Division of Public Health Informatics, Department of Integrative Health Science, Nagoya University Graduate School of Medicine, Nagoya, Japan

[5]Data Science Division, Data Coordinating Center, Department of Advanced Medicine, Nagoya University Hospital, Nagoya, Japan

[6]Division of Neonatology, Center for Maternal-Neonatal Care, Nagoya University Hospital, Nagoya, Japan

Corresponding author: Takafumi Ushida, M.D., Ph.D.

Department of Obstetrics and Gynecology, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, 466-8550, Japan,

Tel: +81-52-744-2261

Fax: +81-52-744-2268

E-mail: u-taka23@med.nagoya-u.ac.jp

Conflict of Interest

The authors have no potential conflicts of interest to disclose.

Funding Information

No funding was received for this study.

Abstract

Introduction: In extremely and very preterm infants, predicting individual risks for adverse outcomes antenatally is challenging but necessary for risk-stratified perinatal management and parents' participation in decision making about treatment. Our aim was to develop and validate prediction models for short-term (neonatal period) and medium-term (3 years of age) outcomes based on antenatal maternal and fetal factors alone.

Material and methods: A population-based study was conducted on 31,157 neonates weighing ≤1,500 g and born between 22 and 31 weeks of gestation registered in the Neonatal Research Network of Japan during 2006–2015. Short-term outcomes were assessed in 31,157 infants and medium-term outcomes were assessed in 13,751 infants among the 31,157 infants. The clinical data were randomly divided into training and validation datasets in a ratio of 2:1. The prediction models were developed by factors selected using stepwise logistic regression from 12 antenatal maternal and fetal factors with the training dataset. The number of factors incorporated into the model varied from 3 to 10, on the basis of each outcome. To evaluate predictive performance, the area under the receiver operating characteristic curve (AUROC) was calculated for each outcome with the validation dataset.

Results: Among short-term outcomes, AUROCs for in-hospital death, chronic lung disease, intraventricular hemorrhage (grade III or IV) and periventricular leukomalacia were 0.85 [95% confidence interval (0.83–0.86)], 0.80 (0.79–0.81), 0.78 (0.75–0.80) and 0.58 (0.55–

0.61), respectively. Among medium-term outcomes, AUROCs for cerebral palsy and developmental quotient of <70 at 3 years of age were 0.66 (0.63–0.69) and 0.72 (0.70–0.74), respectively.

Conclusions: Although the predictive performance of these models varied for each outcome, their discriminative ability for in-hospital death, chronic lung disease, and intraventricular hemorrhage (grade III or IV) was relatively good. We provided a bedside prediction tool for calculating the likelihood of various infant complications for clinical use. To develop these prediction models would be valuable in each country, and these risk assessment tools could facilitate risk-stratified perinatal management and parents' shared understanding of their offspring's subsequent risks.

Keywords

Preterm birth, neonatal outcomes, antenatal counseling, pregnancy complications, prediction model

Abbreviations

AUROC, area under the receiver operating characteristic curve; CI, confidence interval; DC twin, dichorionic twin; DM, diabetes mellitus; DQ, developmental quotient; GDM, gestational diabetes mellitus; HDP, hypertensive disorders of pregnancy; IVH, intraventricular hemorrhage; MC twin, monochorionic twin; NICU, neonatal intensive care unit; NRNJ, Neonatal Research Network of Japan; OR, odds ratio; PVL, periventricular leukomalacia.

Key Message

Risk prediction models based on multiple antenatal factors demonstrated better performance for short- and medium-term outcomes in preterm infants. Accurate antenatal prediction of

individual risks for adverse outcomes could improve parents' shared understanding of their infants' subsequent risks.

Introduction

Preterm birth remains the leading cause of neonatal mortality and morbidity worldwide, occurring in approximately 10% of all pregnancies.(1) Advances in antenatal and neonatal care management have improved survival rates and subsequent neurodevelopmental outcomes among preterm infants.(2, 3) However, extremely and very preterm infants are still particularly susceptible to adverse lifelong health consequences from neonatal period and into adulthood because of prematurity itself and various neonatal complications.(4)

Neonatal mortality and morbidity attributed to prematurity can be reduced through appropriate perinatal interventions provided to both mothers and infants.(5) Individual risks for such diseases must therefore be assessed antenatally for successful primary prevention, and perinatal management needs to be risk stratified to prevent and minimize complications. Accurate antenatal prediction of individual risks for adverse outcomes in extremely and very preterm infants is challenging but essential for both the individual infants and clinicians.(6) Accurate risk prediction may also be advantageous in resource planning and cost-effective management.(7)

