Title

Development and validation of a simplified score to predict neonatal mortality risk among neonates weighing 2000 g or less (NMR-2000): an analysis using data from the UK and The Gambia

Authors

Medvedev, Melissa M
Brotherton, Helen
Gai, Abdou
et al.

Development and validation of a simplified score to predict neonatal mortality risk among neonates weighing 2000 g or less (NMR-2000): an analysis using data from the UK and The Gambia

Melissa M Medvedev, Helen Brotherton, Abdou Gai, Cally Tann, Christopher Gale, Peter Waiswa, Diana Elbourne, Joy E Lawn, Elizabeth Allen

Summary

Background 78% of neonatal deaths occur in sub-Saharan Africa and southern Asia, among which, more than 80% are in low birthweight babies. Existing neonatal mortality risk scores have primarily been developed for high-resource settings. The aim of this study was to develop and validate a score that is practicable for low-income and middle-income countries to predict in-hospital mortality among neonates born weighing 2000 g or less using datasets from the UK and The Gambia.

Methods This analysis used retrospective data held in the UK National Neonatal Research Database from 187 neonatal units, and data from the Edward Francis Small Teaching Hospital (EFSTH), Banjul, The Gambia. In the UK dataset, neonates were excluded if birthweight was more than 2000 g; if the neonate was admitted aged more than 6 h or following discharge; if the neonate was stillborn; if the neonate died in delivery room; or if they were moribund on admission. The Gambian dataset included all neonates weighing less than 2000 g who were admitted between May 1, 2018, and Sept 30, 2019, who were screened for but not enrolled in the Early Kangaroo Mother Care Trial. 18 studies were reviewed to generate a list of 84 potential parameters. We derived a model to score in-hospital neonatal mortality risk using data from 55 029 admissions to a random sample of neonatal units in England and Wales from Jan 1, 2010, to Dec 31, 2016. All candidate variables were included in a complete multivariable model, which was progressively simplified using reverse stepwise selection. We validated the new score (NMR-2000) on 40 329 admissions to the remaining units between the same dates and 14 818 admissions to all units from Jan 1, to Dec 31, 2017. We also validated the score on 550 neonates admitted to the EFSTH in The Gambia.

Findings 18 candidate variables were selected for inclusion in the modelling process. The final model included three parameters: birthweight, admission oxygen saturation, and highest level of respiratory support within 24 h of birth. NMR-2000 had very good discrimination and goodness-of-fit across the UK samples, with a c-index of 0·8859–0·8930 and a Brier score of 0·0232–0·0271. Among Gambian neonates, the model had a c-index of 0·8170 and a Brier score of 0·1688. Predictive ability of the simplified integer score was similar to the model using regression coefficients, with c-indices of 0·8903 in the UK full validation sample and 0·8082 in the Gambian validation sample.

Interpretation NMR-2000 is a validated mortality risk score for hospitalised neonates weighing 2000 g or less in settings where pulse oximetry is available. The score is accurate and simplified for bedside use. NMR-2000 requires further validation using a larger dataset from low-income and middle-income countries but has the potential to improve individual and population-level neonatal care resource allocation.

Funding Bill & Melinda Gates Foundation; Eunice Kennedy Shriver National Institute of Child Health & Human Development; Wellcome Trust; and Joint Global Health Trials scheme of Department of Health and Social Care, Department for International Development, Medical Research Council, and Wellcome Trust.

Introduction

An estimated 2·5 million neonatal deaths occurred in 2018, accounting for 47% of deaths among children younger than 5 years.[1] The burden of neonatal mortality is unequally distributed, with nearly 80% of these deaths occurring in sub-Saharan Africa and southern Asia.[1] Between 2000 and 2015, neonatal mortality declined more slowly than mortality among children aged 1–59 months. This disparity was particularly notable in sub-Saharan Africa, where the annual mortality reduction for newborn babies was less than half of that for 1–59 month-olds.[2] Slow progress in this region might be related to high incidences of preterm birth (<37 completed weeks of gestation) and low birthweight (≤2000 g),[3,4] poor access to care for neonates,[5,6] and health system capacity issues, including

Correspondence to:
Dr Melissa M Medvedev,
Department of Paediatrics,
University of California
San Francisco, San Francisco,
CA 94158, USA
melissa.morgan@lshtm.ac.uk

Research in context

Evidence before this study

2·5 million neonatal deaths occur each year, among which 78% are in sub-Saharan Africa and southern Asia. More than 80% of these deaths occur in babies with a low birthweight who are small because they are preterm, small for their gestational age, or both. There has been slow progress in reducing neonatal mortality, which accounts for nearly half of deaths in children younger than 5 years, highlighting the need for scale-up of effective interventions for neonates who are at risk. We searched PubMed for studies published between Jan 1, 1992, and July 31, 2019, with the search terms "infant, newborn", and "infant mortality", or "infant, newborn, diseases, and mortality", or "infant, premature, diseases, and mortality", or "hospital mortality", and "severity of illness index", or "risk assessment", or "predictive value of tests", or "outcome assessment". Multiple risk scores for neonatal mortality, illness severity, and clinical instability have been developed for intensive care settings. Most of these risk scores are not feasible for low-income and middle-income countries, because they rely on laboratory-derived and therapy-derived parameters that are frequently unavailable, or on clinical observations that are not reliably measurable. A need remains for a highly predictive tool, feasible for use in resource-constrained settings, to help providers objectively assess mortality risk in the most vulnerable babies.

Added value of this study

Using population-wide data from 110 176 neonates admitted to 187 hospitals across England and Wales, this study has derived and validated a mortality risk score for neonates weighing 2000 g or less (NMR-2000). To our knowledge, this is the largest dataset used to develop and validate a neonatal mortality risk score. NMR-2000 uses data on three parameters: birthweight, oxygen saturation (peripheral

capillary oxygen) at admission, and highest level of respiratory support at any point within 24 h of birth. The model had very good discrimination and goodness-of-fit across the development and UK validation samples, with a c-index of 0·8859–0·8930 and a Brier score of 0·0232–0·0271. The simplified integer score, which can be measured and calculated at the bedside, showed predictive ability similar to the model using regression coefficients. In the Gambian dataset, which included 550 neonates at one hospital, the model had good discrimination and overall goodness-of-fit, with a c-index of 0·8170 and a Brier score of 0·1688. The simplified integer score showed similar performance, with a c-index of 0·8082. Complete data for scoring were available for 83% of neonates. These findings indicate that the NMR-2000 is valid for use in health facilities where pulse oximetry is available and underscore the fact that implementation in low-income and middle-income countries would require sensitisation regarding documentation of the three parameters used in the model.

Implications of all the available evidence

To reduce neonatal mortality worldwide, there is an urgent need to scale-up evidence-based interventions targeting the major causes of death. Our risk score could expedite recognition of severe illness and enable targeted delivery of care to small and vulnerable neonates, increasing effectiveness and efficiency of facility-based neonatal care in low-income and middle-income countries. Further research is required to validate NMR-2000 in low-resource settings using a larger sample, and to evaluate its usefulness for clinical decision making. The score has the potential to inform resource use, including nursing workload.

shortes of skilled providers, essential supplies, and basic equipment.[6–8]

More than 80% of neonatal deaths in sub-Saharan Africa and southern Asia occur in babies with a low birthweight.[9] Low birthweight can result from being preterm, being small for their gestational age, or both. Mortality is twice as high in full-term neonates who are small for their gestational age than in full-term neonates who are of average size, and 15 times higher in preterm neonates who are small for their gestational age than in babies with either characteristic alone.[10] The lower the birthweight and gestational age, the higher the mortality risk.[9] Around 86% of neonates born at fewer than 28 weeks' gestation, and 41% of those born at 28–31 weeks, will die without access to intensive care;[11] more than 75% of neonates in sub-Saharan Africa and southern Asia have no access to such care.[7] Estimates suggest that neonatal special care,[5,7] including resuscitation, kangaroo mother care, feeding support or intravenous fluids, and management of respiratory distress, infections, and

jaundice, could prevent 70% of preterm deaths and decrease prematurity-related causes of neonatal mortality by 58%.[12]

