OPEN ACCESS

University of BRISTOL

Gibbison, B., & Reeves, B. C. (2019). Forecasting HOPE: Risk prediction in rare events. *Resuscitation*, *139*, 361-362. https://doi.org/10.1016/j.resuscitation.2019.04.014

Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
10.1016/j.resuscitation.2019.04.014

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at https://www.sciencedirect.com/science/article/pii/S0300957219301285 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

Forecasting HOPE: Risk prediction in rare events

Ben Gibbison[1,2*], Barnaby C Reeves[2, 3]

1. Dept. of Cardiac Anaesthesia and Intensive Care. Bristol Heart Institute. Bristol. UK
2. NIHR Bristol Biomedical Research Centre, University of Bristol. UK.
3. Clinical Trials and Evaluation Unit, Bristol Trials Centre. University of Bristol. UK

ben.gibbison@bristol.ac.uk

* Corresponding author:
Dept of Anaesthesia
Level 7 Queens Building
Bristol Royal Infirmary
Upper Maudlin St
Bristol
BS2 8HW
Tel: +44 (0)7931 568135

Predicting, or *forecasting,* outcomes is difficult and is central to the role of physicians in making clinical management decisions. The integration of knowledge, skills and experience to predict who will respond to a treatment, survive an operation or extract useful benefit from Intensive Care is something many readers of this journal will do daily. This becomes more difficult when both the event and the treatment are rare. In the linked article in this issue of *Resuscitation*[1]*,* Pasquier and colleagues test a tool, the Hypothermia Outcome Prediction after Extracorporeal Life Support (HOPE) score, for predicting outcomes in patients presenting with hypothermic cardiac arrest and treated by extra-corporeal life support (ECLS). True hypothermic cardiac arrest is relatively rare and hypothermic cardiac arrest with access to ECLS is even less frequent. The authors have made substantial progress widening our collective experience and formulating this into an easy to apply tool. Moreover, the way in which the tool has been generated provides some interesting lessons about predicting outcomes from rare events generally *and* more specifically in the context of hypothermic cardiac arrest.

The authors derived the HOPE Score from a previous cohort of patients[2]. This cohort comprised 237 patients from published studies and 49 patients in the care records of two Swiss hospitals. The HOPE score probability of survival was derived from a multiple logistic regression of a larger number of variables and combines information for: age, sex, mechanism of cooling, core temperature, serum K^+ level and duration of CPR. The accuracy of prediction of the score when applied to patients not in the cohort has been uncertain because, until now, the score had not been externally validated. In *this* paper, the authors attempt to provide this validation.

The validation cohort comprised 122 patients, assembled by updating their search of published studies (after excluding patients included in the derivation study) and from the care records of three additional centres. The threshold for offering or not-offering ECLS (≥0.10 probability of survival) represents a decision point that balances the two opposing consequences of misclassifications – providing futile ECLS to a patient with no chance of recovery or failing to provide ECLS to a patient with the potential to recover. These consequences carry different weights and the threshold was chosen to minimise the specific risk of providing futile ECLS and not to minimise *all* misclassifications. The need for a threshold arises from the consequences of futile treatment: poor long term neurological outcome, emotional cost to the family, as well as the direct and indirect economic costs of ECLS.

It is rare to be able to predict outcomes perfectly by applying one threshold. It is therefore unsurprising that there can be instances when the prediction appears wrong. These arise where the underlying distributions of scores associated with the two outcomes overlap *or* when applying the prediction rule outside the precise context in which it was developed and validated. Thus caution should be exercised when using the HOPE score in a deterministic fashion around this threshold value of 0.10. The instance of the young patient discussed by the authors might also make us cautious about using the HOPE score to predict outcome in children.

The need for derivation and validation samples arises because a rule developed on a particular sample will be based in part on *true* underlying associations between predictors and the outcome of interest and in part on *spurious* associations (due to chance) between

other predictors and the outcome. This leads to the performance of the rule being overestimated in the derivation sample. A validation sample is used to estimate the performance of the rule in an independent dataset. However, identifying an independent dataset can be challenging for two reasons, both of which arise from failing to specify the context in which the validated rule will be applied. Firstly, external factors that differ between the derivation and validation samples can influence the performance of the rule and secondly, if the validation sample is not sufficiently independent, then the performance will still be overestimated. Judging whether a validation sample eliminates spurious association is very difficult. This challenge can be illustrated by considering a dataset that is large enough to divide into two subsets – one for derivation and one for validation. Randomly allocating records to one or other subset might appear ideal. However, this randomises spurious as well as true associations – especially if the data were collected in a consistent way. The same might be true if the dataset were to be divided by date of cohort entry (or other non-random method); splitting a dataset by date also introduces the possibility that the performance of the rule might vary over time. To understand how applicable the HOPE score may be, it is important to ask whether the validation dataset should ideally have included records that would not have been included in the derivation dataset – for example, from centres that are less expert in treating hypothermic patients with ECLS.

