



OPEN ACCESS

# Prediction modelling of inpatient neonatal mortality in high-mortality settings

Jalemba Aluvaala <sup>1,2,3</sup>, Gary Collins <sup>4,5</sup>, Beth Maina,<sup>6</sup> Catherine Mutinda,<sup>6</sup> Mary Waiyego,<sup>6</sup> James Alexander Berkley <sup>3,7,8</sup>, Mike English <sup>1,3</sup>

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/archdischild-2020-319217>).

<sup>1</sup>Health Services Unit, KEMRI-Wellcome Trust Research Programme, Nairobi, Kenya

<sup>2</sup>Paediatrics and Child Health, University of Nairobi, Nairobi, Kenya

<sup>3</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

<sup>4</sup>Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, University of Oxford, Oxford, United Kingdom

<sup>5</sup>Oxford University Hospitals, NHS Foundation Trust, Oxford, United Kingdom

<sup>6</sup>Pumwani Maternity Hospital, Nairobi, Kenya

<sup>7</sup>KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

<sup>8</sup>The Childhood Acute Illness & Nutrition (CHAIN) Network, P.O. Box 43640 – 00100, Nairobi, Kenya

## Correspondence to

Dr Jalemba Aluvaala, Health Services Unit, KEMRI-Wellcome Trust Research Programme, P.O. Box 43640 – 00100, Nairobi, Kenya; [jaluvaala@kemri-wellcome.org](mailto:jaluvaala@kemri-wellcome.org)

Received 27 March 2020

Revised 2 August 2020

Accepted 5 September 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY. Published by BMJ.

**To cite:** Aluvaala J, Collins G, Maina B, et al. *Arch Dis Child* Epub ahead of print: [please include Day Month Year]. doi:10.1136/archdischild-2020-319217

## ABSTRACT

**Objective** Prognostic models aid clinical decision making and evaluation of hospital performance. Existing neonatal prognostic models typically use physiological measures that are often not available, such as pulse oximetry values, in routine practice in low-resource settings. We aimed to develop and validate two novel models to predict all cause in-hospital mortality following neonatal unit admission in a low-resource, high-mortality setting.

**Study design and setting** We used basic, routine clinical data recorded by duty clinicians at the time of admission to derive (n=5427) and validate (n=1627) two novel models to predict in-hospital mortality. The Neonatal Essential Treatment Score (NETS) included treatments prescribed at the time of admission while the Score for Essential Neonatal Symptoms and Signs (SENSS) used basic clinical signs. Logistic regression was used, and performance was evaluated using discrimination and calibration.

**Results** At derivation, c-statistic (discrimination) for NETS was 0.92 (95% CI 0.90 to 0.93) and that for SENSS was 0.91 (95% CI 0.89 to 0.93). At external (temporal) validation, NETS had a c-statistic of 0.89 (95% CI 0.86 to 0.92) and SENSS 0.89 (95% CI 0.84 to 0.93). The calibration intercept for NETS was –0.72 (95% CI –0.96 to –0.49) and that for SENSS was –0.33 (95% CI –0.56 to –0.11).

**Conclusion** Using routine neonatal data in a low-resource setting, we found that it is possible to predict in-hospital mortality using either treatments or signs and symptoms. Further validation of these models may support their use in treatment decisions and for case-mix adjustment to help understand performance variation across hospitals.

## INTRODUCTION

Kenya and other low-income to middle-income countries (LMICs) accounted for 99% of global neonatal mortality in 2017. Improved delivery of essential interventions in hospitals is expected to play a key role in achieving Sustainable Development Goal 3.2—a neonatal mortality rate of 12 per 1000 live births or lower.<sup>1–3</sup> A better understanding of hospitals' neonatal mortality and reliable and timely information on how this varies may support efforts to improve hospital care at scale.<sup>4,5</sup>

Well-performing prognostic models can aid clinical decisions and at health system level may support better decisions to improve services.<sup>6</sup> We reviewed existing neonatal prognostic models, and these predominantly use physiological parameters (eg,

## What is already known on this topic?

- Existing neonatal prognostic models are suited for advanced care settings as they use parameters that are not available in routine practice in low-resource settings.
- These parameters include pulse oximetry, blood gases and other laboratory tests.

## What this study adds?

- Application of recent methodological guidance; Prognosis Research Strategy and Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis.
- Using routine neonatal data in a low-resource setting, we found that it is possible to accurately predict in-hospital mortality.
- Further validation of these models may support their use in treatment decisions and for case-mix adjustment to help understand and improve performance variation across hospitals.

blood gas values) that are not available in routine practice in LMICs.<sup>7</sup> However, the use of treatment data is a potentially useful approach to predicting in-hospital neonatal mortality in LMICs, given its greater availability.<sup>7</sup> Current models that use treatments as predictors typically focus on those given in intensive care, and there are limitations in the methods used to develop and validate them.<sup>7</sup> These limitations may be addressed by the selection of treatments relevant to the LMIC setting and application of recommended approaches to prognostic model development and validation.<sup>8</sup>

Candidate treatment predictors are essential interventions included in clinical practice guidelines for in-hospital neonatal care developed by the WHO and adapted for Kenya.<sup>9</sup> Alternative predictors are simple clinical signs recommended for use in routine practice in Kenya,<sup>9</sup> and we have recently shown data on these candidate predictors can be collected.<sup>10</sup>

Our aim was to use these data to develop and validate two models to predict in-hospital neonatal mortality, the Neonatal Essential Treatment Score (NETS) and the Score of Essential Neonatal Symptoms and Signs (SENSS). These data could be combined in one model but as head-to-head comparisons of prognostic models in the same population are rare, we aimed as a start to evaluate

which approach might be most appropriate (based on data availability and performance) for the population of interest.<sup>11</sup>

## METHODS

### Source of data

Data were obtained from a routine data collection system described in detail elsewhere.<sup>12</sup> At the time of this study, neonatal unit data were from one hospital—a large urban maternity hospital in Nairobi County, Kenya.<sup>12</sup> The unit provides essential inpatient care to 60 neonates at any given time with approximately 4500 admitted annually.

