

ENDGAMES

STATISTICAL QUESTION

Prognostic scores

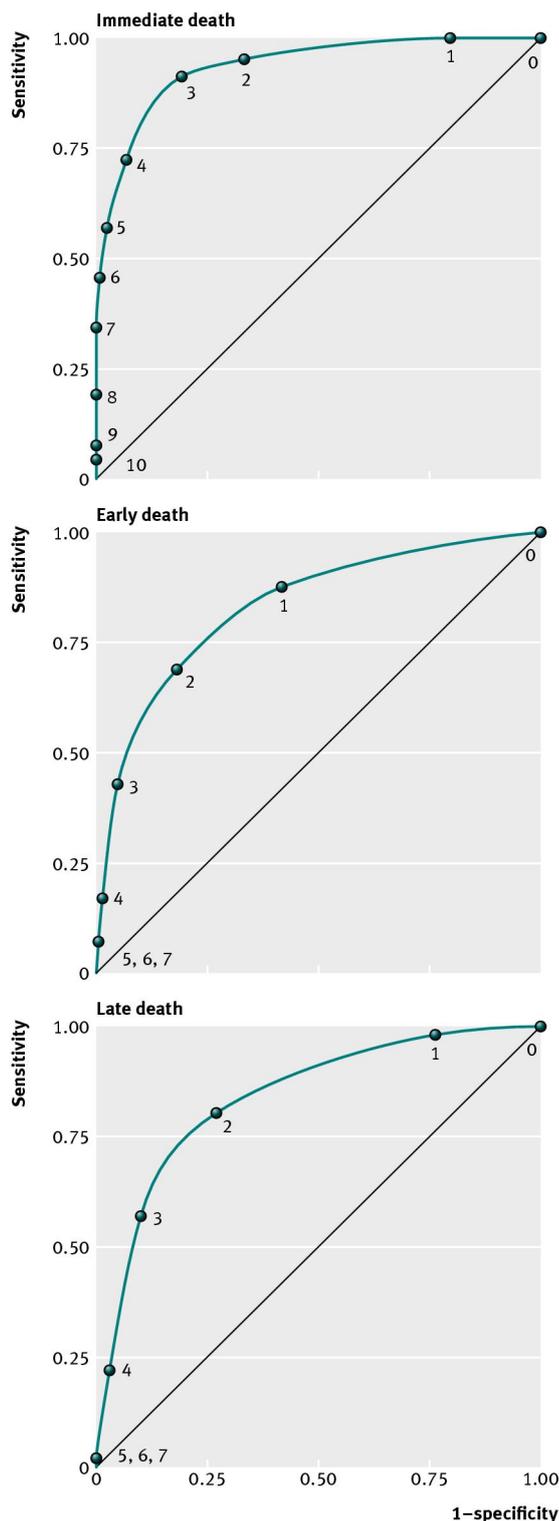
Philip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, London, UK

Researchers investigated clinical indicators of immediate, early, and late mortality in children at admission to hospital. A prospective cohort study design was used. Participants were 8091 children aged more than 90 days admitted to a subSaharan district hospital in Kenya between 1 July 1998 and 30 June 2000. Children were excluded if admitted for trauma or elective procedures. Of the 8091 children admitted 436 (5%) died—60 (14%) died immediately, 193 (44%) died early, and 183 (42%) died late.¹

Separate prognostic models were developed for immediate death (within 4 h of admission), early death (within 4-48 h), and late death (after 48 h). The models were developed from clinical indicators collected prospectively for children in the cohort on admission and at death or discharge.

The prognostic models were validated using data collected from a further cohort of 4802 children aged more than 90 days, admitted to the same hospital between 1 July 2000 and 30 June 2001. Children admitted for trauma and elective admissions were excluded. Of the 4802 children admitted in the validation cohort 222 (5%) died—26 (12%) immediately, 88 (40%) early, and 108 (49%) late. For each child a prognostic score for predicting immediate, early, and late death was derived by summing the total number of clinical indicators present. The performance of the prognostic scores was assessed by receiver operating characteristic curves (figure). The areas under the receiver operating characteristic curves were 0.93 (95% confidence interval 0.92 to 0.94) for immediate, 0.82 (0.80 to 0.83) for early, and 0.82 (0.81 to 0.84) for late deaths.