Various systems for scoring illness severity and mortality risk in neonates have been developed, primarily for high-income settings (appendix p 6). Therapy-based approaches, such as the Neonatal Therapeutic Intervention Scoring System (NTISS),[13] categorise illness severity by the quantity and type of therapies administered. By contrast, the Score for Neonatal Acute Physiology (SNAP),[14,15] the Transport Risk Index of Physiologic Stability (TRIPS),[16] and other physiology-based approaches use objective, measurable parameters that vary with illness severity, such as blood pressure. Related models, such as the Clinical Risk Index for Babies (CRIB),[17,18] combine physiological parameters with perinatal factors, such as birthweight, to provide an overall mortality risk score. SNAP and CRIB are the most widely used systems and have been extensively validated.[19]

Notably, none of the aforementioned systems are practicable for routine use in low-income and middle-income countries (LMICs), because these systems rely on laboratory-derived and therapy-derived measures that are often not available, or on clinical observations that are not reliably measurable in these settings.[20–22] The simplified age-weight-sex (SAWS) score is the only validated neonatal mortality score designed for low-resource settings. Among a derivation cohort of 428 neonates weighing 1500 g or less in Bangladesh and Egypt, the SAWS was reported to have moderate discrimination for in-hospital mortality.[22] To improve the quality of facility-based neonatal care in LMICs, a highly predictive tool, which is feasible for routine use, is needed to help providers objectively assess mortality risk in small babies.

This study has two parts: (1) model development using data from the UK, and (2) model validation using data from the UK and The Gambia. The objectives were to evaluate existing neonatal illness severity and mortality risk scores to select candidate variables for use in the new model; develop and validate a score feasible for use in LMICs to predict in-hospital neonatal mortality risk among neonates weighing 2000 g or less within 24 h of birth; and compare the performance of the novel score (NMR-2000) with that of an existing score (CRIB-II).

Methods

Study design and participants
This retrospective study used data held in the UK National Neonatal Research Database (NNRD) from 187 neonatal units to develop a model for scoring in-hospital neonatal mortality risk in LMICs. The NNRD holds de-identified patient-level data, recorded by health-care providers as part of routine care, from admissions to National Health Service neonatal units in England starting from 2008, and in Wales and Scotland starting from 2012. This study included neonates admitted to units in England and Wales between Jan 1, 2010, and Dec 31, 2017 (appendix p 1). The following exclusion criteria were also applied: birthweight more than 2000 g; being admitted at older than 6 h or following discharge home; neonates who were stillborn; neonates who died in the delivery room; neonates who were moribund (received only comfort care before death; appendix p 1).

As well as data from the NNRD, we used data on neonates in The Gambia. West and central Africa have the highest neonatal mortality worldwide (31 in 1000 livebirths).[1] In The Gambia in 2018, the neonatal mortality (26 in 1000 livebirths) ranked ninth among the 16 countries of west Africa.[1] An estimated 12% of Gambian neonates are born preterm.[3] Edward Francis Small Teaching Hospital (EFSTH) in Banjul is the national referral hospital where the neonatal unit admits around 1400 neonates annually. From 2010 to 2013, case-fatality was 35% overall and prematurity-related complications were the leading cause of death.[23]

The Gambian cohort included all neonates weighing less than 2000 g who were admitted to EFSTH between May 1, 2018, and Sept 30, 2019, who were screened for but not enrolled in the Early KMC (eKMC) trial (NCT03555981). Some routine data, including mode of delivery and treatments administered, were collected from medical charts by trained study personnel. Other data collected as part of the screening process were exported from the trial database, including birthweight, sex, birth plurality, referral status, and peripheral capillary oxygen saturation (SpO$_2$; appendix p 1).

Model development and validation using the UK dataset was approved by the North West–Preston Research Ethics Committee (17/NW/0709), the UK Health Research Authority, and the London School of Hygiene and Tropical Medicine (LSHTM; reference number 14594). Letters were sent to the UK Neonatal Collaborative Lead of all units contributing data to the NNRD, providing information about the study and giving each an opportunity to opt out. Model validation using the Gambian dataset was approved by research ethics committees of the Gambian Government and Medical Research Council Unit The Gambia at the LSHTM (reference number 1643) and LSHTM (reference number 16189). Consent was not obtained, as this was a retrospective study using de-identified data.

Selection of candidate variables
To select the candidate variables for the model, 18 studies describing existing systems for assessing neonatal mortality risk and illness severity were reviewed to generate a list of potential parameters (appendix p 6). Parameters that are typically unavailable, infrequently obtained, or unreliably measured in low-resource settings were excluded (appendix p 1). Remaining parameters were evaluated using the following exclusion criteria: low prevalence in the NNRD (<0·1%); high proportion of missing data in the development dataset (≥20%); not predictive of mortality in neonates who are preterm or have a low birthweight; low prevalence within the first 24 h of life; little evidence to support validity; and concept better represented by an alternative variable (appendix pp 7–9).

Model development
To create the model, we used a development sample of neonates admitted to a random sample of neonatal units in England and Wales from Jan 1, 2010, to Dec 31, 2016. Logistic regression models were derived to model in-hospital mortality risk. Robust standard errors allowed for clustering within units. All candidate variables were included in a complete multivariable model, which was progressively simplified using reverse stepwise selection, with the least statistically significant variable removed at each step. Discrimination was assessed with the c-index, equivalent to the area under the receiver operating characteristic (ROC) curve. A value of 0·5 indicates no predictive ability, 0·8 is considered good, and 1 is perfect.[21]
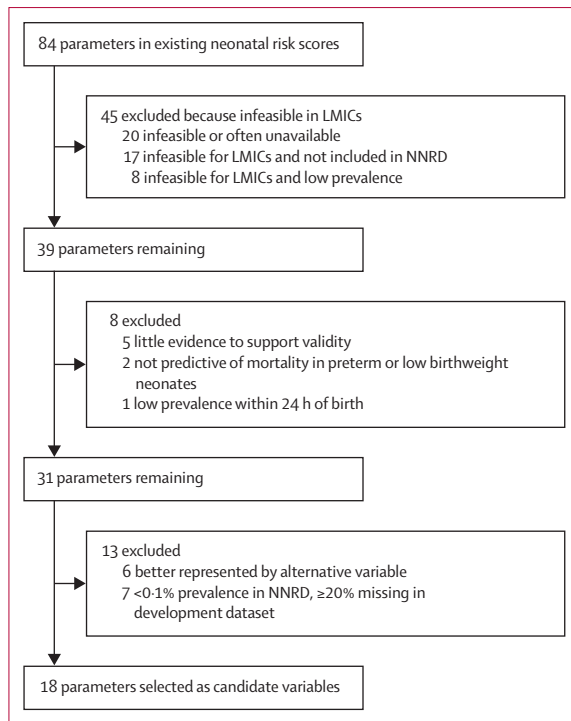
Figure 1: Flow chart showing filtration of parameters from existing risk scores to select candidate variables
LMICs=low-income and middle-income countries. NNRD=National Neonatal Research Database.

Overall goodness-of-fit was assessed with the Brier score and calibration using plots of observed versus predicted risk (appendix p 2). Multiple imputation with chained equations was used to assess the effect of missing data (appendix p 2). The logistic regression model was executed across the imputed datasets, and the resulting β coefficients and c-index were compared with original estimates. A sensitivity analysis excluding neonates whose admission age was unknown (anonymised data derived from calculated difference between birth time and admission time) was done to reassess model performance, because admission at more than 6 h of age was an exclusion criterion. Performance was additionally reassessed following exclusion of neonates who were transferred for any reason because outcome data were not available for these babies. Performance for predicting mortality within 24 h of birth was evaluated in a secondary analysis, because 36% of neonatal deaths occur within this timeframe.[24]

Score development

To develop the score, we assigned the parameters in the final model points proportional to their β regression coefficient values. Whole numbers were used to generate an easily calculable score. We arbitrarily defined low-risk, medium-risk, and high-risk groups (appendix pp 2–3). To assess the calibration of the score to the model using regression coefficients, observed risks in groups and population deciles of scores were derived and compared with mean predicted risks in each group or population decile. We assessed overall predictive ability of the score using the c-index.