### Participants

Two data sets were extracted for this study. The first included admissions from April 2014 to December 2015 (n=9115) and the second included admissions from January 2016 to July 2016 (n=2735). The data capture system randomly selected 60% of the cases for collection of a comprehensive set of clinical and treatment data (the full data set), while for 40% of admissions, fewer variables were collected (the minimum data set). The full data sets from the first (n=5427) and second (n=1627) periods were used for model derivation and validation, respectively.

### Outcome

The outcome was all-cause in-hospital neonatal unit mortality. Outcomes should be assessed blind to predictors to avoid influencing outcome assessment.<sup>8</sup> The data collectors were not blinded but were unaware of these prognostic model analyses.

### Predictors

The treatment predictors included in NETS were use/non-use of supplementary oxygen, enteral feeds, first-line intravenous antibiotics (penicillin and gentamicin) and parenteral phenobarbital (online supplemental table 1).<sup>9</sup> Continuous positive airway pressure, phototherapy, exchange transfusion and kangaroo mother care were rarely prescribed (<1.5%) at admission and were omitted.

For the SENSS model, presence or absence of difficulty feeding, convulsions, indrawing, central cyanosis and floppy/inability to suck were included (online supplemental table 2).<sup>9</sup> We excluded temperature and respiratory rate as these were poorly documented (89% and 65% missingness, respectively). We also excluded bulging fontanelle as it was a rare sign (present in 0.2% of cases in the derivation data set).

We also included birth weight by category (<1 kg, 1.0–<1.5 kg, 1.5–<2.5 kg, 2.5–4.0 kg and >4 kg) and sex in both models. Weight as a continuous predictor is preferable to avoid information loss. However, these categories are based on a priori clinical consensus rather than data driven.<sup>8,9</sup> Gestation at birth was not included due to 70% missingness. This rate of missingness is consistent with previous work in Kenyan hospitals.<sup>13</sup>

### Sample size

Seven predictors (six binary and one categorical) were included in each model. The five birthweight categories required four parameters to be calculated. There was thus a total of 10 parameters in each model against 445 deaths, giving 45 deaths per variable (445/10). This exceeds the recommended ratio of 10 events per prognostic variable and the more recently suggested ratio of 20 events per variable for model derivation.<sup>14</sup> The external validation data set contained 151 in-hospital deaths and 1476 non-events (no deaths). This is more than the minimum recommended 100 events and 100 non-events for validation studies.<sup>15</sup>

### Missing data

For NETS derivation, 587 out of the 5427 observations were excluded as treatment sheets were missing in the patient files (online supplemental table 3), leaving 4840 with 447 in-hospital deaths (9%). In the validation data set, 143 out of 1627 observations had no treatment sheets (online supplemental table 4), leaving 1443 with 137 in-hospital deaths (9.5%). We considered multiple imputation inappropriate, given that entire treatment sheets were missing which would necessitate imputation based on a limited set of non-treatment data such as clinical symptoms.

Predictor missingness in the SENSS derivation data set (online supplemental table 5) ranged from 0.2% (sex and birth weight) to 16% (floppy/unable to suck). In the validation data set, missingness (online supplemental table 5) ranged from 0.1% (sex) to 14% (severe indrawing). We assumed a missing at random (MAR) mechanism for the observed missingness. Multiple imputation using the chained equation approach was implemented for both data sets.<sup>16</sup>

### Statistical analysis methods

Logistic regression without variable selection was used. We derived SENSS using 31 imputed data sets (based on 31% of observations missing at least one variable). Parameter estimates were then combined using Rubin's rule.

Model calibration was assessed by plotting the predicted probability of in-hospital death against the observed proportion. Discrimination was assessed by the c-statistic (equivalent to the area under the receiver operating curve).<sup>15,17</sup> The prediction models were internally validated using bootstrapping to assess any overfitting.

Temporal external validation was done by applying the model coefficients obtained at derivation to the external validation data. For SENSS, this required imputation of 23 data sets (23% of the cohort had one or more missing values).

## RESULTS

The clinical characteristics of all eligible patients are shown in table 1. Slightly more than half of the neonates were male in the derivation and validation data sets. The majority (70%) had a normal birth weight (2.5–4.0 kg), and less than 5% were born outside the facility.

### Derivation model specification

A complete case regression analysis was used for the treatment model. A total of 19 patients were missing outcome, 9 were missing birth weights and 12 were missing sex, leading to a total of 40 observations omitted (table 1). The derivation model therefore included 4800 observations with 445 deaths.

### Model results

Box 1 shows the final equations for NETS and SENSS models in the log odds scale.

### Apparent performance

Figure 1 shows model apparent performance in the derivation data sets. The c-statistic (model discrimination) for NETS was 0.92 (95% CI 0.90 to 0.93) and that for SENSS was 0.91 (95% CI 0.89 to 0.93). The predicted probability of in-hospital death for patients who died was therefore higher than for those who left the unit alive 92% and 91% of the time for the NETS and SENSS, respectively.

