



## Precise mathematics yet hazy predictions: Can validated risk indices help improve patient selection for major elective surgery?

Duminda N. Wijeyesundera, MD, PhD

Received: 18 May 2017 / Accepted: 8 June 2017 / Published online: 16 June 2017  
© Canadian Anesthesiologists' Society 2017

Not uncommonly, anesthesiologists face the scenario of a very high-risk patient being considered for major elective surgery. Many reasonable doctors speculate whether the likelihood of such a patient's survival might be better served with a non-operative treatment option. Perhaps unsurprisingly, there are remarkably few data available to better our understanding of this important issue. In this issue of the *Journal*, McIsaac *et al.* present a population-based retrospective cohort study, by which they sought, in part, to quantify how patients' observed survival after major elective noncardiac surgery compared with their expected survival based on their baseline preoperative health status.<sup>1</sup>

The investigators assembled a cohort of about 450,000 adults who underwent elective noncardiac surgery in Ontario over a 14-year period. They then estimated the risk of these patients dying within the next year using the Hospital-patient One-year Mortality Risk (HOMR) score.<sup>2</sup> The HOMR score uses readily available information at the time of admission to estimate the probability of death within the subsequent year. While the HOMR score was

initially developed using information from administrative databases in Ontario,<sup>2</sup> this relatively simple scoring scheme has shown remarkably stable predictive performance when tested across different hospital settings, countries, and data sources.<sup>3,4</sup> Importantly, the HOMR score was developed and validated in samples that included a broad spectrum of hospitalized patients, including inpatient admissions with varying urgency (i.e., elective, urgent, emergent) and indications (i.e., surgical, medical, trauma, obstetrical).<sup>2,3</sup> McIsaac *et al.* sought to determine the extent of agreement between the observed risks of one-year mortality in elective surgical patients and the risks predicted by the HOMR score.

What were their findings? Overall, the observed mortality in this large cohort of patients undergoing major elective noncardiac surgery was, in relative terms, about 25% lower than that predicted by the HOMR score. This pattern of lower than expected mortality was observed in all but two of the surgical procedures studied, with total cystectomy being the only procedure type with a significantly higher than expected mortality risk. Additionally, the magnitude of the discrepancy between the observed vs predicted mortality risk was greatest among patients with a higher predicted risk of death. Stated otherwise, the overestimation of one-year mortality by the HOMR score was greatest among the highest risk patients. Thus, from a methodological perspective, the HOMR score showed poor *calibration* (Fig. 1), since observed one-year death rates were generally lower than those predicted by the score.<sup>5</sup> Conversely, the score still showed good *discrimination*, since the distribution of scores was still higher in individuals who died vs those who did not.<sup>5</sup> The reported c-statistic of 0.82 indicates that, if the HOMR score of a randomly selected patient who died is compared with that of a randomly selected patient who did not, there