Model validation

We then evaluated both the internal and external validity of the model. Internal validity is the reproducibility of a prediction model for the underlying population from which the data originated.[25] Bootstrap resampling with 1000 samples from within the development sample was used to internally validate the model, estimating optimism-adjusted measures of discrimination and goodness-of-fit in each bootstrap sample (appendix p 3). Performance of the refitted model in each bootstrap sample was compared with that of the refitted model in the original development

---

Panel: Candidate variables evaluated in the modelling process

Clinical signs and observations
- Heart rate at admission
- Respiratory rate at admission
- Temperature at admission
- Oxygen saturation (SpO₂) at admission
- Convulsions within 24 h of birth, defined as the presence of any clinical or electrographic seizures
- Clinically relevant increase in apnoea or brachycardia episodes, oxygen requirement, ventilatory support, or respiratory rate within 24 h of birth*

Therapy-based variables
- Bag-mask resuscitation at delivery
- Intravenous fluids within 24 h of birth
- Antibiotic therapy within 24 h of birth
- Oxygen therapy within 24 h of birth†
- Highest level of respiratory support administered at any point within 24 h of birth‡
- Caffeine (or aminophylline) within 24 h of birth
- Anticonvulsant therapy within 24 h of birth

Perinatal factors
- Sex
- Birthweight
- Gestational age
- Small for gestational age§
- Presence of visually recognisable anomaly at birth¶

SpO₂=peripheral capillary oxygen saturation. FiO₂=fraction of inspired oxygen. *Defined as an increase that was clinically significant enough to necessitate obtaining a culture to evaluate for suspected sepsis, at any point within 24 h of birth. †Defined as delivery of supplemental oxygen (FiO₂ >0·21) via any method at any point within 24 h of birth. ‡Not including initial resuscitation at birth; level 1 defined as nasal cannula or headbox; level 2 defined as continuous positive airway pressure, bilevel or synchronised intermittent positive airway pressure, or invasive ventilation with an endotracheal tube or tracheostomy. §Defined as birthweight less than the 5th percentile for gestational age, using UK-WHO standards.[26] ¶Defined as the presence of one or more of the following: cleft lip or palate; microcephaly; trisomy 13, trisomy 18, or trisomy 21; spina bifida, myelomeningocele, or meningocele; encephalocele; anencephaly; holoprosencephaly or prosencephaly; ambiguous genitalia; hypospadias; absent anus; gastroschisis; exomphalos or omphalocele; achondroplasia; Noonan syndrome.

	Development sample (55 029 eligible neonates; 112 neonatal units)	External validation samples		
		Random (40 329 eligible neonates; 75 neonatal units)	Temporal (14 818 eligible neonates; 167 neonatal units)	Full (55 147 eligible neonates; 173 neonatal units)
Birthweight				
Extremely low birthweight (<1000 g)	7518 (13·7%)	5705 (14·2%)	2238 (15·1%)	7943 (14·4%)
Very low birthweight (1000–1499 g)	15 475 (28·1%)	11 290 (28·0%)	4198 (28·3%)	15 488 (28·1%)
Low birthweight (1500–2000 g)	32 021 (58·2%)	23 324 (57·9%)	8381 (56·6%)	31 705 (57·5%)
Birthweight data missing*	15 (0·03%)	10 (0·02%)	1 (<0·01%)	11 (0·02%)
Gestational age (weeks)				
Extremely preterm (<28)	6969 (12·7%)	5203 (12·9%)	1990 (13·4%)	7193 (13·1%)
Very preterm (28–31)	17 810 (32·4%)	13 108 (32·5%)	4963 (33·5%)	18 071 (32·8%)
Moderate-late preterm (32–36)	28 241 (51·3%)	20 604 (51·1%)	7470 (50·4%)	28 074 (50·9%)
Full term (37–42)	1996 (3·6%)	1408 (3·5%)	393 (2·7%)	1801 (3·3%)
Gestational age data missing	13 (0·02%)	6 (0·02%)	2 (0·01%)	8 (0·01%)
Size at gestation				
Small for gestational age	11 039 (20·1%)	7965 (19·8%)	2816 (19·0%)	10 781 (19·6%)
Size at gestation data missing	16 (0·03%)	10 (0·03%)	2 (0·01%)	12 (0·02%)
Sex				
Male	27 361 (49·9%)	20 307 (50·4%)	7490 (50·6%)	27 797 (50·4%)
Sex data missing	72 (0·1%)	30 (0·07%)	18 (0·1%)	48 (0·09%)
Mode of delivery				
Spontaneous vaginal	16 361 (32·3%)	12 404 (32·6%)	4227 (30·5%)	16 631 (32·0%)
Caesarean section	32 473 (64·1%)	24 404 (64·0%)	9148 (66·1%)	33 552 (64·6%)
Assisted vaginal	1820 (3·6%)	1284 (3·4%)	463 (3·3%)	1747 (3·4%)
Mode of delivery data missing	4375 (8·0%)	2237 (5·5%)	980 (6·6%)	3217 (5·8%)
Multiple birth				
Yes	16 933 (30·8%)	12 056 (29·9%)	4442 (30·0%)	16 498 (29·9%)
Multiple birth data missing	22 (0·04%)	8 (0·02%)	2 (0·01%)	10 (0·02%)
Location of birth				
Inborn†	53 954 (98·1%)	39 481 (98·0%)	14 476 (97·9%)	53 957 (98·0%)
Location of birth data missing	7 (0·01%)	36 (0·09%)	33 (0·2%)	69 (0·1%)
Location of care				
Neonatal intensive care unit‡	24 018 (43·7%)	22 362 (55·3%)	7506 (50·7%)	29 840 (54·1%)
Local neonatal unit§	26 276 (47·8%)	13 541 (33·5%)	6054 (40·9%)	19 541 (35·4%)
Special care baby unit¶	4730 (8·6%)	4538 (11·2%)	1258 (8·5%)	5766 (10·5%)
Location of care data missing	5 (0·01%)	0 (0·0%)	0 (0·0%)	0 (0·0%)
Age at admission (min)				
Median (IQR)	21 (13–33)	21 (13–34)	23 (15–35)	22 (14–34)
Age at admission data missing	5 (0·01%)	0 (0·0%)	0 (0·0%)	0 (0·0%)
Disposition				
Died before discharge	1653 (3·0%)	1306 (3·2%)	395 (2·8%)	1701 (3·1%)
Extremely low birthweight (<1000 g)	1159 (15·4%)	929 (16·3%)	280 (12·5%)	1209 (15·2%)
Very low birthweight (1000–1499 g)	295 (1·9%)	228 (2·0%)	67 (1·6%)	295 (1·9%)
Low birthweight (1500–2000 g)	199 (0·6%)	149 (0·6%)	48 (0·6%)	197 (0·6%)
Died within 24 h of birth	207 (0·4%)	194 (0·5%)	50 (0·3%)	244 (0·4%)
Transferred to another care unit‖	12 793 (23·3%)	10 119 (25·1%)	4268 (30·3%)	14 387 (26·5%)
Disposition data missing	73 (0·1%)	32 (0·1%)	726 (4·9%)	758 (1·4%)
Age at discharge (days)				
Median (IQR)	22 (12–38)	21 (11–36)	19 (10–34)	20 (11–36)
Age at discharge data missing	21 (0·04%)	13 (0·03%)	740 (5·0%)	754 (1·4%)
				(Table 1 continues on next page)