Receiver operating characteristic curves for the prognostic scores for immediate, early, and late death (n=4802). Prognostic scores were categorised as high risk (“positive”) of death if a certain number, or more, clinical indicators were present, otherwise they were categorised as low risk (“negative”). Numbers on curves refer to the cut-off scores between high and low risk categories

The researchers concluded that in children admitted to a sub-Saharan hospital, a small number of simple clinical signs discriminated between those children who died and those who survived after admission to hospital.

Which of the following statements, if any, are true?

- a) The process of validation used is described as external validation
- b) Each clinical indicator was assumed to have equal weight in predicting mortality
- c) Sensitivity was the proportion of children who died and were correctly identified at high risk by their prognostic scores
- d) The prognostic model that best discriminated between death and survival in the validation cohort was that for immediate death

Answers

Statements *b*, *c*, and *d* are true, whereas *a* is false.

The aim of the study was to investigate clinical indicators that predicted death in children after admission to hospital. A separate prognostic model was developed for immediate, early, and late deaths. The scoring systems were developed from 14 clinical indicators collected prospectively for 8091 children aged more than 90 days admitted to a sub-Saharan district hospital in Kenya between 1 July 1998 and 30 June 2000. Further details of the clinical indicators are given in the original article.¹ Not all of the clinical indicators recorded were used in the scoring systems because not all of them helped predict death. Furthermore, the number of clinical indicators that contributed to each scoring system differed, with 10 for immediate deaths and seven for both early and late deaths.

To show that the prognostic models were valuable in predicting mortality, it was not sufficient to show that they successfully predicted death in the original cohort of children admitted between 1 July 1998 and 30 June 2000. It was essential that the models were validated. Validation involved assessing the performance of the prognostic models in a different cohort of children admitted to hospital. This required comparing the observed and predicted death rates for each prognostic model (referred to as calibration), plus quantifying the models’ abilities to distinguish between children who died and those who survived (referred to as discrimination).

Validation is described as internal, temporal, or external depending on the cohort that is used to assess performance. Temporal validation was used in the study above (*a* is false). This involved evaluating the performance of the prognostic models in a cohort of children admitted to the same hospital after the original cohort that had been used to develop the models. Validation was performed using a cohort of 4802 children admitted in the year (1 July 2000 to 30 June 2001) after the two year period in which the original (developmental) cohort had been collated. Internal validation involves splitting a cohort of children admitted to the hospital into two parts—typically in a ratio of 2:1—and then developing the prognostic models on the first group and validating performance on the second group. In the example above, the process of validation can be regarded as temporal or internal validation because, in effect, a cohort was collected over three years and then split temporally. External validation would involve assessing the performance of the prognostic models on a cohort of children admitted to a different hospital, thereby testing the generalisability of the prognostic models.

For each child in the validation cohort, the prognostic models yielded scores on admission to hospital that predicted the risk of immediate, early, and late death. The score from each model was derived by summing the number of clinical indicators present. Counting the number of clinical indicators present, as was done in this study, is the simplest approach to deriving

prognostic scores. This approach assumes that each clinical indicator contributes equally to the outcome of death (b is true), which was probably not true clinically.

Children were stratified by their prognostic scores. If a certain number, or more, of indicators were present then a child was classified as high risk (predicting death), otherwise the child was classified as low risk (predicting survival). The categorised prognostic scores were then compared against the children's outcome (death or survival) after admission. Assessment of the performance of the prognostic scores in predicting death is similar to the assessment of the performance of a screening test against a diagnostic test. Screening tests have been described in a previous question.²

Prognostic scores are rarely 100% accurate in predicting death. A child who died after admission may have been predicted to be at low risk and therefore survive (a false negative), whereas one who survived after admission may have been predicted to be at high risk of death (false positive). As part of the process of validation, to assess the discriminative ability of the prognostic models the optimal prognostic cut-off scores (number of clinical indicators present) in predicting death were investigated. The optimal cut-off score would be the one that best discriminates between children who died and those who survived.

The optimal cut-off between a low and high risk prognostic score was investigated by successively taking the number of clinical indicators from zero up to the maximum included in the prognostic model; all scores smaller than the cut-off score were categorised as low risk, whereas scores equal or greater were categorised as high risk. For each cut-off score the values of sensitivity and specificity were derived. Sensitivity was the proportion of children who died and were correctly identified at high risk by their prognostic scores (c is true). Specificity is the proportion of children who survived after admission and were correctly identified at low risk by their prognostic scores.

The table¹ shows the distribution of the immediate, early, and late mortality prognostic scores by outcome (death and survival) for the validation cohort of 4802 children admitted to hospital between 1 July 2000 and 30 June 2001. For each prognostic score, sensitivity and specificity were derived as described above.

