

Challenges in translating mortality risk to the point of care

Vincent J Major,¹ Yindalon Aphinyanaphongs

Department of Population Health, NYU Langone Health, New York City, New York, USA

Correspondence to

Vincent J Major, NYU Langone Health, New York, NY 10016, USA; vincent.major@nyulangone.org

Accepted 27 August 2019

Despite advances in medicine, prognostication remains inaccurate for many patients. Physicians tend to overestimate survival, even in advanced cancer and terminal illness groups.¹⁻³ Over half of terminally ill patients express they do not want prolonging of life if their quality of life would decline.⁴ End-of-life interventions such as advanced care planning have shown improved adherence to patient's wishes, improvement in satisfaction and reductions in stress, anxiety and depression,⁵ but clinicians remain reluctant to initiate end-of-life discussions with terminal patients if they are currently asymptomatic.⁶ Automated systems can complement clinician judgement to prompt earlier end-of-life discussions.

To this end, predictive analytics is potentially impactful. Many different approaches have been used to estimate mortality risk using factors including severity of illness,⁷ healthcare utilisation⁸ or comorbidities.⁹ However, few works focus on palliative or end-of-life care (PEOLC), and even fewer have translated beyond model validation into prospective testing ultimately affecting clinical care. Instead, PEOLC remains reliant on clinical staff, despite their optimism, for initiation and prioritisation.

The paper by Wegier and colleagues¹⁰ in this issue introduces a new 1-year mortality score—modified Hospitalised-patient One-year Mortality Risk (mHOMR)—designed for broad application at the time of admission. They incorporate mHOMR into two electronic health records (EHRs) to automatically identify patients who may benefit from palliative assessment. Of concern, there is evidence of patient distributional shift at the one site that showed improvement with the intervention. The authors conclude there was an increase in patients who receive palliative care consultations or goals-of-care discussions. However,

the preintervention group appears much healthier, with a 3% in-hospital mortality, compared with the postintervention group (16%). Relatedly, a concomitant shift in patient mix to fewer frail patients is reported (68/100 to 43/97, $p=0.001$; Pearson's χ^2 test with Yates' continuity correction). It is possible, therefore, that the reported increase in palliative assessment is more a function of an increased number of patients qualifying for PEOLC than of any intervention effect.

Nonetheless, this work is commendable for their approach to identify a broad cohort of patients at admission and as one of the few papers that present a predictive modelling system implemented at the point of care. Several issues arise from this work that, if adequately considered, can guide implementation at other institutions, development of similar systems and subsequent effectiveness studies. In particular, this paper raises challenges regarding alignment of recommended interventions with model performance, workload and timing. In addition, as a demonstration of model implementation, the paper warrants a more thorough discussion of model-identified subcohorts and technical barriers to reproduction.

PERFORMANCE AND INTERVENTIONS

Performance of predictive models consists of two related components: overall performance and threshold-specific performance. Measures such as discrimination and calibration assess the model as a whole. However, many applications require separation of patients into two (or more) distinct groups (eg, high risk and low risk) by selecting a defining probability, or threshold for intervention. Several factors are important when selecting a threshold.

The application of an intervention to individual patients is heavily dependent on the model's 'certainty' of a patient



► <http://dx.doi.org/10.1136/bmjqs-2018-009285>



© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Major VJ, Aphinyanaphongs Y. *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2019-009858

being high risk, measured by positive predictive value (PPV). Low PPV is common for broad identification applications, such as screening or populating registries, as it casts a wide net (high sensitivity). Similarly, a PEOLC intervention such as a goals-of-care discussion is straightforward, inexpensive, with little downside and, thus, modest false-positive rates are reasonable. The mHOMR authors acknowledge this reality and select an alert threshold corresponding to specificity=90%, sensitivity=59% with a PPV=36%. That is, two of every three alerted patients will *not* die within 1 year. By contrast, an appropriate threshold for more aggressive PEOLC interventions, such as de-escalation of treatment, would require much higher PPV, likely $\geq 75\%$.

For time and resource-intensive interventions, the volume of patients identified can become a limiting factor and should thus be considered. Prospectively, mHOMR identifies 15.8% and 12.2% of all internal medicine patients admitted across the two sites. This one-in-six volume is sustainable for simpler interventions that are distributable to the care teams, for example, goals-of-care discussions, but would likely not be feasible for more time or resource-intensive interventions. Palliative care consults for more than 10% of patients would likely overwhelm any institution's palliative care team without significant investment in staffing. Regardless of the specific intervention, institutional support is crucial when adding onto the clinical workload.