The British Association of Perinatal Medicine recently updated its recommendations regarding perinatal management of extremely preterm neonates.(8) They emphasized that antenatal risk assessment should be based not only on gestational age or birth weight but also on multiple parameters such as fetal factors, maternal clinical conditions and antenatal therapeutic strategies. A well-designed antenatal prediction model was developed by Tyson *et al* in 2008 to evaluate infants' risks for mortality and profound neurodevelopmental impairment.(9) In addition, they developed a simple and innovative web-based tool to

estimate the likelihood of these outcomes. They updated the model by using recent clinical data; however, it estimated the likelihood of adverse outcomes exclusively for infants born at 22–25 gestational weeks of age.(10) Some previous prediction models are not suitable for antenatal counseling because they include postnatal data such as Apgar score, umbilical cord pH, and body temperature at NICU admission.(11-14)

Currently, very few prediction models have been based on antenatal information without postnatal physiological data.(9, 10) In addition, evidence about risk assessment of various neonatal complications (e.g., respiratory morbidity, sepsis and neurological impairment) other than mortality, or of medium- and long-term outcomes, is limited.

The aim of this study was to develop and validate prediction models for short-term (neonatal period) and medium-term (3 years of age) outcomes in preterm infants that were based on antenatal maternal and fetal factors alone. This risk assessment tool could facilitate appropriate risk-stratified management to accomplish better outcomes and improve parents' understanding of their infants' subsequent risks.

Material and Methods

The data source of this population-based study was the Neonatal Research Network of Japan (NRNJ), a large, multicenter, voluntary collaborative group for improvement of neonatal outcomes and for research on neonatal intensive care that comprises approximately 200 level II and level III neonatal intensive care units (NICUs) throughout Japan. The number of participating NICUs varied during the study period. The NRNJ maintains a database of maternal and neonatal clinical information on extremely and very preterm infants admitted in participating NICUs. Infants eligible for this study were born alive between January 2006 and December 2015 at less than 32 weeks of gestation and weighed ≤1,500 g. We excluded infants born at less than 22 weeks of gestation, infants of higher order multiple births (triplets or more), those with major congenital and/or chromosomal abnormalities, and those with

incomplete medical records (Figure 1). In this study, only women with complete data about antenatal maternal characteristics were included for the prediction models to improve data reliability. Supplementary Table 1 lists the baseline maternal and neonatal characteristics with (n = 31,157) and without complete data (n = 2,533). Information about missing data for each variable is listed in Supplementary Table 1. Informed consent was obtained from all parents at each facility. The basic management for extremely and very preterm infants including prenatal maternal care is unified throughout the participating NICUs; however, the details would be different across the NICUs. Most infants born at 22–23 gestational weeks received active treatment, and palliative care was selected for severe cases such as severe fetal growth restrictions based on the antenatal counseling. However, this information was not documented in the NRNJ database.

A total of 38,028 neonates weighing ≤1,500 g and born at 22–31 weeks' gestation were registered in the database during the study period. Of these infants, 6,871 were not eligible to participate in this study (Figure 1). Short-term outcomes of 31,157 infants were assessed during their NICU stay; of these infants, 24,888 were singletons and 6,269 were twins (2,959 monochorionic [MC] twins and 3,310 dichorionic [DC] twins). Medium-term outcomes at 3 years of age were assessed in 13,751 infants, of whom 11,038 were singletons and 2,713 were twins (1,277 MC twins and 1,436 DC twins). For the other 15,448 infants, medium-term outcomes at 3 years of age were not assessed because of transfer to another hospital or because we could not contact them. Supplementary Table 2 lists the baseline maternal and neonatal characteristics and short-term outcomes of infants whose medium-term outcomes were (n = 13,751) and were not assessed (n = 15,448).

More than 100 variables related to maternal and neonatal information were collected through a manual search at each facility and recorded in the database after identified data were anonymized. Data administrators in the NRNJ data center checked the data quality of

clinical information and asked the data abstractors at each facility to verify the correction of these data if necessary. Maternal information used in this study included maternal age, gestational age at delivery, parity, delivery mode (cesarean section or vaginal delivery), presence of diabetes mellitus or gestational diabetes mellitus, presence of hypertensive disorders of pregnancy (HDP), presence of clinical chorioamnionitis, premature rupture of membrane, antenatal corticosteroid treatment, singleton or twin and chorionicity of twins. Neonatal information included sex, birth weight and short- and medium-term outcomes. Gestational age was estimated on the basis of first-trimester ultrasonography and the date of the mother's last menstrual period.

The following short-term outcomes were evaluated at the end of each infant's stay in the NICU: (1) in-hospital death; (2) respiratory morbidity, including respiratory distress syndrome, chronic lung disease and home oxygen therapy; (3) neurological impairment, including severe intraventricular hemorrhage (IVH; grade III or IV) and periventricular leukomalacia (PVL); (4) neonatal sepsis; (5) necrotizing enterocolitis; (6) treated retinopathy of prematurity; and (7) composite adverse outcomes (in-hospital death, severe IVH or PVL).

The following medium-term outcomes were assessed when children discharged from the NICU were 3 years of age: (1) death after NICU discharge; (2) cerebral palsy; (3) development quotient (DQ) of <70; (4) home oxygen therapy; (5) visual impairment (need for glasses); (6) hearing impairment (need for hearing aids); and (7) composite adverse outcomes (death after NICU discharge, cerebral palsy or DQ of <70).