	Development sample (55 029 eligible neonates; 112 neonatal units)	External validation samples		
		Random (40 329 eligible neonates; 75 neonatal units)	Temporal (14 818 eligible neonates; 167 neonatal units)	Full (55 147 eligible neonates; 173 neonatal units)
(Continued from previous page)				
Variables collected at time of birth				
Bag-mask resuscitation at delivery	24 302 (44·2%)	18 297 (45·4%)	6131 (41·4%)	24 428 (44·3%)
Visually recognisable anomaly	1299 (2·4%)	1151 (2·9%)	309 (2·1%)	1460 (2·7%)
Variables collected at time of admission				
Heart rate (beats per min), mean (SD)	153·4 (18·4)	153·8 (18·7)	154·5 (18·6)	154·0 (18·6)
Heart rate data missing	7197 (13·1%)	3557 (8·8%)	1234 (8·3%)	4791 (8·7%)
Respiratory rate (breaths per min), mean (SD)	53·3 (29·8)	52·4 (12·9)	52·5 (13·2)	52·4 (13·0)
Respiratory rate data missing	9535 (17·3%)	5377 (13·3%)	2008 (13·6%)	7385 (13·4%)
Temperature (°C), mean (SD)	36·6 (0·7)	36·6 (0·7)	36·7 (0·8)	36·6 (0·7)
Temperature data missing	589 (1·1%)	371 (0·9%)	166 (1·1%)	537 (1·0%)
SpO_2 (%), median (IQR)	96 (93–98)	96 (92–99)	96 (93–99)	96 (93–99)
SpO_2 data missing	7787 (14·2%)	4213 (10·4%)	1392 (9·4%)	5605 (10·2%)
Variables collected within 24 h of birth				
Increased apnoea or bradycardia, oxygen, ventilatory support, or respiratory rate	3005 (5·5%)	2458 (6·1%)	903 (6·1%)	3361 (6·1%)
Convulsions	134 (0·3%)	81 (0·2%)	24 (0·2%)	105 (0·2%)
Convulsions data missing	579 (1·1%)	456 (1·1%)	297 (2·0%)	753 (1·4%)
Oxygen therapy	13 998 (25·4%)	10 989 (27·3%)	5178 (34·9%)	16 167 (29·3%)
Highest level of respiratory support				
Nasal cannula or headbox	3722 (7·0%)	2758 (7·0%)	2035 (13·9%)	4793 (8·9%)
CPAP, BiPAP or SiPAP, or invasive ventilation	23 374 (43·8%)	16 676 (42·6%)	6427 (43·8%)	23 103 (42·9%)
Respiratory support data missing	1658 (3·0%)	1161 (2·9%)	150 (1·0%)	1311 (2·4%)
Other interventions				
Intravenous fluids	41 506 (75·4%)	30 468 (75·6%)	11 697 (78·9%)	42 165 (76·5%)
Antibiotic therapy	39 774 (72·3%)	29 877 (74·1%)	11 152 (75·3%)	41 029 (74·4%)
Caffeine citrate	14 276 (25·9%)	10 862 (26·9%)	5438 (36·7%)	16 300 (29·6%)
Anticonvulsant therapy	162 (0·3%)	150 (0·4%)	45 (0·3%)	195 (0·4%)

Data are complete except where missing data are detailed; missing data are the total number of neonates for whom data are not available. Data are n (%) except where otherwise indicated. See panel for definitions of variables. CPAP=continuous positive airway pressure. BiPAP=bilevel positive airway pressure. SiPAP=synchronised intermittent positive airway pressure. SpO_2=peripheral capillary oxygen saturation. *For neonates whose birthweight was missing, admission weight was used to determine eligibility. †Inborn is defined as birth at the hospital of neonatal unit admission. ‡Intensive care units provide care for the sickest neonates who require constant supervision and monitoring, including those born at fewer than 27 weeks' gestational age: care typically includes mechanical ventilation; surgery services offered in some units; care is analogous to American Academy of Paediatrics levels 3 and 4.[27] §Local neonatal units provide full care for the majority of babies more than 27 weeks' gestational age, including short periods of intensive care; therapies provided include continuous monitoring, CPAP, and parenteral nutrition. ¶Special care units provide care for all other babies who could not reasonably be cared for at home; therapies provided include cardiorespiratory monitoring, nasogastric feeding, supplemental oxygen, and phototherapy. ||Transfer to another care unit from the initial unit of neonatal admission.

Table 1: Characteristics of the participants in the data samples from the UK National Neonatal Research Database

sample; estimates of optimism were averaged and subtracted to provide optimism-adjusted measures.

External validity is the generalisability of a model's performance to related populations.[25] The model was evaluated in three external validation samples: the random sample, which included neonates admitted to the units withheld from the development sample; the temporal sample, which included neonates admitted to units in England and Wales from Jan 1, to Dec 31, 2017; and the Gambian sample, which included neonates admitted to EFSTH between May 1, 2018, and Sept 30, 2019. Each sample was used to assess distinctive features of model

performance. The random sample tested performance in different care settings in the UK within the same timeframe, whereas the temporal sample tested performance during a later timeframe. The Gambian sample was used to test performance in a LMIC care setting. We assessed model performance in each validation sample separately and in the UK full (combined random and temporal samples) validation sample. Discrimination was evaluated using the c-index, and goodness-of-fit was evaluated using the Brier score. Calibration was assessed by plotting observed versus predicted risk. We assessed the overall predictive ability of the risk score using the c-index.

In the Gambian sample, we redefined low-risk, medium-risk, and high-risk groups to account for increased case fatality in this sample compared with the UK samples (appendix p 3). Observed risks in groups and population deciles of scores were derived and compared with mean predicted risks in each group or population decile of the Gambian sample.

Comparison with the CRIB II score

The NNRD did not include all the variables required for calculation of CRIB, SNAP, SNAP-II, SNAPPE-II, TRIPS, or TRIPS-II scores (appendix pp 7–9); therefore, CRIB-II was selected for comparison with NMR-2000 (appendix p 3). Because CRIB-II has only been validated for use in neonates born up to 32 weeks' gestation,[18] we compared c-indices for CRIB-II and NMR-2000 among neonates born at 32 weeks' gestation or earlier in the full validation sample. All analyses were completed using Stata (version 15).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

For the selection of candidate variables, 18 studies were reviewed to generate a list of 84 potential parameters. 45 (53·6%) of 84 parameters were considered infeasible for LMICs and were excluded, among which, 25 (55·5%) also had a low prevalence or were not included in the NNRD (figure 1). Eight (9·5%) of 84 parameters were excluded because the evidence was scarce, because the parameters had poor predictive ability in preterm neonates or neonates with low birthweight, or because the parameter had a low prevalence within 24 h of birth. 18 candidate variables were selected for inclusion in the modelling process (panel).

110176 neonates were included in the UK development and validation samples. Characteristics of the samples and participants are shown in table 1. More than half (56·6–58·2%) of the neonates had low birthweight (1500–2000 g), 28·0–28·3% had very low birthweight (1000–1499 g), and 13·7–15·1% had extremely low birthweight (<1000 g). Around half (50·4–51·3%) of the neonates were moderate-late preterm (32–36 weeks) and one-third (32·4–33·5%) were very preterm (28–31 weeks). Overall case-fatality was similar across samples (2·8–3·2%). Case-fatality of neonates with extremely low birthweight in the temporal sample (280 [12·5%] of 2238) was lower than in the other samples. No neonatal units declined to contribute data.

Characteristics of the 550 neonates in the Gambian validation sample are shown in table 2. Among the 550 neonates, 298 (54·2%) had a low birthweight,

	Neonates with available data*	Neonates with characteristic
Birthweight†	550 (100%)	··
Extremely low birthweight (<1000 g)	··	63 (11·5%)
Very low birthweight (1000–1499 g)	··	189 (34·4%)
Low birthweight (1500 to 2000 g)	··	298 (54·2%)
Sex†	549 (99·8%)	··
Male sex	··	261 (47·5%)
Mode of delivery†	488 (88·7%)	··
Spontaneous vaginal	··	342 (70·1%)
Caesarean section	··	140 (28·7%)
Assisted vaginal	··	6 (1·2%)
Multiple birth†	550 (100%)	··
Yes	··	142 (25·8%)
Inborn‡†	549 (99·8%)	··
Yes	··	299 (54·5%)
Died before discharge†	520 (94·5%)	··
Total	··	215 (41·4%)
Extremely low birthweight (<1000 g)	··	55/61 (90·2%)§
Very low birthweight (1000–1499 g)	··	93/179 (52·0%)§
Low birthweight (1500–2000 g)	··	67/280 (23·9%)§
Oxygen saturation at admission	513 (93·3%)	··
SpO₂ (%) at admission¶	··	92% (83–96)
Highest level of respiratory support within 24 h of birth†	494 (89·8%)	··
Nasal cannula	··	294 (59·5%)
CPAP ventilation	··	53 (10·7%)

Data are n (%), n/N (%), median (IQR). See panel for definition of variables. CPAP=continuous positive airway pressure. SpO₂=peripheral capillary oxygen saturation. *Out of the total 550 neonates. †Information is included on routine admission forms. ‡Defined as birth at the study hospital. §Proportion of babies in each birthweight category who died (outcome data were not available for 5·5% of babies). ¶Information collected for the trial (eKMC trial).