**Table 1** Characteristics of patients included in model derivation and external validation

Characteristic	Derivation				External validation			
	NETS, n=4840		SENSS, n=5427*		NETS, n=1443		SENSS, n=1627*	
	n†	%	n†	%	n†	%	n†	%
<b>Sex</b>								
Male	2605	54	2942	54	850	59	962	60
Missing	12	0.3			2	0.1		
<b>Birth weight (kg)</b>								
<1	31	0.6	32	0.6	10	1	10	1
1.0–<1.5	115	2	136	3	40	3	45	3
1.5–<2.5	1043	22	1182	22	316	22	361	22
2.5–4.0	3438	71	3848	71	1002	69	1126	69
>4.0	204	4	229	4	74	5	85	5
Missing	9	0.2			1	0.1		
<b>Mode of delivery</b>								
Spontaneous vaginal	2697	57	3107	57	897	63	1018	63
Assisted vaginal	1	0.02	6	0.1	0	0	1	0.1
Breech	40	1	102	2	19	1	34	2
Caesarean section	1999	42	2212	41	509	36	574	35
Missing	145	3			18	1		
<b>Outborn§‡</b>								
Yes	107	2	123	2	57	4	60	4
Missing	0	0			0	0		
<b>HIV exposure</b>								
Exposed	287	6	338	6	74	5	93	6
Missing	277	6			80	6		
<b>Outcome</b>								
Alive	4374	90	4918	91	1300	90	1476	91
Dead	447	9	509	9	137	9	151	9
Missing	19	0.4			6	0.4		

\*Data presented are after multiple imputation. The multiple imputation filled in the missing values while preserving the pattern of distribution observed in the original data sets (online supplemental table 8).

†Denominators for the variables obtained by subtracting the missing data from the sample (4840 for derivation, 1443 for external validation)

‡Outborn refers to neonates admitted to the unit having been born either in another facility, at home or on the way to hospital

NETS, Neonatal Essential Treatment Score; SENSS, Score for Essential Symptoms and Signs.

### Internal validation

There was a small difference between the apparent performance and the optimism-corrected performance, indicating

negligible overfitting (table 2). The original coefficients were therefore taken forward for external validation without shrinkage.

### Box 1 Logistic regression models for NETS and SENSS

#### NETS

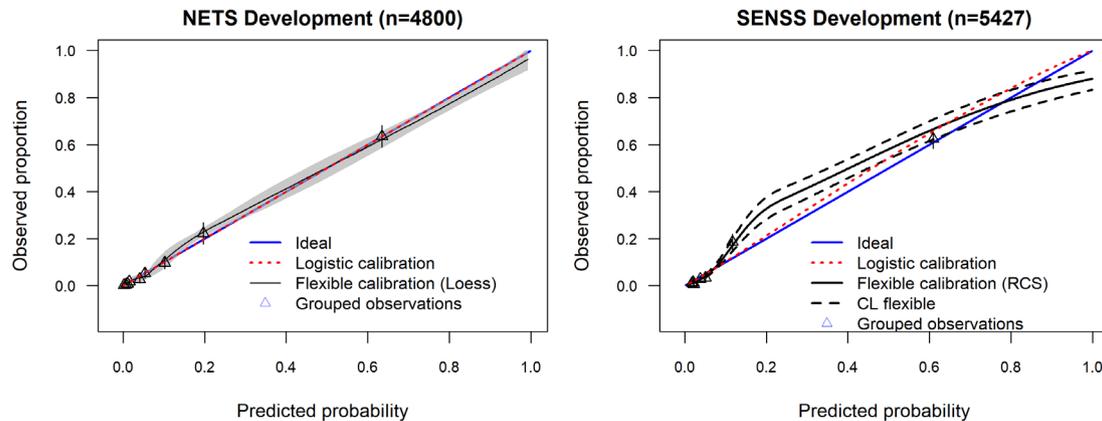
Linear predictor =  $-4.1521 + 5.6836 * ELBW + VLBW + 1.4186 * LBW - 0.2927 * Macrosomia - 0.3125 * Male + 1.3695 * Antibiotics + 1.3256 * Fluids - 1.9135 * Feeds + 0.6142 * Oxygen + 2.5947 * Phenobarbital$

#### SENSS

Linear predictor =  $-3.8583 + 5.7580 * ELBW + 3.7082 * VLBW + 0.9232 * LBW - 0.4918 * Macrosomia - 0.1336 * Male + 1.3596 * Difficulty feeding + 1.3977 * Convulsion + 1.9790 * Indrawing + 0.9584 * Cyanosis + 1.6266 * Floppy unable to suck$

For each variable, the presence of the indicator takes a value of 1 and absence takes a value of 0. The coefficients are summated to give the linear predictor, which is then converted to predicted probability of in-hospital mortality.

ELBW, extremely low birth weight; LBW, low birth weight; NETS, Neonatal Essential Treatment Score; SENSS, Score of Essential Neonatal Symptoms and Signs; VLBW, very low birth weight.



**Figure 1** Calibration plot for NETS and SENSS in the derivation data sets. NETS, Neonatal Essential Treatment Score; SENSS, Score for Essential Neonatal Symptoms and Signs; RCS, Restricted Cubic Splines; CL, Confidence Limits (95%).