Pregnancy-related and neonatal complications in the NRNJ database were defined previously.(15) HDP was defined as blood pressure exceeding 140/90 mm Hg after 20 weeks of gestation.(16) Clinical chorioamnionitis was defined as a maternal body temperature of ≥38.0°C and at least one of the following: (1) maternal heart rate of ≥100 bpm; (2) uterine tenderness; (3) malodorous vaginal discharge; and (4) white blood cell count of ≥15,000 cells/μL.(17) Respiratory distress syndrome was diagnosed on the basis of both clinical and

radiological features. Chronic lung disease was defined as oxygen requirement at 36 weeks' postmenstrual age. IVH and PVL were diagnosed according to findings on intracranial ultrasonography or brain magnetic resonance imaging. The severity of IVH was graded 1 to 4 according to the Papile's classification.(18) Neonatal sepsis was diagnosed on the basis of clinical symptoms and blood culture results. Necrotizing enterocolitis was defined as stage II or higher of Bell's criteria.(19) Treatments for retinopathy of prematurity included laser photocoagulation, cryotherapy or anti-vascular endothelial growth factor therapy. DQ was assessed with the Kyoto Scale of Psychological Development when children were approximately 36 months of age.(20) This scale is widely used and accepted in Japan, and its accuracy was reported to be comparable with that of the Bayley Scales of Infant and Toddler Development, Third Edition.(21)

Statistical Analyses

We used the split-sample approach to develop and internally validate the risk prediction models in this study. Data from 31,157 infants were randomly assigned to either training or validation datasets in a ratio of 2:1; thus data from 20,771 infants were assigned to the training dataset and those of 10,386 infants to the validation dataset. Overall, 12 clinical factors incorporated into the prediction model were selected based on literature reviews: maternal age, gestational age, parity, delivery mode, diabetes mellitus/gestational diabetes mellitus, HDP, clinical chorioamnionitis, premature rupture of membranes, antenatal corticosteroid treatment, plurality of pregnancy (singleton or twin), chorionicity (MC or DC), infant's sex and birth weight.(9, 11, 22-24) The association of each clinical factor with each outcome was assessed with univariate logistic regression analysis. To develop prediction models for each outcome, we performed multivariate logistic regression analysis with stepwise forward selection in the training dataset; the dependent variable was each outcome status (control = 0, case = 1) and independent variables included the clinical factors. The significance level for inclusion in and exclusion from the model construction was $p < 0.05$.

To develop prediction models for each outcome, we used stepwise forward selection, starting with 12 factors. The number of factors incorporated into the model varied from 3 to 10, on the basis of each outcome. The multicollinearity between these factors was assessed before model development. All of the variance inflation factor (VIF) values were less than 5 in this study.(25) The prediction models for infants' outcomes were developed in the training dataset; then, we evaluated their predictive performance in the validation dataset by calculating the area under the receiver operating characteristic curve (AUROC) with a 95% confidence interval (CI). The probability of infants' outcomes was estimated with a formula developed by logistic regression models. The risk calculation tool was developed with Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) on the basis of the formula for each outcome (Supplementary File 1). We then established cutoff points of the probability of each outcome as the point closest to the top-left part of the plot with perfect sensitivity or specificity. We evaluated the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of each model. To evaluate the model calibration, we created a calibration plot. We performed the Hosmer-Lemeshow test, which is a measure of deviation between observed and predicted outcomes in deciles of predicted risk.

We then performed two additional analyses to evaluate the predictive performance of the models, one based on gestational age alone and the other based on clinical data of infants born between 22 and 27 weeks of gestation registered in the NRNJ during 2011–2015.

Maternal and neonatal characteristics were compared using the chi-squared test or Student's *t* test. A *p*-value of <0.05 reflected statistical significance. All statistical analyses were performed with SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) and R software version 3.5.0 (The R Foundation, Vienna, Austria; www.r-project.org). ROC analysis was performed using pROC of the R package.(26)

Ethical approval

This study was approved by the Institutional Ethics Board of Nagoya University (approval

number 2018-0026 on 9 May, 2018), and the use of this database was permitted by the Japan Neonatal Network Executive Committee.