Table 2: Characteristics of participants in the Gambian validation sample

189 (34·4%) had a very low birthweight, and 63 (11·5%) had an extremely low birthweight. 142 (25·8%) of 550 neonates were multiple births (eg, twins), 299 (54·5%) of 549 were inborn, and 215 (41·4%) of 520 died.

The full model (18 variables) had a c-index of 0·9223 in the development sample (n=41514). After stepwise elimination, the final model included three variables (table 3), with a c-index of 0·8883 and a Brier score of 0·0232 (table 4). Complete data on all three variables were available for 46108 (83·8%) of 55029 neonates in the development sample. After imputation of missing values for predictor variables (n=54956), the resulting β coefficients were nearly identical to original estimates (appendix p 9) and model performance was unchanged (c-index 0·8894; appendix p 2). Admission age was uncertain for 5 (0·01%) of 55029 neonates; in a sensitivity analysis excluding these neonates, there was no change in performance (c-index 0·8886, Brier score 0·0232).

	β coefficient	95% confidence interval*	Integer-points†	c-index
Birthweight (g)	–0·0032	–0·0035 to –0·0029	Birthweight/100	0·8540
Highest respiratory support within first 24 h	..	..	..	0·7529
Nasal cannula or headbox	0·3167	–0·1055 to 0·7389‡	–1	..
CPAP, BiPAP or SiPAP, or invasive ventilation	1·6214	1·2682 to 1·9746‡	–5	..
SpO₂ at admission	–0·0390	–0·0455 to –0·0326	..	0·6712
<80% (reference level)	..	..	0§	..
80–89%	–0·7694	–1·0093 to –0·5294	2§	..
90–100%	–1·3697	–1·6019 to –1·1376	4§	..
Constant	2·6142¶	1·7655 to 3·4629	..	..

n=46 108. CPAP=continuous positive airway pressure. BiPAP=bilevel positive airway pressure. SiPAP=synchronised intermittent positive airway pressure. SpO₂=peripheral capillary oxygen saturation. *p<0·0001 for estimates for all variables. †Calculated by multiplying the β coefficient by a constant (–3·13) and rounding to the nearest integer. The reciprocal of the coefficient for birthweight divided by 100 ([1/–0·0032]/100=–3·13) was used as the constant to retain the exact birthweight (per 100 g) in the score. ‡p<0·0001 for overall effect of level of respiratory support. §The continuous SpO₂ parameter was categorised into clinically meaningful categorical variables; the β coefficients of these variables were multiplied by the constant to obtain integer-points; the reference level (<80%) was assigned zero points. ¶Reflects β coefficient for constant in model including SpO₂ as a continuous variable.

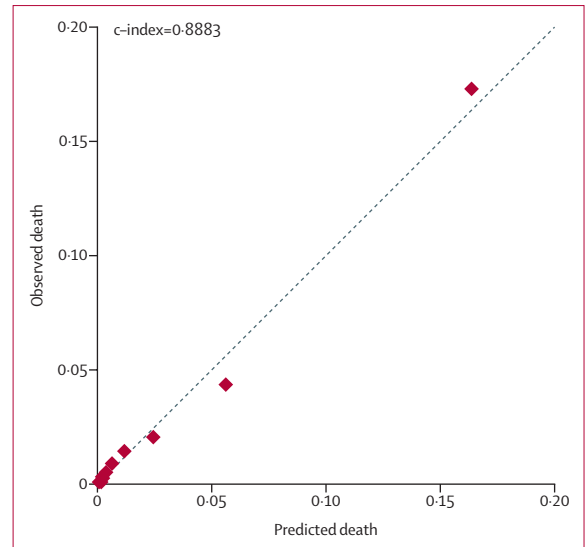*Table 3*: Derivation logistic model for the NMR-2000 score

	Development sample (n=46 108)		External validation samples			
	Original	Optimism-adjusted*	Random (n=35 193)	Temporal (n=12 653)	Full (n=47 846)	Gambian (n=457)
Brier score	0·0232	0·0233	0·0271	0·0240	0·0263	0·1688
c-index	0·8883	0·8882	0·8930	0·8859	0·8912	0·8170

*Because optimism-adjusted estimates of the c-index and Brier score were nearly identical to the original estimates, no adjustments were made to the model coefficients.

Table 4: Model performance in the development and validation samples

Figure 2: Predicted versus observed death for population deciles by predicted risk in the development sample
n=46 108. Graph created using pmcalplot.

12 793 (23·3%) of 54 956 neonates were transferred from the unit of admission to another care unit; an analysis excluding these neonates showed improved performance (c-index 0·9150, Brier score 0·0255; appendix p 2). Predictive accuracy for mortality within 24 h (c-index 0·8858, Brier score 0·0037) was nearly identical to that for in-hospital mortality. Because availability of SpO₂ monitoring is variable in LMIC settings, we tested a related variable (clinically relevant increase in apnoea or bradycardia, oxygen requirement, ventilatory support, or respiratory rate); however, this variable was not associated with in-hospital mortality (c-index 0·5061). A plot of observed versus predicted mortality risk in the development sample is shown in figure 2.

Birthweight was the most predictive variable in the model (c-index 0·8540). The reciprocal of the coefficient for birthweight divided by 100 ([1/–0·0032]/100=–3·13) was used as the constant to enable retention of exact birthweights in score calculation (table 3),[28] thereby improving predictive ability. The score range for low risk was set at 16 or more, for medium risk at 6–15, and for high risk at 5 or fewer points (appendix p 9). An example score form is shown in figure 3. Among 46 108 neonates from the development sample with complete data on the

three variables included in the final model, 27 289 (59·1%) were designated as low risk, 17 215 (37·3%) as medium risk, and 1640 (3·6%) as high risk. Observed risks were 0·3% (95% CI 0·3–0·4) for low risk, 4·1% (3·8–4·4) for medium risk, and 27·3% (25·2–29·5) for high risk, with a c-index of 0·8875. Mean predicted risks derived from regression coefficients were 0·2% (SD 0·2) for low risk, 4·6% (SD 4·2) for medium risk, and 23·5% (SD 8·8) for high risk. Observed risks across population deciles by score were similar to the risks predicted with regression coefficients (appendix p 10).

After bootstrap resampling, optimism-adjusted estimates of c-index and Brier score were nearly identical to the original measures; thus, no adjustments were made to the coefficients. In the random validation sample, complete data on all three parameters were available for 35 193 (87·3%) of 40 329 neonates, for the temporal validation sample the data were available for 12 653 (85·4%) of 14 818 neonates, for the full validation sample they were available for 47 846 (86·8%) of 55 147 neonates, and for the Gambian validation sample complete data on all three parameters were available for 457 (83·1%) of 550 neonates. The model showed very good performance across the UK validation samples (c-index 0·8859–0·8930) and good performance in the Gambian validation sample (c-index 0·8170; Brier score 0·1688; table 4). Performance was similar among neonates weighing 1500 g or less in the Gambian sample (c-index 0·8069, Brier score 0·1753).

Graphical plots showed a high level of agreement between observed and predicted mortality risk across the external validation samples (figure 4). Applying the empirical optimal cutpoint of 3·9% based on the Youden Index gave moderately high sensitivity (79·1–81·6) and

specificity (81·0–82·9), with high negative predictive value (99·3) and low positive predictive value (11·4–12·2; appendix p 10).