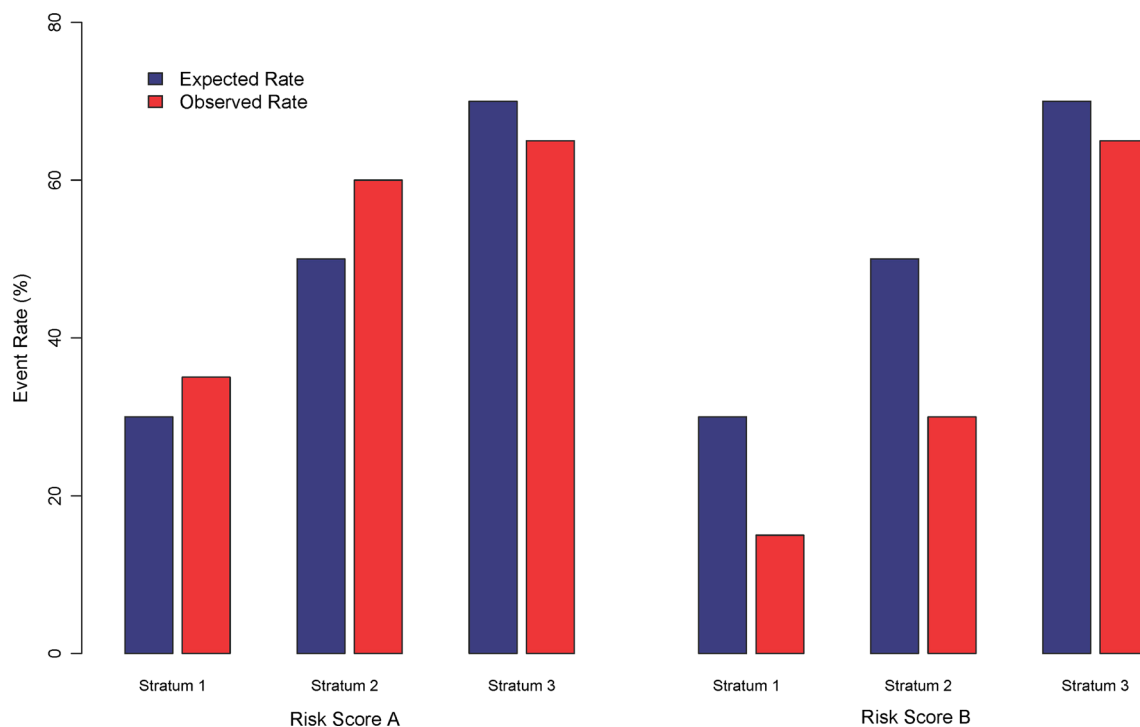
---

D. N. Wijeyesundera, MD, PhD (✉)  
Li Ka Shing Knowledge Institute of St. Michael's Hospital,  
30 Bond Street, Toronto, ON M5B 1W8, Canada  
e-mail: d.wijeyesundera@utoronto.ca

D. N. Wijeyesundera, MD, PhD  
Department of Anesthesia and Pain Management, Toronto  
General Hospital, Toronto, ON, Canada

D. N. Wijeyesundera, MD, PhD  
Department of Anesthesia, University of Toronto, Toronto, ON,  
Canada

D. N. Wijeyesundera, MD, PhD  
Institute of Health Policy Management and Evaluation,  
University of Toronto, Toronto, ON, Canada



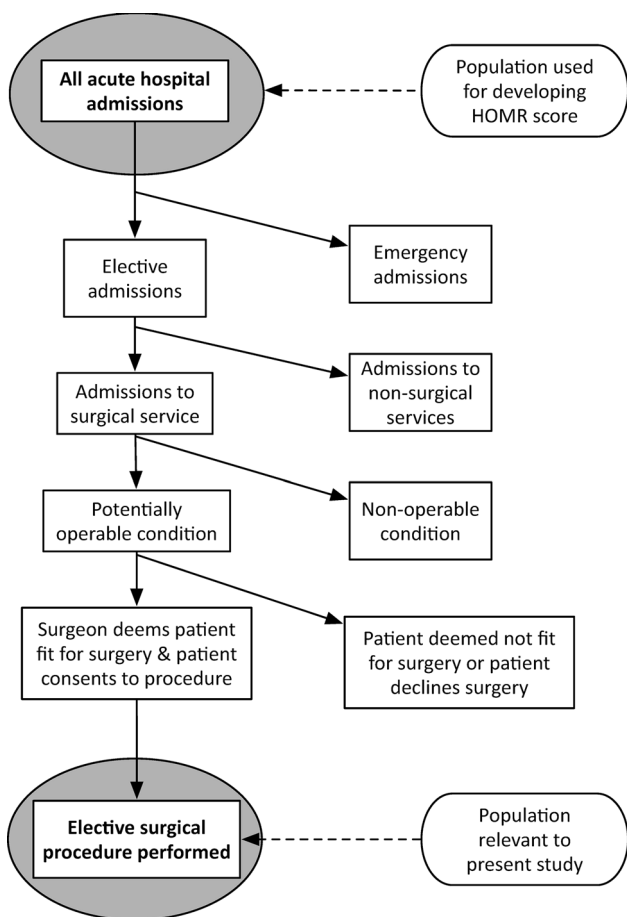
**Fig. 1** Assessing the calibration of predictive indices. Calibration refers to how well event rates in observed outcomes agree with event rates predicted by a predictive index such as the Hospital-patient One-year Mortality Risk score.<sup>2,5</sup> Fig. 1 presents observed (red columns) and predicted (blue columns) event rates for two different hypothetical predictive indices. Each index categorizes patients into one of three risk strata, with the expected risks increasing incrementally from strata 1 to 3. The red columns indicate that both predictive indices perform reasonably well with respect to *discrimination*, in that the observed event rates differ across the strata. Nonetheless, when comparing event rates for expected *vs* observed outcomes, risk score A has better overall *calibration*, in that observed event rates are similar to expected rates across all three risk strata. By comparison, strata 1 and 2 in risk score B show poorer calibration, with considerable overprediction of event rates. Stratum 3 in risk score B is better calibrated, as evidenced by the similarity between expected and observed event rates in this stratum. The graph was plotted using the R Statistical Language version 3.2.1 (Vienna, Austria). Reproduced with permission from: *Wijeyesundera DN. Predicting outcomes: is there utility in risk scores? Can J Anesth 2016; 63: 148-58*<sup>5</sup>