Results

The baseline maternal and obstetric characteristics are listed in Table 1. Short- and medium-term outcomes between the training and validation datasets are listed in Table 2. No variables were significantly different between the two datasets (Table 1 and 2). Table 3 lists the predictive performance of each model and the numbers of antenatal factors (of 12) incorporated into the models. The number of factors incorporated into the model varied from 3 to 10, on the basis of each outcome. Supplementary Table 3 lists the crude odds ratios of the 12 maternal and fetal factors. Supplementary Table 4 lists the adjusted odds ratios of the incorporated factors. The crude and adjusted odds ratios for birth weight are listed in 100-g increments in birth weight. Supplementary Table 5 lists the formulas for probability of each outcome. Among short-term outcomes, AUROCs for in-hospital death, chronic lung disease, IVH (grade III or IV) and PVL were 0.85 (95% CI 0.83–0.86), 0.80 (95% CI 0.79–0.81), 0.78 (95% CI 0.75–0.80) and 0.58 (95% CI 0.55–0.61), respectively (Table 3). Among medium-term outcomes, AUROCs for cerebral palsy and DQ of <70 were 0.66 (95% CI 0.63–069) and 0.72 (95% CI 0.70–0.74), respectively. Supplementary Table 6 shows the performance of each model and the threshold of probability for each model. The threshold of probability of each outcome is calculated on the basis of the receiver operating characteristic curve. The formula for calculating the probability of each outcome is presented in Supplementary Table 5. Supplementary Figure 1 shows the calibration plot of each outcome. The risk calculation tool of infants' short- and medium-term outcomes was developed with Microsoft Excel (Supplementary File 1). The probabilities of all outcomes are calculated at once if the values of the 12 factors are entered in the formulas. Medium-term outcomes were calculated among NICU survivors; thus the probability of these outcomes is based on the assumption that an

infant survives to NICU discharge.

We evaluated the predictive performance of models if the models were based on gestational age alone. In comparison with the original models, the AUROCs of the models based on gestational age alone were either numerically similar or slightly higher (Supplementary Table 7).

We then evaluated the difference in predictive performance between the original models and the additional models that were based on clinical data of infants born between 22 and 27 weeks of gestation registered in the NRNJ during 2011–2015. We found that the additional models for most short- and medium-term outcomes (22–27 weeks of gestation) had less discriminative ability than the original models (22–31 weeks of gestation) (Supplementary Table 8).

Discussion

In this study, we sought to develop and validate models for predicting short- and medium-term outcomes of infants born at 22–31 weeks of gestation and to provide a bedside prediction tool for calculating the likelihood of various complications for clinical use. The main findings of this study were that our prediction models for several outcomes performed well and that the prediction models based on multiple factors more accurately estimated complications than did the models based on gestational age alone. Thus the multifactorial models would be of practical use for perinatal caregivers in clinical situations to optimize perinatal care and to inform parents of various risks and expected outcomes from the neonatal period to early childhood. These models can be applied for the Japanese or women in some countries with similar perinatal medical standards; however, these predictive models can be developed through our methodology in each country with its own neonatal database.

Shared decision making is vital in neonatal care, especially for extremely preterm neonates, and it should be based on accurate risk assessment in antenatal counseling.(27)

Because the causes of preterm birth are multifactorial and heterogeneous, and because the backgrounds of individual neonates vary, the British Association of Perinatal Medicine recommended risk assessment based on both gestational age and multiple factors associated with maternal and intrauterine fetal health.(8) One of the unique aspects of our prediction models is that infants' backgrounds (e.g., maternal characteristics, pregnancy complications and antenatal treatment) were incorporated into the models to improve predictive performance. Our study demonstrated better predictive performance of models for most of the short- and medium-term outcomes than that of the models based on gestational age alone.

Another unique point was that we did not include neonatal physiological data such as Apgar score, umbilical cord blood pH and base excess, which are often used for predicting neonatal mortality.(11-13) Risk-scoring systems such as the Clinical Risk Index for Babies II (CRIB II) and the Score for Neonatal Acute Physiology (SNAP) have demonstrated validity and reliability as tools to predict early neonatal mortality in extremely preterm infants.(12, 13) The CRIB II estimates the risk of mortality on the basis of five clinical factors: gestational age, birth weight, sex, body temperature at NICU admission and base excess within the first 12 h of life. Although these scoring systems have shown better prediction than those based on a single factor alone such as birth weight or gestational age, postnatal information is required for most.(28) Thus, these scoring systems cannot be applied for antenatal risk assessment.

In accordance with previous studies, adverse outcomes in extremely and very preterm infants were more accurately predicted on the basis of multiple risk factors than on the basis of gestational age alone.(9, 10) In addition, our models showed better predictive performance than did an earlier model reported in a study of neonatal mortality.(9) According to a recent systematic review, CRIB had the best discriminative ability for predicting predischarge mortality (AUROC 0.88 [95% CI 0.86–0.90]).(28) The predictive performance of our models was comparable with that of CRIB, which requires postnatal information.

However, various factors that affect the model performance must be considered in order to compare AUROCs between different prediction models. As shown in our additional analysis (Supplementary Table 8), the range of gestational ages and timeframes may significantly affect the predictive performance of the models (e.g., for in-hospital death, AUROC 0.849 vs. 0.781), even though the same methods were applied. In this study, inclusion criteria included infants born between 22 and 31 gestational weeks, which is different from previously published models (Tyson *et al:* 22–25 gestational weeks; Parry *et al:* 22–30 gestational weeks).(9, 12) In addition, rates of adverse events varied considerably according to outcome (e.g., in-hospital death 6.3% and respiratory distress syndrome 67.9%), gestational age, and study populations. These factors that affect predictive performance must be considered when the performance of different models is compared in different settings.