### External validation

For NETS, there was a higher proportion of prescription of all the five treatments in the external validation data set compared with the derivation data set (online supplemental table 6). A smaller proportion of deaths was observed in those who had a prescription of intravenous antibiotics, intravenous fluids, oxygen and phenobarbital in the external validation data set.

For SENSS, among the patients who died, the proportion with clinical signs present was higher in the derivation data set compared with the external validation data set for four out of the five signs included as predictors (online supplemental table 7).

The two models demonstrated similar discrimination at external validation (figure 2). The c-statistic was 0.89 (95% CI 0.86 to 0.92) for NETS (compared with 0.92 after internal validation) and 0.89 (95% CI 0.84 to 0.93) for SENSS (compared with 0.91 after internal validation). However, greater deterioration was observed for calibration. The calibration intercept dropped to  $-0.72$  (95% CI  $-0.96$  to  $-0.49$ ) for NETS and  $-0.33$  (95% CI  $-0.56$  to  $-0.11$ ) for SENSS.

### DISCUSSION

The NETS and SENSS models demonstrated similar performance at derivation; c-statistic (discrimination) for NETS was 0.92 (95% CI 0.90 to 0.93) and that for SENSS was 0.91 (95% CI 0.89 to 0.93). There was minimal overfitting at derivation for both. At external validation, NETS had a c-statistic of 0.89 (95% CI 0.86 to 0.92) and SENSS had a c-statistic of 0.89 (95% CI 0.84 to 0.93), while the calibration intercept for NETS dropped to  $-0.72$  (95% CI  $-0.96$  to  $-0.49$ ) and that for SENSS dropped to  $-0.33$  (95% CI  $-0.56$  to  $-0.11$ ).

Model predictors were preselected based on their availability in clinical practice.<sup>9 18</sup> No selection procedures were employed for two reasons. First, the ideal case is a limited number of

preselected predictors entered into a full model.<sup>15</sup> In addition, a more parsimonious model (fewer predictors) may be expected to have better predictive performance.<sup>15</sup> However, besides the 0.8 threshold for good discrimination, there is no consensus on relative discriminatory ability or difference in calibration when comparing alternative models.<sup>19</sup>

Models from high-income settings such as Clinical Risk index for Babies include predictors that are not available in our data set, which is likely the best in Kenya at the time.<sup>7 20</sup> Houweling *et al*<sup>21</sup> published models for neonatal mortality from India, Nepal and Bangladesh. Besides the Simplified Age–Weight–Sex score and the NMR-2000, these appear to be the only models applicable to LMIC settings published to date.<sup>7 22 23</sup> They used population surveillance data, which differ from NETS and SENSS, which used routine hospital data.<sup>12</sup> In addition, time of death encompassed 0–28 days in contrast to NETS and SENSS, where only deaths occurring in the neonatal unit (most occurring in the first week) were considered. These are potentially applicable to the Kenyan context but are prediction of neonatal deaths in the general population rather than in-hospital. The Simplified Age–Weight–Sex model was developed for neonates with gestational age of  $\leq 33$  weeks and a birth weight  $\leq 1500$  g, precluding further comparison with NETS and SENSS.<sup>22</sup> Similarly, the NMR-2000 is restricted to a birth weight of  $\leq 2000$  g, age less than 6 hours and includes pulse oximetry, which may not be routinely available.<sup>23 24</sup>

NETS and SENSS at derivation had a good ability to distinguish between neonates who died in hospital and those who did not (c-statistics of 0.92, 95% CI 0.90 to 0.93 for NETS and 0.91, 95% CI 0.89 to 0.93 for SENSS) (figure 1). Houweling *et al* derived models based on four time points: at the start of pregnancy, start of delivery, after birth and start of delivery (including multiple delivery). Only the after-birth model demonstrated an area under the curve (AUC) greater than 0.8 (pooled average 0.83, 95% CI 0.80 to 0.89).<sup>21</sup>

At internal validation, the optimism adjusted c-statistic was 0.916 (NETS) and 0.901 (SENSS) compared with the original 0.918 and 0.902, respectively. Similarly, there was minimal change in the respective calibration intercept and slope (table 2). In our systematic review on neonatal treatment intensity scores, none of the 10 studies included performed internal validation.<sup>7</sup> Houweling and colleagues conducted cross-validation and again only the after-birth model had an AUC greater than 0.8 (pooled average 0.83, 95% CI 0.79 to 0.86).<sup>21</sup>

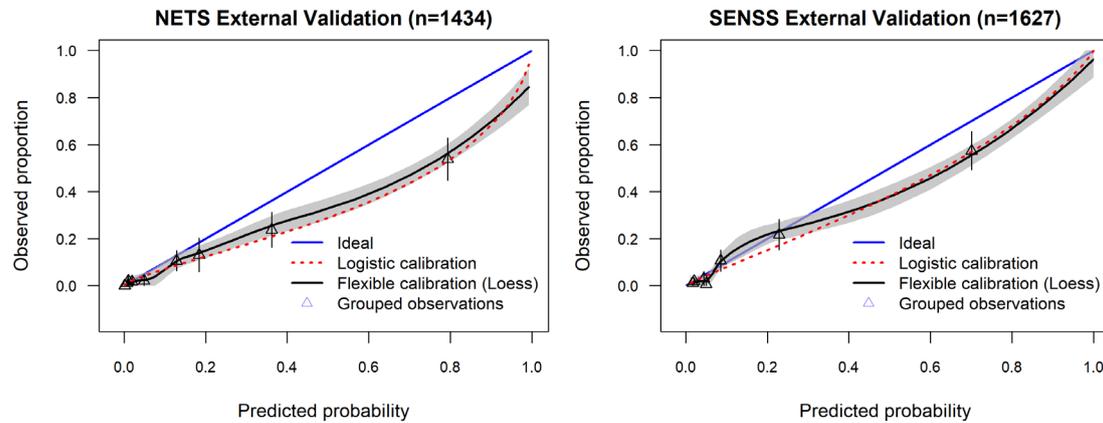
The most useful aspect of prediction model performance is in an external population to assess generalisability (external