Évaluation de l'étalonnage des indices de prédiction. L'étalonnage fait référence à la mesure dans laquelle les taux d'événements, dans les résultats observés, correspondent aux taux d'événements prédits par un indice de prédiction tel que le score HOMR.<sup>2,5</sup> La figure 1 présente les taux d'événements observés (colonnes rouges) et prédits (colonnes bleues) pour deux indices de prédiction hypothétiques différents. Chaque indice catégorise les patients dans l'une de trois strates de risque, les risques attendus augmentant de façon progressive des strates 1 à 3. Les colonnes rouges indiquent que les deux indices de prédiction sont raisonnablement fiables en matière de *discrimination*, en ce que les taux d'événements observés sont différents d'une strate à l'autre. Toutefois, lorsqu'on compare les taux d'événements en ce qui touche aux résultats attendus *vs* observés, la cote de risque A affiche un meilleur étalonnage global, car les taux d'événements observés sont semblables aux taux attendus dans les trois strates de risque. Comparativement, les strates 1 et 2 de la cote de risque B affichent un moins bon étalonnage, étant donné leur sur-prédiction considérable des taux d'événements. La strate 3 de la cote de risque B est mieux étalonnée, comme le démontre la similitude entre les taux d'événements attendus et observés dans cette strate. Ce graphique a été dessiné à l'aide de R Statistical Language, version 3.2.1 (Vienne, Autriche). Reproduit avec la permission de : *Wijeyesundera DN. Predicting outcomes: is there utility in risk scores? Can J Anesth 2016; 63: 148-58*<sup>5</sup>

is an 82% probability that the score would be higher in the individual who died.

Thus, patients undergoing major elective noncardiac surgery in Ontario generally appear to have much better one-year survival than would be predicted by a previously well-validated predictive score. The key question: Why? There are several possibilities. The operative procedure itself may have improved patients' one-year survival above that predicted by the HOMR score. For example, successful surgical resection of colon cancer or surgical treatment of abdominal aortic aneurysms could plausibly and

significantly improve long-term survival. In addition, the referral process for elective surgery might have selected patients with better prognoses than would be expected based solely on HOMR scores. As depicted in Fig. 2, multiple steps were entailed for inclusion in McIsaac *et al.*'s study cohort. Some of the factors influencing these steps are included in the HOMR score, e.g., admission urgency or the primary admitting service; hence, their prognostic impact can be accounted for in the score. Conversely, several other prognostically important factors are incompletely accounted for in the risk score.



**Fig. 2** Patient selection process for major elective surgery. While the Hospital-patient One-year Mortality Risk (HOMR) score was derived in a broad sample of hospitalized inpatients, McIsaac *et al.*'s study cohort was restricted to those selected for major elective surgery.<sup>1</sup> The refinement of the larger population to the elective surgical population is based on several factors, not all of which are accounted for in the HOMR score. This process could plausibly have resulted in a selection bias that affected the calibration of the risk score<sup>6</sup> *Processus de sélection des patients pour les chirurgies non urgentes majeures. Alors que le score HOMR (Hospital-patient One-year Mortality Risk) a été dérivé d'un vaste échantillon de patients hospitalisés, la cohorte d'étude de McIsaac et coll. était restreinte aux patients sélectionnés pour une chirurgie non urgente majeure.<sup>1</sup> La réduction sélective d'une population plus importante à la population de chirurgie non urgente se fonde sur plusieurs facteurs, lesquels ne sont pas tous pris en compte par le score HOMR. Ce processus pourrait potentiellement avoir causé un biais de sélection qui a affecté l'étalonnage de la cote de risque<sup>6</sup>*