In addition to workload considerations, any intervention must also be appropriately matched to the population's expected survival for effective implementation. Many PEOLC systems, including mHOMR, focus on 'early' identification of patients at risk of dying in 1 year, often to prompt palliative care, citing the fact that palliative care is often initiated too near a patient's death (median (IQR) of 59 (13, 200) days¹¹). Alternatively, identification of 'sentinel hospitalizations,'¹² the turning point of a patient's trajectory, has been proposed as an appropriate moment to prompt initiation of palliative care.¹³ However, patients follow different trajectories that warrant PEOLC interventions at different times and the inability to distinguish patients with weeks or many months to live complicates recommendation of any single intervention.

The authors of mHOMR describe multiple potential interventions but do not recommend any single one, instead prompting a general palliative approach in their notification. In our opinion, goals-of-care discussion is the only intervention feasible for all patients identified by mHOMR given the time horizon, and predictive performance selected. Naturally, some patients will receive other PEOLC interventions, with or without goals-of-care. The authors report combined rates of either goals-of-care discussion or palliative care consult, allowing clinicians to choose. Concerningly, the authors describe a potential 'ceiling effect'

related to the high baseline palliative care/goals-of-care rates at site 2, suggesting clinicians are self-moderating possibly due to a lack of 'certainty' or that universal goals-of-care is implausible in this broad population. A useful approach to further investigate would be to employ provider agreement to the patient's risk as a process measure. With a low PPV, a long time horizon and an ambiguous intervention, mHOMR may not be persuasive enough to change a clinician's opinion, obstructing effective implementation.

APPLYING MHOMR IN PRACTICE

In restricting intervention to internal medicine patients, the investigators implement a constrained version of mHOMR. Because the original model was developed on 15 services and included interaction terms using service, many coefficients rely on services other than internal medicine (14 other services \times 3 interaction terms=42 of the original 60 coefficients). When implementation is restricted to only include internal medicine, these parameters become irrelevant and are effectively dropped from the model. This may cause model performance to degrade to a point no longer comparable to that initially reported.

MODEL-IDENTIFIED SUBCOHORTS

Existing methods to prompt PEOLC typically estimate long-term mortality in cohorts of older adults using administrative or geriatric assessment data¹⁴⁻¹⁷ but few have translated into widespread clinical practice.¹⁸ One limitation of methods targeting older adults is the explicit omission of younger adults. Avati *et al*¹⁹ used deep learning methods for the prediction of 3-12 months' all-cause mortality in patients of all ages (but received criticism for their experimental design and ill-defined intervention). mHOMR is similarly derived in a general cohort of adult patients. However, the model identifies an older, frail population (mean (SD) age=83 (7.8); 55% frail), similar to other work that explicitly selects an older population.

Prediction-based systems learn patterns in data. Any model can overfit particular parameters, or combinations of data, which can produce counter-intuitive results, for example, surprising high or low risk patients at a given threshold. The advantage of simple models like mHOMR (in contrast to machine learning 'black boxes') is the potential to explore the model coefficients to identify outlier cohorts and potentially restrict use of these untrustworthy predictions. mHOMR only uses nine data elements which minimises the number of data combinations and simplifies identification of outlier cohorts that do exist.

We reproduced mHOMR for all possible data combinations (omitting hospital service) in search of outlier cohorts (code available at: github.com/vincent-major/reproducing-mHOMR). Overall, we find that age is such an important factor that mHOMR automatically identifies patients as high risk with little

other data. For instance, a 77-year-old male patient with zero prior emergency department (ED) visits or admissions by ambulance, living independently at home, presenting electively and without an intensive care unit (ICU) stay will be identified as *high risk* due to his age. At the opposite extreme, acute admissions of high utilisers will be identified as low risk below a certain age. A patient with two prior admissions by ambulance, two prior ED visits, presenting by ambulance to the ED from home with home care services, as a 30-day readmission, and admitted directly to the ICU, will be *low risk* unless he is at least 58 or she is at least 64. mHOMR's reliance on age may be caused by a lack of alternative parameters, such as disease state, that may be more influential than age for particular subgroups.

OPERATIONAL IMPLEMENTATION

From a technical perspective, there are several hurdles to generalise this work into other institutions. First, historical data may be limited in regions with fragmented healthcare systems that may lead to systematically lower risk estimates. Second, mHOMR is calculated inside the EHR which necessitates reproduction, including parameter mapping of non-standardised fields such as hospital service and current living situation. Third, the delivery of the notification may not be easily reproduced at other institutions, as described in this study: the electronic sign-out tool was effective at one site but could not be implemented at the second. Moreover, both the sign-out and email methods have limitations and the preferred notification may vary by institution and recipient type. Overall, widespread deployment of systems like mHOMR will require EHR vendors to develop a toolbox of configurable functionalities and willingness of individual institutions to invest technical and clinical resources.