Of interest is that PVL could not be appropriately predicted using our models and the predictive performance was not satisfactory. This may indicate that the underlying mechanisms of PVL are complicated and the designated factors that we incorporated into our models may not be strongly associated with the pathophysiological processes of PVL. Likewise, the lower event rate may also have been associated with the poor predictive performance.

One strength of this study is that the sample size was quite large, and the most recent data for the past 10 years were used for analyses; thus our models could be adapted to actual clinical situations. In addition, we extended the range of gestational age at birth (22–31 weeks) and the scope of complications that included not only mortality but also various complications in the short- and medium-term, comparing a previous model.(9) Our study demonstrated that our multifactorial models have a greater discriminative ability in predicting mortality and morbidities than the simple models that are based on gestational age alone, and

the ability of our models to predict mortality is similar to those of CRIB II and SNAP, which require postnatal information.(28) Second, to make these models usable in clinical practice, we provided a bedside tool in the form of a Microsoft Excel file that can calculate the likelihood of various complications at once with data about antenatal maternal and fetal characteristics (Supplementary File 1). Although the accuracy of predictions about medium-term outcomes might be inferior, information about approximate probabilities of these outcomes may be helpful for the infants' parents. Future studies may provide insight into the underlying mechanism of neonatal complications by evaluating the selected factors in each prediction model.

Several limitations in this study should also be acknowledged. First, the medium-term outcomes were documented in only approximately 40% of children who were discharged from NICUs; thus selection bias may have been present. Second, the detailed reasons for not receiving antenatal corticosteroids, for the selection of delivery mode and for the decision to deliver early were unavailable. Some of the reasons for early delivery may include bleeding from placenta previa, placenta abruption, spontaneous preterm birth and iatrogenic preterm birth as a result of pregnancy-unrelated complications, such as cancer and heart disease. According to previous reports, the reasons for preterm birth may also affect subsequent offspring outcomes.(29, 30)

We used data on birth weight when we developed these models because we could not obtain ultrasound data on estimated fetal weight before delivery. In actual clinical use of these models, fetal sex and body weight must be estimated with ultrasonography, which has a potential for error. However, the error can be small and acceptable for predicting the outcomes.(31)

For this study, we selected factors among 12 antenatal factors by stepwise univariate logistic regression models to construct each prediction model for better predictive performance. However, we could not consider interaction between factors (e.g., HDP with

versus without clinical chorioamnionitis) to develop better models. According to a recent study, a prediction model that involved a machine-learning strategy showed better performance than a model obtained by a traditional statistical approach.(14) However, that model requires postnatal physiological data (e.g., Apgar score). Further research is necessary for better predictive performance with a machine-learning strategy.

Our predictive models can be used for neonates born in Japan or other countries that have similar medical standards because the standard management for extremely and very preterm neonates varies across countries in general.(32) However, the methods of developing prediction models are not complicated. Thus we believe that the development of these models in other countries using our methods with the use of their own perinatal database is valuable and can facilitate better antenatal management in clinical practice.

Conclusion

This study established and validated risk prediction models for short- and medium-term outcomes in extremely and very preterm infants. Although the predictive performance of these models varies for each outcome, their discriminative ability for in-hospital death, chronic lung disease, and intraventricular hemorrhage (grade III or IV) is relatively good, which indicates that these models will be of great use for both clinicians and parents in daily clinical situations. We also suggest estimating the likelihood of infants' adverse outcomes with the use of multiple risk factors, not gestational age or birth weight alone. The development of these prediction models in each country would be meaningful.

Acknowledgments

The authors thank Enago (www.enago.jp) for the English language review.

Tweetable Abstract

Risk prediction models based on multiple antenatal factors demonstrated better performance for short- and medium-term outcomes in extremely and very preterm infants.

Author Contributions

TU, YM and MN contributed to the concept and design of the study. MN and YK performed the statistical analyses. TU drafted the first version of the manuscript. TU, YM, KI, TNK, NN, MH, HK, and TK were involved in analyzing and interpreting the data. TU, YM, MN, and TK critically reviewed the manuscript, and all authors approved for the final version of the manuscript.