For example, while the HOMR score does account for preexisting cancer, it does not capture the extent to which the cancer might be surgically resectable. Thus, the score cannot differentiate between patients with potentially resectable pancreatic cancer (i.e., those who would be included in the present study cohort) vs those deemed inoperable (i.e., those who would be excluded from the cohort and suffer a very high risk of one-year mortality). Similarly, the HOMR score incompletely captures poor

functional capacity or the severity of underlying comorbidities, e.g., coronary artery disease or heart failure. Importantly, surgeons' selection of patients for elective surgery is, in part, influenced by these same incompletely measured but prognostically important factors. Thus, within the same strata of HOMR scores, the patients selected by surgeons for elective surgery are likely to have systematically better prognoses. This possibility is supported by the observation that the discrepancy between observed vs predicted mortality risks was greatest among patients with higher predicted risks of death. Specifically, unmeasured prognostically important factors are more likely to impact whether a surgeon selects a patient with multiple significant comorbidities for surgery vs an otherwise healthy patient with well-controlled hypertension. Consequently, the prediction of one-year mortality by the HOMR score is more likely to be biased among higher risk patients, as was observed by McIsaac *et al.* Overall, the impact of including patients in the study cohort based on incompletely measured factors is a form of *selection bias* that affects the calibration of the HOMR score.<sup>6</sup> Thus, while the HOMR score is well calibrated in a broad group of hospitalized patients with a significant proportion of surgical patients,<sup>2,3</sup> the narrower restriction to elective surgical patients could plausibly have worsened its calibration.

How should we interpret these data? *First*, there remains uncertainty regarding how best to quantify the expected survival of high-risk patients who are uncertain whether to proceed with surgery, a non-operative alternative, or no treatment for an underlying condition. McIsaac *et al.*'s findings indicate that the HOMR score, while a well-validated general predictive index, should likely not be used to estimate the risk of one-year mortality in such patients. For now, randomized trials may be the best approach to help inform our understanding of this issue. The preference for surgery vs non-operative treatment will likely vary depending on the underlying surgical condition and the intervention being evaluated. For example, transcatheter aortic valve replacement is superior to medical therapy with respect to improving survival in very high-risk patients with symptomatic aortic stenosis,<sup>7</sup> yet endovascular aortic repair is not superior to ongoing surveillance with respect to improving overall survival in high-risk patients with abdominal aortic aneurysms.<sup>8</sup> *Second*, anesthesiologists should be broadly reassured that, in general, patient selection for major elective surgery seems to occur in a reasonably rational manner. Even after accounting for known predictors of one-year mortality, surgeons appear to select a healthier spectrum of patients. While more work clearly remains to be done to advance our understanding of appropriate patient selection for surgery, for now, clinical judgement is doing a reasonably good job.

## Des calculs précis, mais des prédictions vagues : les indices de risque validés peuvent-ils améliorer la sélection des patients pour les chirurgies non urgentes majeures?

Bien souvent, les anesthésiologistes sont confrontés à des situations où ils doivent évaluer si un patient à très haut risque peut ou non subir une chirurgie non urgente majeure. Bon nombre de docteurs tout à fait raisonnables spéculent quant à savoir si la probabilité de survie d'un tel patient ne serait pas mieux servie avec une option thérapeutique non chirurgicale. Sans surprise peut-être, les données disponibles pour améliorer notre compréhension de cette question importante brillent par leur rareté. Dans ce numéro du *Journal*, McIsaac *et coll.* présentent une étude de cohorte rétrospective dans laquelle ils ont tenté, en partie, de quantifier la survie observée des patients après une chirurgie majeure non cardiaque et non urgente par rapport à leur survie anticipée selon leur état de santé préopératoire de base.<sup>1</sup>