**Table 2** Evaluation of the NETS and SENSS models for optimism after bootstrapping

Parameter	Calibration				Discrimination*	
	Intercept		Slope			
	NETS	SENSS	NETS	SENSS	NETS	SENSS
Original	0	0	1	1	0.918	0.902
Corrected	$-0.062$	$-0.029$	0.979	0.986	0.916	0.901

\*c-statistic of the logistic regression model.

NETS, Neonatal Essential Treatment Score; SENSS, Score for Essential Symptoms and Signs.



**Figure 2** Calibration curves for NETS and SENSS in the external validation data sets. NETS, Neonatal Essential Treatment Score; SENSS, Score for Essential Neonatal Symptoms and Signs.

validation).<sup>17</sup> After temporal external validation, SENSS had a better calibration slope (0.90, 95% CI 0.78 to 1.01) than NETS (0.76, 95% CI 0.65 to 0.87). This means that across the range of predicted risks, the SENSS predictions were closer to the observed outcomes in the external validation population. However, there was deterioration in performance at external validation for both models. Three factors may explain this: (1) differences in definition and measurement of predictors and outcomes, (2) differences in the case mix and (3) inclusion of fewer patients in validation compared with derivation.<sup>25</sup>

The prescription of treatments (NETS predictors) and recognition of signs and symptoms of severe illness (SENSS predictors) are different clinical skills. It is possible a systematic difference in practices as junior clinicians changed contributed to the difference in model performance. In addition, the proportion of deaths per predictor for both models were less in the external validation data set (online supplemental tables 6 and 7). However, the magnitude of these differences varied between NETS and SENSS data sets, which might have contributed to the differences in model calibration observed at external validation. Finally, the validation data sets were smaller and spanned a shorter period (7 months vs 18 months). This might have resulted in case-mix difference due to random variation.

### Limitations

Variation in treatments may also be influenced by time, availability of resources and level of care (eg, intensive care).<sup>7</sup> However, the use of standard treatment guidelines reduces this variation, and the model may be updated as practice changes.<sup>6,7,9</sup> The missingness of SENSS predictors necessitated multiple imputation. It is hard to eliminate missingness in large observational data sets, and imputation provides an approach to manage this while trying to continuously improve data quality through interventions like audit and feedback.<sup>12,26</sup> Temporal external validation of NETS and SENSS does not translate to generalisability in other neonatal units as this can only be assessed by externally validating the model with such external data (geographical external validation). However, there exists within Kenya the potential to conduct geographical external validation using routine data from other neonatal units.<sup>27</sup> After validation, application to individual patients should be preceded by impact studies (preferably cluster randomised trials) that evaluate effectiveness and safety.<sup>6</sup>

### Implications

Using routine neonatal data in a low-resource setting, we found that it may be possible to predict in-hospital mortality using either treatments or signs and symptoms. Using treatments as predictors (NETS) had the advantage of availability of data in contrast to the signs and symptoms (SENSS), which required imputation. Prediction of in-hospital mortality can be used for case-mix adjustment and potentially to inform treatment decisions for individual patients (such as referral to higher-level facilities based on risk of death). Case-mix adjustment of in-hospital mortality is the most important as this is a vital component in the exploration of health system performance at scale in the delivery of care for small and sick neonates. The NETS model may be more suited for case-mix adjustment, given our experience that that it is much easier to identify treatment data than clinical signs where data can only be obtained from retrospective data.

**Twitter** Gary Collins @GScollins

**Acknowledgements** The authors thank and acknowledge the input of the data management team of the Clinical Information Network at the KEMRI-Wellcome Trust Research Programme, Nairobi. We also acknowledge the support and contributions of the Kenya Paediatric Association plus the staff and management of Pumwani Maternity Hospital, Nairobi, Kenya.

**Contributors** JA, GC, JAB and ME conceived the study. JA designed the data collection platform and implemented it with BM, MW and CM. JA analysed the data with support from GC, JAB and ME. JA wrote the manuscript with support from GC, JAB and ME. All authors revised the manuscript, approved the final version and agreed to be accountable for the findings.

**Funding** This work was primarily supported by a Wellcome Trust Senior Fellowship (#097170) awarded to ME. Additional support was provided by Health Systems Research Initiative joint grant provided by the Department for International Development, UK, Economic and Social Research Council, Medical Research Council and Wellcome Trust (grant number MR/M015386/1 and a Wellcome Trust core grant awarded to the KEMRI-Wellcome Trust Research Programme (#092654).

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** Ethical approval was provided by the KEMRI Scientific and Ethical Review Committee (SERU 3459)

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. The source data are owned by the Kenyan Ministry of Health, County Governments and as the data might be used to deidentify hospitals the study authors are not permitted to share the source data directly. Users who wish to reuse the source data can make a request initially through the KEMRI-Wellcome Trust Research Programme data governance committee. This committee will supply contact information for the KEMRI Scientific and Ethical Review unit, County Governments and individual hospitals as appropriate. The KEMRI-Wellcome Trust Research

Programme data governance committee can be contacted on: [dgc@kemri-wellcome.org](mailto:dgc@kemri-wellcome.org).