References

1. Chawanpaiboon S, Vogel JP, Moller AB, Lumbiganon P, Petzold M, Hogan D, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. The Lancet Global health. 2019;7(1):e37-e46.
2. D'Onofrio BM, Class QA, Rickert ME, Larsson H, Långström N, Lichtenstein P. Preterm birth and mortality and morbidity: a population-based quasi-experimental study. JAMA Psychiatry. 2013;70(11):1231-40.
3. Norman M, Hallberg B, Abrahamsson T, Björklund LJ, Domellöf M, Farooqi A, et al. Association Between Year of Birth and 1-Year Survival Among Extremely Preterm Infants in Sweden During 2004-2007 and 2014-2016. JAMA. 2019;321(12):1188-99.
4. Raju TNK, Pemberton VL, Saigal S, Blaisdell CJ, Moxey-Mims M, Buist S. Long-Term Healthcare Outcomes of Preterm Birth: An Executive Summary of a Conference Sponsored by the National Institutes of Health. The Journal of pediatrics. 2017;181:309-18.e1.
5. Jois RS. Understanding long-term neurodevelopmental outcomes of very and extremely preterm infants: A clinical review. Australian journal of general practice. 2019;48(1-2):26-32.
6. Lemyre B, Moore G. Counselling and management for anticipated extremely preterm birth. Paediatrics & child health. 2017;22(6):334-41.
7. Richardson DK, Zupancic JA, Escobar GJ, Ogino M, Pursley DM, Mugford M. A critical review of cost reduction in neonatal intensive care. II. Strategies for reduction. J Perinatol. 2001;21(2):121-7.
8. Mactier H, Bates SE, Johnston T, Lee-Davey C, Marlow N, Mulley K, et al. Perinatal management of extreme preterm birth before 27 weeks of gestation: a framework for practice.

424 Archives of disease in childhood Fetal and neonatal edition. 2020;105(3):232-9.
425 9. Tyson JE, Parikh NA, Langer J, Green C, Higgins RD. Intensive care for extreme
426 prematurity--moving beyond gestational age. The New England journal of medicine.
427 2008;358(16):1672-81.
428 10. Rysavy MA, Horbar JD, Bell EF, Li L, Greenberg LT, Tyson JE, et al. Assessment of an
429 Updated Neonatal Research Network Extremely Preterm Birth Outcome Model in the Vermont
430 Oxford Network. JAMA pediatrics. 2020;174(5):e196294.
431 11. Ambalavanan N, Carlo WA, Tyson JE, Langer JC, Walsh MC, Parikh NA, et al.
432 Outcome trajectories in extremely preterm infants. Pediatrics. 2012;130(1):e115-25.
433 12. Parry G, Tucker J, Tarnow-Mordi W. CRIB II: an update of the clinical risk index for
434 babies score. Lancet (London, England). 2003;361(9371):1789-91.
435 13. Zupancic JA, Richardson DK, Horbar JD, Carpenter JH, Lee SK, Escobar GJ.
436 Revalidation of the Score for Neonatal Acute Physiology in the Vermont Oxford Network.
437 Pediatrics. 2007;119(1):e156-63.
438 14. Podda M, Bacciu D, Micheli A, Bellù R, Placidi G, Gagliardi L. A machine learning
439 approach to estimating preterm infants survival: development of the Preterm Infants Survival
440 Assessment (PISA) predictor. Scientific reports. 2018;8(1):13743-.
441 15. Ushida T, Kotani T, Sadachi R, Hirakawa A, Hayakawa M, Moriyama Y, et al. Antenatal
442 Corticosteroids and Outcomes in Preterm Twins. Obstetrics and gynecology. 2020;135(6):1387-97.
443 16. Watanabe. K, Naruse. K, Tanaka. K, Metoki. H, Suzuki. Y. Outline of Definition and
444 Classification of “Pregnancy induced Hypertension (PIH)”. Hypertension Research In Pregnancy.
445 2013;1:3-4.
446 17. Lencki SG, Maciulla MB, Eglinton GS. Maternal and umbilical cord serum interleukin
447 levels in preterm labor with clinical chorioamnionitis. Am J Obstet Gynecol. 1994;170(5 Pt
448 1):1345-51.
449 18. Papile LA, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal
450 and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. The
451 Journal of pediatrics. 1978;92(4):529-34.
452 19. Bell MJ, Ternberg JL, Feigin RD, Keating JP, Marshall R, Barton L, et al. Neonatal
453 necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. Annals of surgery.
454 1978;187(1):1-7.
455 20. Koyama T, Osada H, Tsujii H, Kurita H. Utility of the Kyoto Scale of Psychological
456 Development in cognitive assessment of children with pervasive developmental disorders.
457 Psychiatry and clinical neurosciences. 2009;63(2):241-3.
458 21. Kono Y, Yonemoto N, Kusuda S, Hirano S, Iwata O, Tanaka K, et al. Developmental
459 assessment of VLBW infants at 18 months of age: A comparison study between KSPD and Bayley
460 III. Brain & development. 2016;38(4):377-85.
461 22. Sugai MK, Gilmour S, Ota E, Shibuya K. Trends in perinatal mortality and its risk