The discriminatory ability of the simplified integer score was similar to the model using regression coefficients, with c-indices of 0·8903 in the full validation sample and 0·8082 in the Gambian validation sample (appendix p 11). Among the 47 846 neonates in the UK full validation sample, 28 565 (59·7%) were designated as low risk, 17 407 (36·4%) as medium risk, and 1874 (3·9%) as high risk. Observed risks for these categories were 0·4% (95% CI 0·3–0·5) for low risk, 4·8% (4·5–5·1) for medium risk, and 29·7% (27·7–31·8) for high risk. In the Gambian validation sample, the score range for low risk was set at 23 or more, for medium risk at 17–22, and for high risk at 16 or fewer points (appendix pp 3, 11). Among the 457 neonates in the Gambian sample for whom data on all three parameters were available, 28 (6·1%) were designated as low risk, 215 (47·1%) as medium risk, and 214 (46·8%) as high risk. Observed risks were 10·7% (95% CI 3·5–28·5) for low risk, 21·4% (16·4–27·4) for medium risk, and 68·2% (61·7–74·1) for high risk. Mean predicted risks derived from regression coefficients were 9·4% (SD 1·9) for low risk, 22·3% (SD 8·5) for medium risk, and 67·4% (SD 18·4) for high risk. Observed risks across population deciles by score were similar to those predicted with coefficients (appendix p 11).

Comparison of areas under the ROC curves for NMR-2000 (c-index 0·8523 [95% CI 0·8336–0·8710]) and CRIB-II (c-index 0·7443 [95% CI 0·7153–0·7733]) among 10 812 neonates born at 32 weeks' gestation or earlier (figure 5) indicated that discriminatory performance of NMR-2000 was superior to that of CRIB-II (p<0·0001).

Discussion

This population-wide study, including data from 110 176 newborn babies at 187 hospitals in the UK and 550 newborn babies at one hospital in The Gambia, has derived and validated NMR-2000 for predicting in-hospital mortality. A strength of this work is that, to our knowledge, this is the largest dataset that has been used to develop and validate a neonatal mortality risk score. Among neonates born at 32 weeks' gestation or earlier, the discriminatory ability of NMR-2000 was superior to that of CRIB-II, one of the most widely used neonatal risk scores.

Performance of the NMR-2000 simplified integer score, which can be measured and calculated at the bedside, was similar to that of the model using regression coefficients. The three parameters used in the score can be feasibly collected in LMIC settings. Although sub-Saharan Africa and southern Asia account for 78% of the world's neonatal deaths,[1] existing risk scores have primarily been developed for intensive care settings and often require complex calculations. In LMICs, where parameters in widely used scores are typically not available nor reliably measurable,[20,21] NMR-2000 could
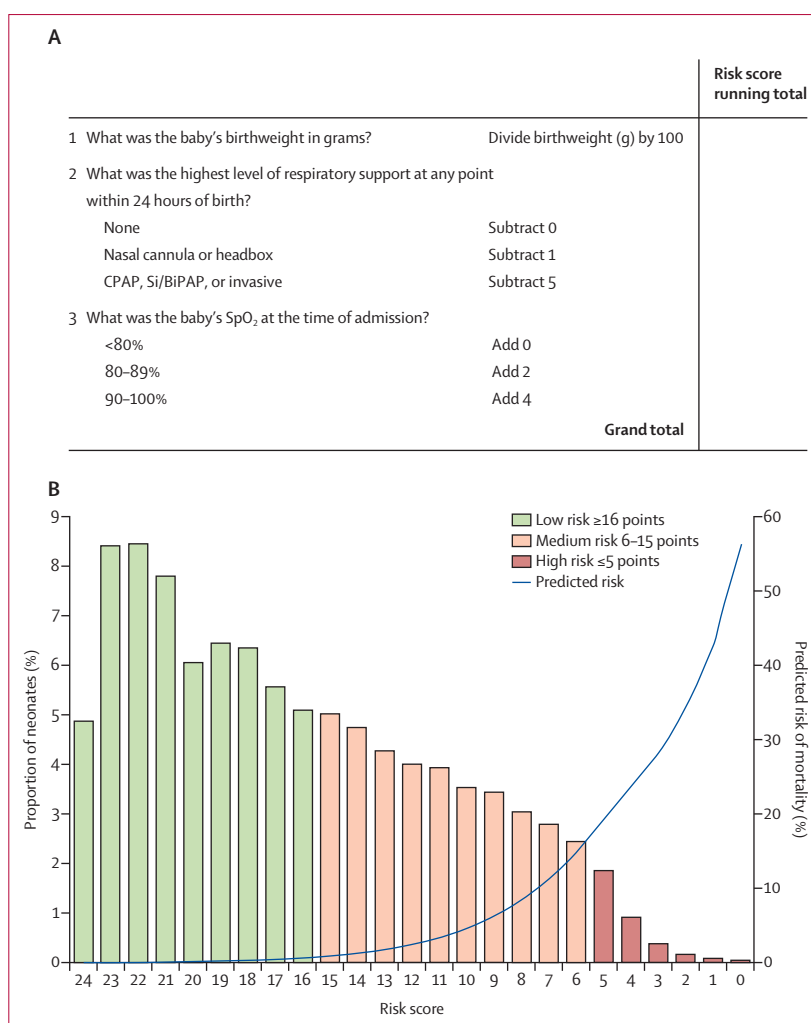
Figure 3: NMR-2000, a simplified risk score to predict mortality amongst neonates weighing 2000 g or less
(A) An example mortality risk score for clinical use. (B) Predicted risk of in-hospital mortality plotted against risk scores in the development sample; bar data indicate proportion of neonates; blue line indicates predicted mortality risk. CPAP=continuous positive airway pressure. BiPAP=bilevel positive airway pressure. SiPAP=synchronised intermittent positive airway pressure.

support shared decision making by enabling providers to objectively assess illness severity.[29] The score could be used in clinical trials to assess eligibility and compare participants.[29] Additionally, NMR-2000 could inform service delivery planning by identifying bottlenecks in care provision.[6] Given that 73% of neonatal deaths occur within the first 7 days of life,[24] early recognition of severe illness and rapid initiation of evidence-based interventions are crucial to promoting survival.[5,9,12]

The c-index was 0·8859–0·8930 across the development and UK validation samples, suggesting that NMR-2000 can discriminate neonates who will die from neonates who will survive. This level of performance is similar to that of commonly used neonatal mortality scores in high-resource settings. Discriminatory ability at the time of model derivation ranged from 0·87 for TRIPS-II[16] to 0·92 for CRIB-II.[18] Similar to TRIPS-II, NMR-2000 can be
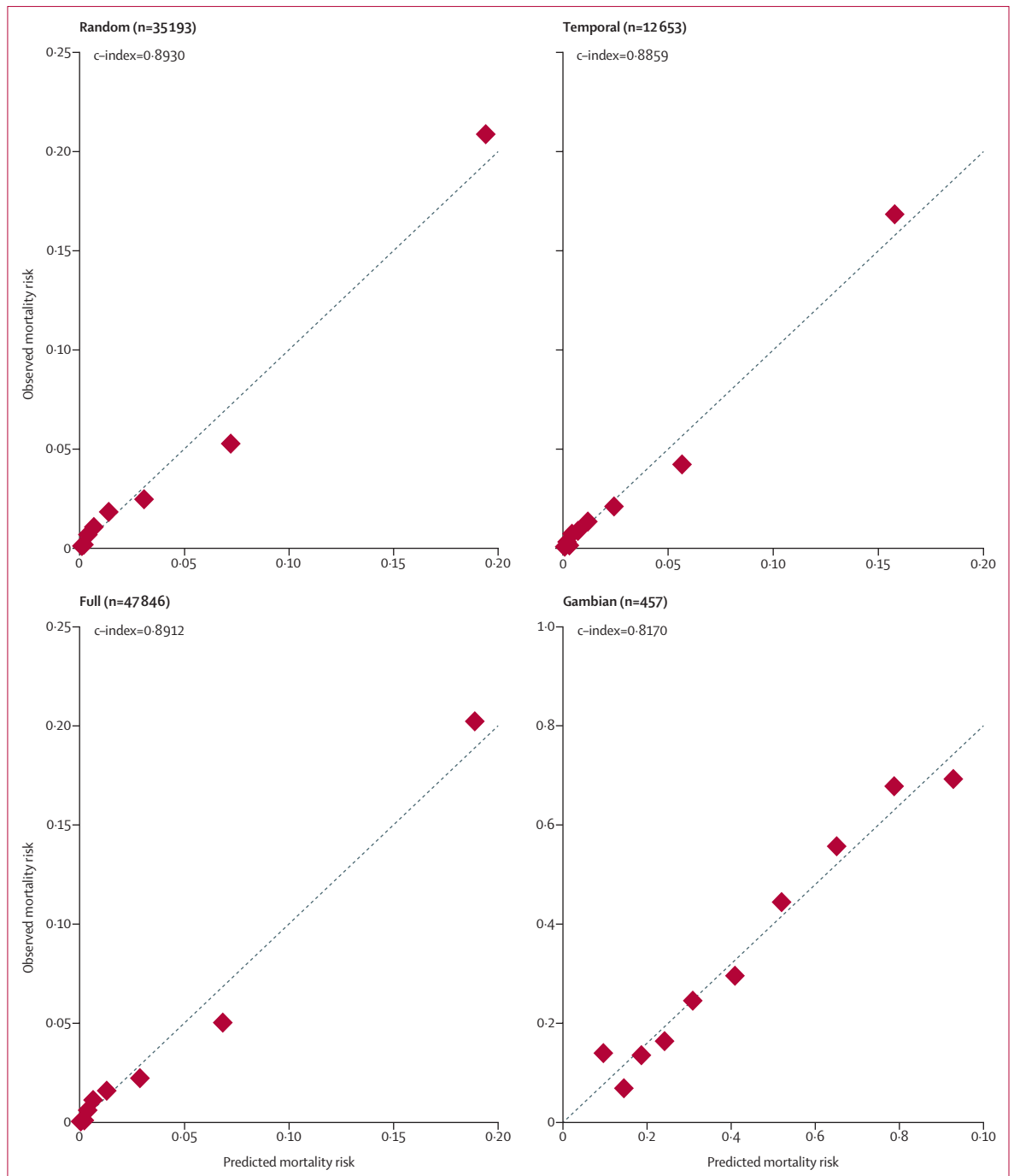
***Figure 4:* Predicted versus observed mortality risk for population deciles in the random, temporal, full, and Gambian validation samples**
Graphs created using pmcalplot.