TIMING WITHIN PATIENT TRAJECTORY

Though we recognise the paper's aim was to assess feasibility, the authors present results that inadvertently challenge mHOMR's utility and future efforts to evaluate PEOLC notifications. As previously described, the identified population is generally elderly with a mean age of 83, an age that is comparable to the life expectancy for Ontario.²⁰ Moreover, 38% of alerted patients were admitted with a current do not resuscitate order and 11% died during their hospitalisation, making the notification irrelevant to them. A comprehensive evaluation should include rates of (1) hospice care, (2) readmission, and (3) patients not willing to engage in PEOLC. Future effectiveness studies should report both process measures describing provider agreement and adherence to the recommendations as well as patient-centred outcome measures of quality and intensity of care.

mHOMR succeeds in demonstrating implementation and initial evaluation of a mortality prediction

model to prompt PEOLC interventions. We anticipate growing application of prediction models driven by increasingly sophisticated algorithms. With model performance, reproduction, implementation and transportability²¹ to external populations will continue to be key challenges to success, especially for more complex models. Before that, we must consider the broader clinical application: whether the notifications come at an impactful time in a patient's trajectory; how, when and to whom the notification is delivered; the patient volume necessary for the feasibility of planned intervention; the data driving each high risk prediction; and performance expectations when applied to different populations. Without addressing these challenges, implementation of predictive modelling will continue to be ad hoc, struggling to reproduce in other populations and systems.

Contributors VJM and YA both contributed to the writing and revision of the manuscript and both have approved the final version submitted.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCES

- Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ* 2000;320:469–73.
- Glare P, Virik K, Jones M, *et al.* A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ* 2003;327:195–8.
- Amano K, Maeda I, Shimoyama S, *et al.* The accuracy of physicians' clinical predictions of survival in patients with advanced cancer. *J Pain Symptom Manage* 2015;50:139–46.
- Singer PA, Martin DK, Kelner M. Quality end-of-life care: patients' perspectives. *JAMA* 1999;281:163–8.
- Detering KM, Hancock AD, Reade MC, *et al.* The impact of advance care planning on end of life care in elderly patients: randomised controlled trial. *BMJ* 2010;340:c1345.
- Keating NL, Landrum MB, Rogers SO, *et al.* Physician factors associated with discussions about end-of-life care. *Cancer* 2010;116:998–1006.
- Knaus WA, Draper EA, Wagner DP, *et al.* Apache II: a severity of disease classification system. *Crit Care Med* 1985;13:818–29.
- Makar M, Ghassemi M, Cutler DM, *et al.* Short-Term mortality prediction for elderly patients using Medicare claims data. *Int J Mach Learn Comput* 2015;5:192–7.
- Charlson M, Szatrowski TP, Peterson J, *et al.* Validation of a combined comorbidity index. *J Clin Epidemiol* 1994;47:1245–51.
- Wegier P, Koo E, Ansari S, *et al.* mHOMR: a feasibility study of an automated system for identifying inpatients having an elevated risk of 1-year mortality. *BMJ Qual Saf* 2019. doi:10.1136/bmjqs-2018-009285. [Epub ahead of print: 28 Jun 2019].

- 11 Seow H, O'Leary E, Perez R, *et al.* Access to palliative care by disease trajectory: a population-based cohort of Ontario decedents. *BMJ Open* 2018;8:e021147.
- 12 Lin RJ, Adelman RD, Diamond RR, *et al.* The sentinel hospitalization and the role of palliative care. *J Hosp Med* 2014;9:320–3.
- 13 Rajaram A, Morey T, Dosani N, *et al.* Palliative care in the twenty-first century: using advanced analytics to Uncloak insights from big data. *J Palliat Med* 2019;22:124–5.
- 14 Teno JM, Harrell FE, Knaus W, *et al.* Prediction of survival for older hospitalized patients: the help survival model. *J Am Geriatr Soc* 2000;48:S16–24.
- 15 Levine SK, Sachs GA, Jin L, *et al.* A prognostic model for 1-year mortality in older adults after hospital discharge. *Am J Med* 2007;120:455–60.
- 16 Dramé M, Novella JL, Lang PO, *et al.* Derivation and validation of a mortality-risk index from a cohort of frail elderly patients hospitalised in medical wards via emergencies: the SAFES study. *Eur J Epidemiol* 2008;23:783–91.
- 17 Bari MD, Balzi D, Roberts AT, *et al.* Prognostic Stratification of Older Persons Based on Simple Administrative Data: Development and Validation of the "Silver Code," To Be Used in Emergency Department Triage. *J Gerontol A Biol Sci Med Sci* 2010;65A:159–64.
- 18 Yourman LC, Lee SJ, Schonberg MA, *et al.* Prognostic indices for older adults: a systematic review. *JAMA* 2012;307:182–92.
- 19 Avati A, Jung K, Harman S, *et al.* Improving palliative care with deep learning. *BMC Med Inform Decis Mak* 2018;18(Suppl 4):122.
- 20 Statistics Canada. *Table 13-10-0389-01: life expectancy, at birth and at age 65, by sex, three-year average, Canada, provinces, territories, health regions and peer groups, 2019.*
- 21 Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515–24.