The performance of each prognostic model in predicting mortality was investigated by constructing receiver operating characteristic curves (figure) using the validation cohort. A receiver operating characteristic curve, described in a previous question,³ is the plot of sensitivity against (1 minus specificity) for each cut-off score. The curve allows the association between sensitivity and specificity to be examined as the cut-off point between a low and high risk prognostic score changes. The value

of (1 minus specificity) is the proportion of children who survived after admission and were identified incorrectly as high risk by the prognostic model. The value of (1 minus specificity) is referred to as the "false positive rate."

If a cut-off point between a low and high risk prognostic score predicted mortality with 100% accuracy, then both sensitivity and specificity would equal 1, and the false positive rate would equal 0. In that case the receiver operating characteristic curve would pass through the top left hand corner of the figure. The curve would start at the origin, go vertically up the y axis to a sensitivity of 1.0, and then horizontally across to a false positive rate of 1.0. A prognostic model will rarely be 100% accurate. Therefore, the closer the receiver operating characteristic curve is to the upper left corner, the higher the overall accuracy of the prognostic model across all potential cut-off points. Although the researchers did not suggest the optimal cut-off value for each prognostic model, the one closest to the top left hand corner is usually chosen.

The overall accuracy of each of the three prognostic models for predicting mortality can be compared by looking at the area beneath each curve. The model with the curve closest to the top left corner—and therefore the greatest area beneath the curve—is the best at discriminating between those children who died and those who survived. The prognostic model for immediate death had the greatest area under the curve, which suggests that it is the best model for discriminating between those children who died and those who survived in the validation cohort (d is true). The area under the curve, sometimes referred to as the c index, may be used to summarise discrimination in the validation process. The discriminative ability of the prognostic models should be assessed by comparing the areas under the receiver operating curves for the validation cohort with those of the developmental cohort. However, the researchers did not provide these data.

The researchers did not comment on the calibration of the prognostic models in the validation cohort. However, it was concluded that the clinical indicators chosen may be useful in predicting death in children at admission to hospital. Nonetheless, further research is needed before the prognostic models could be used in risk assessment.

Competing interests: None declared.

- 1 Berkley JA, Ross A, Mwangi I, Osier FHA, Mohammed M, Shebbe M, et al. Prognostic indicators of early and late death in children admitted to district hospital in Kenya: cohort study. *BMJ* 2003;326:361.
- 2 Sedgwick P. Screening tests: indices of performance. *BMJ* 2011;342:d3763.
- 3 Sedgwick P. Receiver operating characteristic curves. *BMJ* 2011;343:d4302.

Cite this as: *BMJ* 2014;348:g282

© BMJ Publishing Group Ltd 2014

Table

Table 1 | Distribution of prognostic scores and outcome in the validation cohort of 4802 children admitted to hospital between 1 July 2000 and 30 June 2001

Score	Immediate death					Early death*					Late death*				
	Admitted (n)	Died (n)	Survived (n)	Sensitivity	Specificity	Admitted (n)	Died (n)	Survived (n)	Sensitivity	Specificity	Admitted (n)	Died (n)	Survived (n)	Sensitivity	Specificity
0	698	0	698	1.00	0.00	2779	11	2768	1.00	0.00	15	0	15	1.00	0.00
1	2196	1	2195	1.00	0.15	1101	16	1085	0.88	0.59	1081	2	1079	1.00	0.00
2	657	1	656	0.96	0.64	668	23	645	0.69	0.82	2310	18	2292	0.98	0.23
3	612	5	607	0.92	0.79	183	23	160	0.43	0.95	869	26	843	0.81	0.72
4	217	4	213	0.73	0.92	55	10	45	0.17	0.99	347	38	309	0.57	0.90
5	72	3	69	0.58	0.97	15	5	10	0.06	1.00	170	22	148	0.22	0.97
6	34	3	31	0.46	0.98	1	0	1	0.00	1.00	10	2	8	0.02	1.00
7	31	4	27	0.35	0.99	0	0	0	0.00	1.00	0	0	0	0.00	1.00
8	12	3	9	0.19	1.00	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
9	2	1	1	0.08	1.00	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	1	1	0	0.04	1.00	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Total	4532†	26	4506			4802	88	4714			4802	108	4694		

*NA=not applicable.

†Prognostic scores for immediate death were missing for 270 children owing to incomplete records on clinical indicators.