assessed at any point within the first 24 h of life and could be repeated if the level of respiratory support increases. The performance of NMR-2000 (c-index 0·8523) was superior to that of CRIB-II (0·7443) among neonates born at 32 weeks' gestation or earlier. The model also showed very good predictive ability for mortality within 24 h of birth (c-index 0·8858), which is notable because 37% of

neonatal deaths in sub-Saharan Africa occur within this timeframe.[24] The NMR-2000 model showed a high level of agreement between observed and predicted deaths, as assessed by calibration plots, in the development and validation samples. Calibration plots are the preferred method for assessing calibration.[25] Previous neonatal scores, including NTISS,[13] SNAPPE-II,[15] CRIB-II,[18] and

TRIPS-II,[16] were reported to have good calibration for predicting in-hospital mortality using the Hosmer-Lemeshow test. However, such results should be interpreted with caution given the limitations of this test, which include subjective and imprecise grouping of babies as well as inability to denote the directionality of miscalibration when incongruities are detected.[25]

In the Gambian validation sample, the NMR-2000 model had good discrimination and overall goodness-of-fit, with c-index of 0·8170 and Brier score of 0·1688. Complete data were available for 83% of neonates. The calibration plot showed a strong agreement between observed and predicted mortality. These findings suggest that NMR-2000 is valid for use in LMIC settings where pulse oximetry is available. Discrimination of the SAWS score, developed for neonates weighing 1500 g or less in low-resource settings, at the time of validation (c-index 0·679–0·698)[22] was decreased relative to NMR-2000 among Gambian neonates weighing 1500 g or less (c-index 0·8069). Notably, neither goodness-of-fit nor calibration were reported for the SAWS score.[22] Further, SAWS relies on accurate assessment of gestational age, which can be challenging in LMICs because of late presentation for antenatal care, poor recall of last menstrual period, and unavailability of ultrasonography.[30] Case-fatality of Gambian neonates in this study is similar to that reported from a previous study at EFSTH (35% overall, 58% for neonates with a very low birthweight),[23] and higher than studies at similar hospitals in Ghana (20% overall),[31] Nigeria (14–20% overall),[32,33] and Burkina Faso (15% overall).[34]

Among the three NMR-2000 parameters, all except SpO_2 are included on routine admission forms at EFSTH, Gambia (at time of screening),[23] as well as standard forms at government hospitals in Kenya, Malawi, and Tanzania. We were able to obtain SpO_2 data for the Gambian sample primarily because these data were being collected as part of the eKMC trial screening process. Variability in the implementation of routine pulse oximetry is a crucial gap in low-resource neonatal units.[35,36] In a study of nearly 7500 neonates admitted to 11 hospitals in Nigeria, hypoxaemia increased the adjusted odds of mortality by six times, and clinical signs (eg, chest in-drawing, grunting) poorly predicted hypoxaemia.[37] Furthermore, expansion of neonatal inpatient care, often of variable quality and frequently inclusive of unmonitored 100% oxygen supplementation, has placed sub-Saharan Africa on the brink of an epidemic of retinopathy of prematurity.[38] Widespread availability of SpO_2 monitoring and improved coverage of screening for and treatment of retinopathy of prematurity will be essential to control the incidence of visual loss in affected neonates. In LMIC settings, successful implementation of NMR-2000 would require sensitisation around recording the three parameters. Several studies have highlighted issues surrounding the collection of data on neonatal care in LMICs, including variable uptake of standard admission forms;[36] incomplete documentation of assessments, monitoring, and therapies
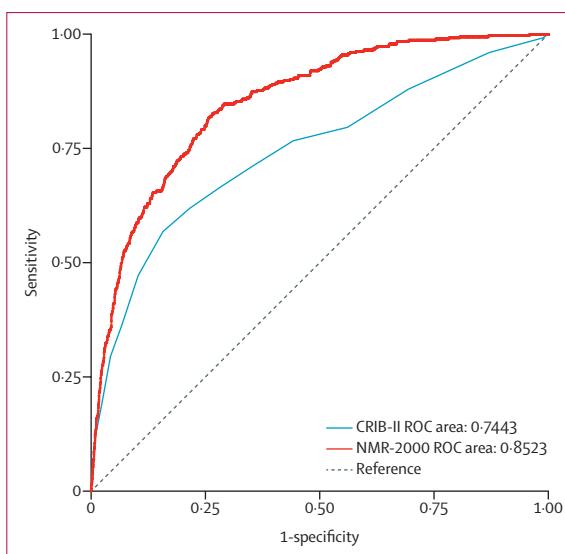
Figure 5: Comparison of areas under the ROC curves for CRIB-II and NMR-2000
This analysis includes neonates born at 32 weeks' gestation or earlier in the full validation sample (n=10 812). ROC=receiver operating characteristic.

prescribed;[20,23,36] and low capacity of data systems to capture information on neonates who die soon after birth or are transferred to another facility.[4] Increasing the quality and coverage of data is crucial to promote actions to improve neonatal survival, and will require coordination across different levels of the health-care system.

One strength of this study is our use of a large and purposely-selected UK dataset, which enabled maximisation of model performance. One limitation is that the Gambian dataset was small and limited to a single hospital; research is required to validate the model using a larger LMIC dataset. Several candidate variables in the development sample had a considerable proportion of missing data, including admission heart rate, respiratory rate, and SpO_2. It was not possible to compare NMR-2000 with the CRIB, SNAP, SNAP-II, SNAPPE-II, or TRIPS-II scores, because the NNRD did not include all parameters required for their calculation. Because pulse oximetry is not always available in low-resource neonatal units, the usefulness of the NMR-2000 score in such settings could be limited. The NNRD did not include clinical signs of respiratory distress that could be tested as a potential proxy for SpO_2. We tested a related variable (clinically relevant increase in apnoea or bradycardia, oxygen requirement, ventilatory support, or respiratory rate); however, this variable was not associated with mortality. The use of respiratory support level as a parameter could affect the performance of the model. Administration of therapies varies in line with variations in clinical practice[19] and resource availability and so might not reflect true therapeutic requirements.[36] In LMICs, delivery systems for oxygen therapy and CPAP might be unavailable or non-functional, and related supplies (eg, nasal cannulas) might be out of stock.[8,35,36]

Research is required to validate the NMR-2000 score in low-resource settings using a sufficiently sized dataset, and to evaluate its usefulness for supporting clinical decision making.[29] A follow-up study using a large, multihospital dataset from Kenya is planned. Nurses have essential roles as frontline providers of neonatal care; however, there is a severe shortage of neonatal nurses in LMICs.[6,7] Future research could explore the model's ability to inform resource use,[13] particularly nursing workload.