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iDs

Jalemba Aluvaala <http://orcid.org/0000-0002-0851-3711>

Gary Collins <http://orcid.org/0000-0002-2772-2316>

James Alexander Berkley <http://orcid.org/0000-0002-1236-849X>

Mike English <http://orcid.org/0000-0002-7427-0826>

#### REFERENCES

- United Nations/General Assembly. *Transforming our world: the 2030 agenda for sustainable development*, 2015.
- 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.
- Hug L, Alexander M, You D, et al. National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: a systematic analysis. *Lancet Glob Health* 2019;7:e710–20.
- WHO. *Systems thinking for health systems strengthening*, 2009.
- Kihuba E, Gathara D, Mwinga S, et al. Assessing the ability of health information systems in hospitals to support evidence-informed decisions in Kenya. *Glob Health Action* 2014;7:7.
- Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis research strategy (progress) 3: prognostic model research. *PLoS Med* 2013;10:e1001381.
- Aluvaala J, Collins GS, Maina M, et al. A systematic review of neonatal treatment intensity scores and their potential application in low-resource setting hospitals for predicting mortality, morbidity and estimating resource use. *Syst Rev* 2017;6:248.
- Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55–63.
- Ministry of Health, Government of Kenya. *Basic paediatric protocols for ages up to 5 years*, 2016.
- Tuti T, Bitok M, Paton C, et al. Innovating to enhance clinical data management using non-commercial and open source solutions across a multi-center network supporting inpatient paediatric care and research in Kenya. *J Am Med Inform Assoc* 2016;23:184–92.
- Collins GS, Moons KGM. Comparing risk prediction models. *BMJ* 2012;344:e3186.
- Maina M, Aluvaala J, Mwaniki P, et al. Using a common data platform to facilitate audit and feedback on the quality of hospital care provided to sick newborns in Kenya. *BMJ Glob Health* 2018;3:e001027.
- Aluvaala J, Nyamai R, Were F, et al. Assessment of neonatal care in clinical training facilities in Kenya. *Arch Dis Child* 2015;100:42–7.
- Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol* 2016;76:175–82.
- Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. New York: Springer, 2009.
- Harrell Jr FE, with contributions from Charles Dupont and many others. *Hmisc: Harrell Miscellaneous*. R package version 4.0-3. 2017.
- 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.
- Opiyo N, English M. What clinical signs best identify severe illness in young infants aged 0–59 days in developing countries? A systematic review. *Arch Dis Child* 2011;96:1052–9.
- Dorling JS, Field DJ. Value and validity of neonatal disease severity scoring systems. *Arch Dis Child Fetal Neonatal Ed* 2008;93:F80–2.
- Aluvaala J, Collins GS, Maina B, et al. Competing risk survival analysis of time to in-hospital death or discharge in a large urban neonatal unit in Kenya. *Wellcome Open Res* 2019;4:96.
- Houweling TAJ, van Klaveren D, Das S, et al. A prediction model for neonatal mortality in low- and middle-income countries: an analysis of data from population surveillance sites in India, Nepal and Bangladesh. *Int J Epidemiol* 2019;48:186–98.
- 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.
- Medvedev MM, Brotherton H, Gai A, 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. *Lancet Child Adolesc Health* 2020;4:299–311.
- Enoch AJ, English M, Clinical Information Network, et al. Variability in the use of pulse oximeters with children in Kenyan hospitals: a mixed-methods analysis. *PLoS Med* 2019;16:e1002987.
- Toll DB, Janssen KJM, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61:1085–94.
- Tuti T, Bitok M, Malla L, et al. Improving documentation of clinical care within a clinical information network: an essential initial step in efforts to understand and improve care in Kenyan hospitals. *BMJ Glob Health* 2016;1:e000028.
- English M, Ayieko P, Nyamai R, et al. What do we think we are doing? How might a clinical information network be promoting implementation of recommended paediatric care practices in Kenyan hospitals? *Health Res Policy Syst* 2017;15:4.

**Supplementary Table 1:** Essential treatments prescribed at admission that were included as predictors

Variable name	Description	Definition	
		Yes	No*
<b>Intravenous antibiotics</b>	Penicillin AND gentamicin are first line antibiotics prescribed together for possible Severe Bacterial Infection.	Prescribed	Not prescribed
<b>Intravenous fluids</b>	Any prescription of an infusion of intravenous fluid e.g. 10% Dextrose	Prescribed	Not prescribed
<b>Enteral feeding</b>	Expressed breast milk or formula fed through nasogastric tube or cup	Prescribed	Not prescribed
<b>Oxygen</b>	Supplemental oxygen provided by any method e.g. nasal cannula	Prescribed	Not prescribed
<b>Phenobarbital</b>	Prescription of phenobarbital	Prescribed	Not prescribed

\*Reference category in the model

**Supplementary Table 2:** Symptoms and signs of severe illness considered as candidate predictors

Variable name	Description†	Definition	
		Yes	No*
<b>Difficulty feeding</b>	Symptom: reported history of difficulty feeding	Present	Absent
<b>Convulsions</b>	Symptom: reported episode(s) of convulsions	Present	Absent
<b>Indrawing</b>	Sign: severe (deep) lower chest indrawing	Present	Absent
<b>Central cyanosis</b>	Sign: bluish or greyish discolouration of the tongue	Present	Absent
<b>Floppy/inability to suck</b>	Sign: measure of altered consciousness in young infants (age 0-59 days)	Present	Absent