Les chercheurs ont rassemblé une cohorte de quelques 450 000 adultes ayant subi une chirurgie non cardiaque et non urgente en Ontario sur une période de 14 ans. Ils ont ensuite estimé le risque que ces patients meurent au cours de l'année suivante en utilisant le score HOMR (pour *Hospital-patient One-year Mortality Risk*, soit Risque de mortalité à un an du patient hospitalisé).<sup>2</sup> Le score HOMR utilise des informations facilement disponibles au moment de l'admission afin d'estimer la probabilité de décès au cours de l'année subséquente. Alors que le score HOMR a d'abord été mis au point en utilisant des informations tirées de bases de données administratives en Ontario,<sup>2</sup> ce système de gradation relativement simple a démontré une performance de prédiction remarquablement stable lorsqu'il a été testé dans différents contextes hospitaliers, pays et sources de données.<sup>3,4</sup> Fait important, le score HOMR a été mis au point et validé à l'aide d'échantillons qui comprenaient une vaste gamme de patients hospitalisés, notamment les admissions de patients en milieu hospitalier selon divers degrés d'urgence (c.-à-d. non urgent, urgent ou très urgent) et diverses indications (c.-à-d. chirurgicale, médicale, de trauma, obstétricale).<sup>2,3</sup> McIsaac *et coll.* ont cherché à déterminer l'ampleur de la concordance entre les risques observés de mortalité à un an chez les patients de chirurgie non urgente et les risques tels que prédits par le score HOMR.

Qu'ont-ils découvert? Globalement, la mortalité observée dans cette importante cohorte de patients subissant des

chirurgies non cardiaques non urgentes majeures était, en termes relatifs, environ 25 % *plus basse* que celle prédite par le score HOMR. Cette tendance de mortalité plus basse que prévue a été observée pour toutes les interventions chirurgicales étudiées sauf deux, la cystectomie totale étant le seul type d'intervention démontrant un risque de mortalité significativement plus élevé que prévu. En outre, l'ampleur de la divergence entre le risque de mortalité observé vs prédit était la plus importante chez les patients ayant un risque prédit plus élevé de décès. En d'autres termes, la surestimation de la mortalité à un an par le score HOMR était la plus importante chez les patients à plus haut risque. Ainsi, d'un point de vue méthodologique, le score HOMR a démontré un faible *étalonnage* (figure 1), étant donné que les taux de décès observés à un an étaient généralement plus bas que ceux prédits par le score.<sup>5</sup> Toutefois, le score a tout de même démontré une bonne *discrimination*, étant donné que la distribution des scores était tout de même plus élevée chez les personnes qui sont décédées que chez les personnes qui ont survécu.<sup>5</sup> La statistique C rapportée de 0,82 indique que, si le score HOMR d'un patient décédé choisi au hasard est comparé au score d'un patient non décédé également choisi de façon aléatoire, la probabilité que le score de la personne décédée soit plus haut que celui de l'autre est de 82 %.

Dès lors, les patients subissant une chirurgie non cardiaque non urgente majeure en Ontario semblent en général avoir un taux de survie à un an bien meilleur que ce qui serait prévu par un score prédictif précédemment bien validé. La question clé est donc : pourquoi? Plusieurs réponses sont possibles. L'intervention chirurgicale peut, en soi, avoir amélioré le taux de survie à un an des patients de telle façon à ce qu'il augmente par rapport au taux initial prédit par le score HOMR. Par exemple, il serait plausible que la résection chirurgicale réussie d'un cancer du côlon ou le traitement chirurgical d'un anévrisme de l'aorte abdominale puisse améliorer la survie à long terme d'un patient de manière significative. En outre, le processus de référence pour les chirurgies non urgentes pourrait avoir permis la sélection de patients ayant de meilleurs pronostics qu'anticipés si l'on se fondait exclusivement sur les scores HOMR. Comme l'illustre la figure 2, plusieurs étapes étaient nécessaires pour admettre l'inclusion d'un patient dans la cohorte de l'étude de McIsaac *et coll.* Certains des facteurs influençant ces étapes sont pris en compte par le score HOMR, par ex., l'urgence de l'admission ou le service d'admission primaire; ainsi, le score peut tenir compte de leur impact pronostique. À l'inverse, plusieurs autres facteurs importants d'un point de vue pronostique ne sont pas pris en compte de façon exhaustive dans la cote de risque.