462 factors in Japan: Analysis of vital registration data, 1979-2010. Sci Rep. 2017;7:46681.
463 23. Pontello D, Ianni A, Driul L, Della Martina M, Veronese P, Chiandotto V, et al. Prenatal
464 risk factors for intraventricular hemorrhage, neonatal death and impaired psychomotor
465 development in very low birth weight infants. Minerva ginecologica. 2008;60(3):223-9.
466 24. Hatzidaki E, Giahnakis E, Maraka S, Korakaki E, Manoura A, Saitakis E, et al. Risk
467 factors for periventricular leukomalacia. Acta Obstet Gynecol Scand. 2009;88(1):110-5.
468 25. Daoud JI. Multicollinearity and Regression Analysis. Journal of Physics: Conference
469 Series. 2017;949:012009.
470 26. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-
471 source package for R and S+ to analyze and compare ROC curves. BMC bioinformatics.
472 2011;12:77.
473 27. Gee RE, Corry MP. Patient engagement and shared decision making in maternity care.
474 Obstetrics and gynecology. 2012;120(5):995-7.
475 28. McLeod JS, Menon A, Matusko N, Weiner GM, Gadepalli SK, Barks J, et al. Comparing
476 mortality risk models in VLBW and preterm infants: systematic review and meta-analysis. J
477 Perinatol. 2020;40(5):695-703.
478 29. Garite TJ, Combs CA, Maurel K, Das A, Huls K, Porreco R, et al. A multicenter
479 prospective study of neonatal outcomes at less than 32 weeks associated with indications for
480 maternal admission and delivery. Am J Obstet Gynecol. 2017;217(1):72.e1-.e9.
481 30. Morken NH, Källen K, Jacobsson B. Outcomes of preterm children according to type of
482 delivery onset: a nationwide population-based study. Paediatric and perinatal epidemiology.
483 2007;21(5):458-64.
484 31. Milner J, Arezina J. The accuracy of ultrasound estimation of fetal weight in comparison
485 to birth weight: A systematic review. Ultrasound. 2018;26(1):32-41.
486 32. Helenius K, Sjörs G, Shah PS, Modi N, Reichman B, Morisaki N, et al. Survival in Very
487 Preterm Infants: An International Comparison of 10 National Neonatal Networks. Pediatrics.
488 2017;140(6).
489
490
491
492
493
494
495
496
497
498
499

Figure Legends

Figure 1. Flow diagram of the study population. Data on 38,028 neonates born at gestational ages of 22 0/7 to 31 6/7 weeks and weighing 1,500 g or less at birth were collected from the Neonatal Research Network of Japan database from 2006 to 2015. A total of 6,871 neonates were excluded. In cases of incomplete maternal information, data were lacking with regard to maternal age (n = 661), parity (n = 248), delivery mode (n = 124), antenatal corticosteroid treatment (n = 444), gestational diabetes or diabetes mellitus (n = 343), hypertensive disorders of pregnancy (n = 215), clinical chorioamnionitis (n = 1,326), premature rupture of membranes (n = 240) or sex of infant (n = 11) or a combination of two or more of these. Among the remaining 31,157 neonates whose short-term outcomes were evaluated, 13,751 were evaluated at 3 years of age for medium-term outcomes. MC, monochorionic; DC, dichorionic.

Supplementary Figure 1. A calibration plot with distribution of the predicted probabilities for individuals with and without outcomes. Validity of clinical models in the validation dataset. Distributions of predicted probabilities are shown separately for infants with and without each short- and medium-term outcome. Dots indicate observed proportion of disease by tenths of the predicted probability with 95% confidence intervals.

Table 1. Demographic and obstetric characteristics of the study population

Variable	Training dataset (n = 20,771)	Validation dataset (n = 10,386)	*p*-value	Total (n = 31,157)
Maternal characteristics				
Maternal age (years)	31.6 ± 5.4	31.7 ± 5.4	0.21	31.7 ± 5.4
Gestational age (weeks)	27.8 ± 2.5	27.8 ± 2.4	0.66	27.8 ± 2.5
Primiparity	10,712 (51.6)	5,340 (51.4)	0.79	16,052 (51.5)
Cesarean section	15,993 (77.0)	7,988 (76.9)	0.87	23,981 (77.0)
GDM or DM	643 (3.1)	331 (3.2)	0.66	974 (3.1)
Hypertensive disorders of pregnancy	3,866 (18.6)	1,912 (18.4)	0.66	5,778 (18.5)
Clinical chorioamnionitis	4,302 (20.7)	2,187 (21.1)	0.48	6,489 (20.8)
Premature rupture of membranes	6,933 (33.4)	3,411 (33.1)	0.66	10,374 (33.3)
Antenatal corticosteroid treatment	11,416 (55.0)	5,684 (54.7)	0.70	17,100 (54.9)
Twins	4,185 (20.1)	2,084 (20.1)	0.86	6,269 (20.1)
Monochorionic twins/total twins	2,006 (47.9)	953 (45.7)	0.10	2,959 (47.2)
Year of delivery			1.00	
2006–2010	9,838 (47.4)	4,920 (47.4)		14,758 (47.4)
2011–2015	10,933 (52.6)	5,466 (52.6)		16,399 (52.6)
Neonatal characteristics				
Male	10,885 (52.4)	5,384 (51.8)	0.35	16,269 (52.2)
Birth weight (g)	975 ± 299	970 ± 299	0.22	973 ± 299

GDM, gestational diabetes mellitus; DM, diabetes mellitus. Data are presented as means ± standard deviations or n (%). *P* values were calculated using the chi-squared test or Student's *t* test.