\*Documented that the patient did not have the symptom or sign present

† Adapted from Opiyo N, English M. *What clinical signs best identify severe illness in young infants aged 0-59 days in developing countries? A systematic review. Archives of Disease in Childhood. 2011;96(11):1052-9.*

**Supplementary Table 3:** Clinical characteristics of neonates included in the model development data set

Characteristic	Random sample(n=5427)		NETS* development data set(n=4840)		Excluded from NETS (n=587) <sup>†</sup>	
	n	%	n	%	n	%
<b>1. Sex</b>						
Male	2937	<b>54</b>	2605	<b>54</b>	332	<b>57</b>
Missing	13	<b>0.2</b>	12	<b>0.3</b>	1	<b>0.2</b>
<b>2. Birthweight (kg)</b>						
<1	32	<b>0.6</b>	31	<b>0.6</b>	1	<b>0.2</b>
1-<1.5	136	<b>3</b>	115	<b>2</b>	21	<b>4</b>
1.5-<2.5	1180	<b>22</b>	1043	<b>22</b>	137	<b>23</b>
2.5-4	3841	<b>71</b>	3438	<b>71</b>	403	<b>69</b>
>4	229	<b>4</b>	204	<b>4</b>	2	<b>4</b>
Missing	9	<b>0.2</b>	9	<b>0.2</b>	0	
<b>3. Mode of delivery</b>						
Spontaneous vaginal	3011	<b>57</b>	2697	<b>57</b>	314	<b>55</b>
Assisted vaginal	3	<b>0.1</b>	1	<b>0.02</b>	2	<b>0.3</b>
Breech	43	<b>1</b>	40	<b>1</b>	3	<b>1</b>
Caesarean section	2212	<b>42</b>	1999	<b>42</b>	255	<b>44</b>
Missing	158	<b>3</b>	145	<b>3</b>	13	<b>2</b>
<b>4. Outborn<sup>‡</sup></b>						
Yes	123	<b>2</b>	107	<b>2</b>	16	<b>3</b>
Missing	0		0	-	0	-
<b>5. Apgar score (5min)</b>						
0-3	116	<b>3</b>	112	<b>2</b>	-	-
4-6	602	<b>13</b>	593	<b>13</b>	-	-
7-10	3992	<b>85</b>	3918	<b>85</b>	-	-
Missing	717	<b>11</b>	217	<b>4</b>	500	<b>86</b>
<b>7. HIV exposure</b>						
Exposed	319	<b>6</b>	287	<b>6</b>	32	<b>6</b>
Missing	305	<b>6</b>	277	<b>6</b>	28	<b>5</b>
<b>8. Outcome</b>						
Discharged alive	4838	<b>89</b>	4318	<b>89</b>	520	<b>89</b>
Dead	508	<b>9</b>	447	<b>9</b>	61	<b>10</b>
Referred	62	<b>1</b>	56	<b>1</b>	6	<b>1</b>
Missing	19	<b>0.4</b>	19	<b>0.4</b>	0	-

\* NETS: Neonatal Essential Treatment Score

<sup>†</sup> Excluded due to missing treatment sheets in the patient file<sup>‡</sup> Outborn refers to neonates admitted to the unit having been born either in another facility, at home or on the way to hospital

**Supplementary Table 4:** Comparison of patients included in NETS derivation and external validation

Characteristic	NETS* derivation (n=4840)		NETS* external validation (n=1443)	
	n	%	n	%
<b>1. Sex</b>				
Male	2605	54	850	59
Missing	12	0.3	2	0.1
<b>2. Birthweight (kg)</b>				
<1	31	0.6	10	1
1-<1.5	115	2	40	3
1.5-<2.5	1043	22	316	22
2.5-4	3438	71	1002	69
>4	204	4	74	5
Missing	9	0.2	1	0.1
<b>3. Mode of delivery</b>				
Spontaneous vaginal	2697	57	897	63
Assisted vaginal	1	0.02	0	0
Breech	40	1	19	1
Caesarean section	1999	42	509	36
Missing	145	3	18	1
<b>4. Outborn ‡</b>				
Yes	107	2	57	4
Missing	0	0	0	0
<b>5. Apgar score (5)</b>				
0-3	112	2	28	2
4-6	593	13	192	14
7-10	3918	85	1122	84
Missing	217	4	69	5
<b>7. HIV exposure</b>				
Exposed	287	6	74	5
Missing	277	6	80	6
<b>8. Outcome</b>				
Discharged alive	4318	89	1262	88
Dead	447	9	137	10
Referred	56	1	38	3
Missing	19	0.4	6	0.4
<b>9. Admission Diagnosis</b>				
Birth Asphyxia	2011	42	494	34
Preterm/LBW	979	20.1	257	18
Neonatal sepsis	495	10	172	12
Respiratory D. Syndrome	331	7	103	7
Neonatal jaundice	272	6	141	10
Others	513	11	332	23