The NMR-2000 is a simplified risk score, validated for high-resource and low-resource settings where pulse oximetry is available, to accurately predict in-hospital mortality among neonates weighing 2000 g or less. By enabling providers to objectively assess illness severity, this tool could contribute to improvements in the quality of care delivered in LMIC facilities. Early recognition of severe illness and rapid initiation of evidence-based interventions are crucial to promoting survival of small and vulnerable neonates.

References
1 UN Inter-agency Group for Child Mortality Estimation. Levels & trends in child mortality: report 2019. https://www.unicef.org/media/60561/file/UN-IGME-child-mortality-report-2019.pdf (accessed Nov 1, 2019).

2 Liu L, Oza S, Hogan D, et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet* 2016; **388:** 3027–35.

3 Chawanpaiboon S, Vogel JP, Moller AB, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob Health* 2019; **7:** e37–46.

4 Blencowe H, Krasevec J, de Onis M, et al. National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: a systematic analysis. *Lancet Glob Health* 2019; **7:** e849–60.

5 Lawn JE, Davidge R, Paul VK, et al. Born too soon: care for the preterm baby. *Reprod Health* 2013; **10** (suppl 1)**:** S5.

6 Dickson KE, Simen-Kapeu A, Kinney MV, et al. Every Newborn: health-systems bottlenecks and strategies to accelerate scale-up in countries. *Lancet* 2014; **384:** 438–54.

7 Moxon SG, Lawn JE, Dickson KE, et al. Inpatient care of small and sick newborns: a multi-country analysis of health system bottlenecks and potential solutions. *BMC Pregnancy Childbirth* 2015; **15** (suppl 2)**:** S7.

8 Morgan MC, Spindler H, Nambuya H, et al. Clinical cascades as a novel way to assess physical readiness of facilities for the care of small and sick neonates in Kenya and Uganda. *PLoS One* 2018; **13:** e0207156.

9 Lawn JE, Blencowe H, Oza S, et al. Every Newborn: progress, priorities, and potential beyond survival. *Lancet* 2014; **384:** 189–205.

10 Katz J, Lee AC, Kozuki N, et al. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. *Lancet* 2013; **382:** 417–25.

11 Blencowe H, Lee AC, Cousens S, et al. Preterm birth-associated neurodevelopmental impairment estimates at regional and global levels for 2010. *Pediatr Res* 2013; **74** (suppl 1)**:** 17–34.

12 Bhutta ZA, Das JK, Bahl R, et al. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? *Lancet* 2014; **384:** 347–70.

13 Gray JE, Richardson DK, McCormick MC, Workman-Daniels K, Goldmann DA. Neonatal therapeutic intervention scoring system: a therapy-based severity-of-illness index. *Pediatrics* 1992; **90:** 561–67.

14 Richardson DK, Gray JE, McCormick MC, Workman K, Goldmann DA. Score for neonatal acute physiology: a physiologic severity index for neonatal intensive care. *Pediatrics* 1993; **91:** 617–23.

15 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.

16 Lee SK, Aziz K, Dunn M, et al. Transport Risk Index of Physiologic Stability, version II (TRIPS-II): a simple and practical neonatal illness severity score. *Am J Perinatol* 2013; **30:** 395–400.

17 The International Neonatal Network. The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993; **342:** 193–98.

18 Parry G, Tucker J, Tarnow-Mordi W. CRIB II: an update of the clinical risk index for babies score. *Lancet* 2003; **361:** 1789–91.

19 Richardson D, Tarnow-Mordi WO, Lee SK. Risk adjustment for quality improvement. *Pediatrics* 1999; **103** (suppl E)**:** 255–65.

20 Morgan MC, Nambuya H, Waiswa P, et al. Kangaroo mother care for clinically unstable neonates weighing ≤2000 g: is it feasible at a hospital in Uganda? *J Glob Health* 2018; **8:** 010701.

21 Dorling JS, Field DJ. Value and validity of neonatal disease severity scoring systems. *Arch Dis Child Fetal Neonatal Ed* 2008; **93:** F80–82.

22 Rosenberg RE, Ahmed S, Saha SK, et al. Simplified age-weight mortality risk classification for very low birth weight infants in low-resource settings. *J Pediatr* 2008; **153:** 519–24.

23 Okomo UA, Dibbasey T, Kassama K, et al. Neonatal admissions, quality of care and outcome: 4 years of inpatient audit data from The Gambia's teaching hospital. *Paediatr Int Child Health* 2015; **35:** 252–64.

24 Oza S, Cousens SN, Lawn JE. Estimation of daily risk of neonatal death, including the day of birth, in 186 countries in 2013: a vital-registration and modelling-based study. *Lancet Glob Health* 2014; **2:** e635–44.

25 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; **35:** 1925–31.

26 Royal College of Paediatrics and Child Health. UK-WHO growth chart—0–4 years. https://www.rcpch.ac.uk/resources/uk-who-growth-charts-0-4-years (accessed Jan 16, 2020).

27 Department of Health. Toolkit for high quality neonatal services. http://www.londonneonatalnetwork.org.uk/wp-content/uploads/2015/09/Toolkit-2009.pdf (accessed Jan 29, 2020).

28 Malin GL, Morris RK, Riley R, Teune MJ, Khan KS. When is birthweight at term abnormally low? A systematic review and meta-analysis of the association and predictive ability of current birthweight standards for neonatal outcomes. *BJOG* 2014; **121:** 515–26.

29 Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013; **10:** e1001381.

30 Lee AC, Mullany LC, Ladhani K, et al. Validity of newborn clinical assessment to determine gestational age in Bangladesh. *Pediatrics* 2016; **138:** e20153303.

31 Owusu BA, Lim A, Makaje N, Wobil P, SameAe A. Neonatal mortality at the neonatal unit: the situation at a teaching hospital in Ghana. *Afr Health Sci* 2018; **18:** 369–77.

32 Omoigberale AI, Sadoh WE, Nwaneri DU. A 4 year review of neonatal outcome at the University of Benin Teaching Hospital, Benin City. *Niger J Clin Pract* 2010; **13:** 321–25.

33 Ekwochi U, Ndu IK, Nwokoye IC, Ezenwosu OU, Amadi OF, Osuorah D. Pattern of morbidity and mortality of newborns admitted into the sick and special care baby unit of Enugu State University Teaching Hospital, Enugu State. *Niger J Clin Pract* 2014; **17:** 346–51.

34 Kouéta F, Yé D, Dao L, Néboua D, Sawadogo A. Neonatal morbidity and mortality in 2002–2006 at the Charles de Gaulle pediatric hospital in Ouagadougou (Burkina Faso). *Sante* 2007; **17:** 187–91 (in French).

35 Morgan MC, Maina B, Waiyego M, et al. Pulse oximetry values of neonates admitted for care and receiving routine oxygen therapy at a resource-limited hospital in Kenya. *J Paediatr Child Health* 2018; **54:** 260–66.

36 Aluvaala J, Nyamai R, Were F, et al. Assessment of neonatal care in clinical training facilities in Kenya. *Arch Dis Child* 2015; **100:** 42–47.

37 Graham H, Bakare AA, Ayede AI, et al. Hypoxaemia in hospitalised children and neonates: a prospective cohort study in Nigerian secondary-level hospitals. *EClinicalMedicine* 2019; **16:** 51–63.

38 Gilbert C, Malik ANJ, Nahar N, et al. Epidemiology of ROP update—Africa is the new frontier. *Semin Perinatol* 2019; **43:** 317–22.