Table 2. Short- and medium-term outcomes of the study population

Variable	Training dataset (n = 20,771)	Validation dataset (n = 10,386)	*p*-value	Total (n = 31,157)
Short-term outcomes				
In-hospital death	1,305/20,763 (6.3)	653/10,383 (6.3)	0.99	1,958/31,146 (6.3)
Respiratory distress syndrome	14,060/20,721 (67.9)	7,069/10,362 (68.2)	0.51	21,129/31,083 (68.0)
Chronic lung disease	4,986/18,858 (26.4)	2,518/9,392 (26.8)	0.51	7,504/28,250 (26.6)
Home oxygen therapy	1,357/19,229 (7.1)	696/9,592 (7.3)	0.54	2,053/28,821 (7.1)
Intraventricular hemorrhage (III or IV)	1,126/20,650 (5.5)	516/10,323 (5.0)	0.09	1,642/30,973 (5.3)
Periventricular leukomalacia	735/20,652 (3.6)	408/10,329 (4.0)	0.09	1,143/30,981 (3.7)
Sepsis	1,939/20,684 (9.4)	961/10,342 (9.3)	0.81	2,900/31,026 (9.3)
Necrotizing enterocolitis	415/20,716 (2.0)	197/10,354 (1.9)	0.55	612/31,070 (2.0)
Treated retinopathy of prematurity	3,488/19,453 (17.9)	1,769/9,751 (18.1)	0.66	5,257/29,204 (18.0)
Composite adverse outcomes	2,618/20,763 (12.6)	1,312/10,383 (12.6)	0.95	3,930/31,146 (12.6)
Medium-term outcomes	(n = 9,159)	(n = 4,592)		(n = 13,751)
Death after NICU discharge	80/9,159 (0.9)	37/4,592 (0.8)	0.68	117/13,751 (0.9)
Cerebral palsy	749/8,792 (8.5)	375/4,413 (8.5)	0.97	1,124/13,205 (8.5)
Developmental quotient of <70	1,177/6,863 (17.1)	592/3,430 (17.3)	0.89	1,769/10,293 (17.2)
Home oxygen therapy	184/8,486 (2.2)	94/4,249 (2.2)	0.87	278/12,735 (2.2)
Visual impairment	270/5,598 (4.8)	136/2,794 (4.9)	0.93	406/8,392 (4.8)
Hearing impairment	59/6,434 (0.9)	32/3,217 (1.0)	0.71	91/9,651 (0.9)
Composite adverse outcomes	1,742/9,120 (19.1)	886/4,567 (19.4)	0.68	2,628/13,687 (19.2)

Short-term composite adverse outcomes: in-hospital death, intraventricular hemorrhage (III or IV) and/or periventricular leukomalacia; medium-term composite adverse outcomes: death after discharge from the neonatal intensive care unit (NICU), cerebral palsy and/or developmental quotient of <70. Data are presented as n (%).

Table 3. Predictive performance for each outcome

	AUROC (95% CI)	Number of factors
Short-term outcomes		
In-hospital death	0.85 (0.83–0.86)	7
Respiratory distress syndrome	0.71 (0.70–0.72)	10
Chronic lung disease	0.80 (0.79–0.81)	8
Home oxygen therapy	0.81 (0.80–0.83)	8
Intraventricular hemorrhage (III or IV)	0.78 (0.75–0.80)	6
Periventricular leukomalacia	0.58 (0.55–0.61)	5
Sepsis	0.72 (0.71–0.74)	7
Necrotizing enterocolitis	0.78 (0.76–0.81)	4
Treated retinopathy of prematurity	0.79 (0.77–0.80)	5
Composite adverse outcomes	0.74 (0.73–0.76)	8
Medium-term outcomes		
Death after NICU discharge	0.71 (0.64–0.79)	4
Cerebral palsy	0.66 (0.63–0.69)	7
Developmental quotient of <70	0.72 (0.70–0.74)	5
Home oxygen therapy	0.78 (0.73–0.82)	4
Visual impairment	0.71 (0.67–0.75)	3
Hearing impairment	0.62 (0.52–0.71)	3
Composite adverse outcomes	0.70 (0.68–0.72)	7

The numbers of antenatal factors (of 12) incorporated into the models are listed. Short-term composite adverse outcomes: in-hospital death, intraventricular hemorrhage (III or IV) and/or periventricular leukomalacia; medium-term composite adverse outcomes: death after discharge from the neonatal intensive care unit (NICU), cerebral palsy and/or developmental quotient of <70. AUROC, area under the receiver operating characteristic curve; CI, confidence interval.