There is a difference between the patients reported in published datasets or treated in high-volume centres and hypothermic patients at other centres where treatment decisions about ECLS may also be made. Successful implementation and outcome of ECLS in the emergency department relies on a well-rehearsed pathway, which is unusual for rare events. (There

were 6,000 episodes of all extracorporeal cardiopulmonary resuscitation (ECPR) worldwide in 2018 (not just hypothermia) spread across 9,000 centres[3]; published 10-year experiences at two *specialist* centres yielded 11[4] and 48[5] patients respectively.) Since many of these cases will be concentrated in centres with a specialist interest, expecting the HOPE score to achieve a similar level of prediction in *small* volume centres is probably unrealistic.

The HOPE validation dataset is relatively small. This limitation leads to imprecise estimates of its performance - confidence intervals for predictive values of living or dying given a particular HOPE score are not reported. Many risk prediction tools (eg EuroSCORE[6], P-POSSUM[7]) do not publish the CIs with their predicted risk – they should do. The CI, as well as the point estimate, describes the risk of misclassifications and is important when interpreting and applying a rule. Confidence intervals can change with the predicted risk (the EuroSCORE has narrow CIs in low risk patients and the CIs increase as the risk increases). Clinicians must have CIs presented alongside the calculated risk to be fully informed and adjust their decision making.

Centres should reflect on how they currently make decisions about implementing ECLS. It is clear that using a serum potassium level alone to guide the decision to use ECLS in hypothermic cardiac arrest is inadequate and leads to futile care in a number of cases. We recommend that the HOPE score be widely adopted to inform decisions about using ECLS. Centres should systematically record the information required to calculate the score, hold these data securely and make them available for subsequent analysis. However, applying the HOPE score threshold systematically means that this data can only inform whether the threshold for treatment can be *raised* (i.e. avoiding some additional instances of futile care)

since ECLS will not be offered to patients with a score below the threshold. By stratifying centres according to their experience or volume, analysis of these data should inform whether the survival threshold should be set at a *higher* level for low volume centres to avoid futile care. These data will not address the issue of the validity of using the HOPE score threshold in children, where the concern is that a *lower* threshold may be appropriate. Nonetheless, collecting the data required for the HOPE score would still be worthwhile for all children for whom a decision is made about treatment with ECLS.

Both the number of centres using and the number of patients receiving ECLS are rising[3] and we must begin to risk stratify its use on the basis of evidence rather than positively biased case reports leading to enthusiasm-based use. Whilst the HOPE score is not a perfect prediction tool, it is better than the way in which decisions are currently made and can only be improved by being updated with more recent, independent 'test' data.

Acknowledgement

BG and BR are supported by the UK NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the review authors and not necessarily those of the NHS, the NIHR or the UK Department of Health

References

1. Pasquier M, Rousson V, Darocha T, et al. Hypothermia Outcome Prediction after Extracorporeal Life Support for Hypothermic Cardiac Arrest Patients: an external Validation of the HOPE Score. *Resuscitation* 2019

2. Pasquier M, Hugli O, Paal P, et al. Hypothermia outcome prediction after extracorporeal life support for hypothermic cardiac arrest patients: The HOPE score. *Resuscitation* 2018;126:58-64.

3. Registry E. ELSO Registry - International Summary 2019 [Available from: https://www.elso.org/Registry/Statistics/InternationalSummary.aspx accessed 21.3.19 2019.

4. Khorsandi M, Dougherty S, Young N, et al. Extracorporeal Life Support for Refractory Cardiac Arrest from Accidental Hypothermia: A 10-Year Experience in Edinburgh. *The Journal of emergency medicine* 2017;52(2):160-68.

5. Debaty G, Moustapha I, Bouzat P, et al. Outcome after severe accidental hypothermia in the French Alps: A 10-year review. *Resuscitation* 2015;93:118-23.

6. Chalmers J, Pullan M, Fabri B, et al. Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. *European Journal of Cardio-Thoracic Surgery.* 2013;43(4):688-94.

7. Prytherch DR, Whiteley MS, Higgins B, et al. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. *Br J Surg* 1998;85(9):1217-20.