\* NETS: Neonatal Essential Treatment Score

**Supplementary Table 5:** Distribution of predictors and missingness in the SENSS datasets

Characteristic	Category	SENSS* Derivation		External Validation	
		N= 5427		N= 1627	
		n	%	n	%
<b>1.Sex</b>	Male	2937	54	961	58
	Missing	13	0.2	2	0.1
<b>2.Birthweight(kg)</b>	<1	32	0.5	10	0.6
	1-<1.5	136	3	45	3
	1.5-<2.5	1180	22	361	22
	2.5-4	3841	71	1125	69
	>4	229	4	85	5
	Missing	9	0.2	9	0.2
<b>3.Difficulty feeding</b>	Yes	225	5	134	9
	Missing	575	11	199	12
<b>4.Convulsions</b>	Yes	29	0.5	25	2
	Missing	571	11	199	12
<b>5.Central cyanosis</b>	Yes	222	5	64	5
	Missing	570	11	216	13
<b>6.Severe indrawing</b>	Yes	429	9	62	5
	Missing	620	11	234	14
<b>7.Floppy/unable to suck</b>	Yes	530	11	359	26
	Missing	876	16	222	13
<b>8.Outcome</b>	Dead	508	9	151	9
	Missing	19	0.4	6	0.4

\*SENSS: Score for Essential Neonatal Symptoms and Signs

**Supplementary Table 6:** Distribution of NETS predictors by in-hospital mortality

Predictors	NETS <sup>a</sup> External validation(n=1443)						NETS <sup>a</sup> Development (n=4840)					
	Prescribed		Dead <sup>†</sup>		Missing <sup>‡</sup>		Prescribed		Dead <sup>†</sup>		Missing <sup>‡</sup>	
	n	%	n	%	n	%	n	%	n	%	n	%
Intravenous antibiotics	780	54	123	16	1	0.1	1892	39	383	20	4	0.2
Intravenous fluids	495	34	116	23	1	0.2	869	18	295	34	1	0.1
Enteral feeds	598	41	7	1	2	0.3	1381	29	11	1	8	0.6
Oxygen	105	7	105	24	0	0	295	6	120	41	0	0
Phenobarbital	72	5	21	25	1	1.5	69	1	42	61	0	0
None	196	14	9	5	3	1.5	834	17	18	2	8	1

\* NETS: Neonatal Essential Treatment Score

† Denominators obtained by subtracting the number missing outcome data for each predictor

‡ Missing outcome data for each predictor

**Supplementary Table 7:** Distribution of SENSS predictors by in-hospital mortality

Predictor	SENSS Derivation				External validation			
	(n=5427) <sup>†</sup>				(n=1627) <sup>‡</sup>			
	Present		Dead <sup>§</sup>		Present		Dead <sup>§</sup>	
	n	%	n	%	n	%	n	%
<b>1.Difficulty feeding</b>	264	5	110	42	146	9	42	29
<b>2.Convulsions</b>	38	1	17	45	29	2	7	24
<b>3.Central cyanosis</b>	255	5	141	56	79	5	35	44
<b>4.Severe indrawing</b>	494	9	266	54	69	4	40	59
<b>5.Floppy/unable to suck</b>	647	12	297	46	413	25	127	31

\*SENSS: Score for Essential Neonatal Symptoms and Signs

† One of the 31 multiply imputed data sets

‡ One of the 23 multiply imputed data sets

§ Number and proportion dead amongst those with the predictor present

**Supplementary Table 8:** SENSS predictors in the original and imputed data sets

Characteristic	Category	SENSS <sup>†</sup> Derivation (N <sup>†</sup> =5427)				External validation (N <sup>†</sup> =1627)			
		Original		Imputed <sup>‡</sup>		Original		Imputed <sup>‡</sup>	
		n	%	n	%	n	%	n	%
<b>1. Sex</b>	Male	2937	<b>54</b>	2942	<b>54</b>	961	<b>58</b>	962	<b>59</b>
<b>2. Birthweight(kg)</b>	<1	32	<b>0.5</b>	32	<b>0.6</b>	10	<b>0.6</b>	10	<b>0.6</b>
	1-<1.5	136	<b>3</b>	136	<b>3</b>	45	<b>3</b>	45	<b>3</b>
	1.5-<2.5	1180	<b>22</b>	1182	<b>22</b>	361	<b>22</b>	361	<b>22</b>
	2.5-4	3841	<b>71</b>	3848	<b>71</b>	1125	<b>69</b>	1126	<b>69</b>
	>4	229	<b>4</b>	229	<b>4</b>	85	<b>5</b>	85	<b>5</b>
<b>3. Difficulty feeding</b>	Yes	225	<b>5</b>	264	<b>5</b>	134	<b>9</b>	146	<b>9</b>
<b>4. Convulsions</b>	Yes	29	<b>0.5</b>	38	<b>0.5</b>	25	<b>2</b>	29	<b>2</b>
<b>5. Central cyanosis</b>	Yes	222	<b>5</b>	255	<b>5</b>	64	<b>5</b>	79	<b>5</b>
<b>6. Severe indrawing</b>	Yes	429	<b>9</b>	494	<b>9</b>	62	<b>5</b>	69	<b>4</b>
<b>7. Floppy/unable to suck</b>	Yes	530	<b>11</b>	647	<b>11</b>	359	<b>26</b>	413	<b>25</b>
<b>8. Outcome</b>	Dead	508	<b>9</b>	509	<b>9</b>	151	<b>9</b>	151	<b>9</b>

\* SENSS: Score for Essential Neonatal Symptoms and Signs

† Total number of patients in the data set

‡ Imputation models included auxiliary variables: mode of delivery, place of delivery, HIV exposure status and documentation of Apgar score. Outcome was also included. Multiple imputation using the chained equation approach was implemented using logistic regression for binary variables while categorical variables were imputed with polytomous logistic regression. There were no continuous variables.