Par exemple, bien que le score HOMR tienne compte d'un cancer préexistant, il ne peut saisir l'ampleur de la résection chirurgicale possible du cancer. Ainsi, le score ne peut

distinguer les patients ayant un cancer du pancréas potentiellement résecable (c.-à-d. les patients qui seraient inclus dans la cohorte de l'étude dont il est question ici) des patients jugés inopérables (c.-à-d. ceux qui seraient exclus de l'étude et courraient un risque très élevé de mortalité à un an). De la même manière, le score HOMR ne donne qu'une image incomplète d'une mauvaise capacité fonctionnelle ou de la gravité des comorbidités sous-jacentes, comme par exemple une coronaropathie ou une insuffisance cardiaque. Il est important de souligner que la sélection des patients pour chirurgie non urgente par les chirurgiens est en partie influencée par ces facteurs, certes mesurés de façon incomplète, mais importants d'un point de vue pronostique. Donc, dans la même strate de score HOMR, il est probable que les patients sélectionnés pour chirurgie non urgente par les chirurgiens manifestent des pronostics systématiquement meilleurs. Cette possibilité est étayée par l'observation que la divergence entre les risques de mortalité observés vs prédits était la plus flagrante chez les patients ayant des risques prédits de décès plus élevés. Plus spécifiquement, des facteurs importants d'un point de vue pronostique – mais non mesurés – auront probablement un impact plus important sur la sélection d'un patient présentant plusieurs comorbidités importantes pour la chirurgie par rapport à un patient autrement en bonne santé avec une hypertension bien maîtrisée. Par conséquent, la prédiction de mortalité à un an par le score HOMR court un risque plus élevé de biais chez les patients à risque plus élevé, comme l'ont observé McIsaac *et coll.* Globalement, l'impact de l'inclusion de patients dans la cohorte d'étude selon des facteurs mesurés de façon incomplète est une forme de *biais de sélection* qui affecte l'étalonnage du score HOMR.<sup>6</sup> Ainsi, bien que le score HOMR soit bien étalonné pour évaluer un vaste groupe de patients hospitalisés comptant une importante proportion de patients chirurgicaux,<sup>2,3</sup> la restriction plus étroite aux patients de chirurgie non urgente pourrait, de façon tout à fait plausible, avoir quelque peu faussé son étalonnage.

Comment interpréter ces données? *Premièrement*, l'incertitude demeure quant à savoir comment quantifier au mieux la survie de base attendue des patients à risque élevé qui ne sont pas certains de devoir aller de l'avant avec une chirurgie, une alternative non chirurgicale, ou ne subir aucun traitement pour une condition sous-jacente. Les observations de McIsaac *et coll.* indiquent que le score HOMR, tout en étant un indice de prédiction générale bien validé, ne devrait probablement pas être utilisé pour estimer le risque de mortalité à un an de tels patients. Pour le moment, les études randomisées constituent la meilleure approche disponible pour nous aider mieux comprendre cette question. La préférence pour la chirurgie vs un traitement non chirurgical va probablement varier selon la condition chirurgicale sous-jacente et l'intervention évaluée. Par

exemple, un remplacement de valve aortique par voie percutanée est supérieur à un traitement médical quant à l'amélioration de la survie chez les patients à risque très élevé atteints de sténose aortique symptomatique,<sup>7</sup> mais une réparation par technique endovasculaire n'est pas supérieure à une surveillance continue quant à l'amélioration de la survie globale chez les patients à risque élevé atteints d'anévrismes de l'aorte abdominale.<sup>8</sup> *Deuxièmement*, les anesthésiologistes devraient être rassurés que, la plupart du temps, la sélection des patients pour une chirurgie non urgente majeure semble se faire de façon à la fois raisonnable et rationnelle. Même en ayant tenu compte des prédicteurs connus de mortalité à un an, les chirurgiens semblent sélectionner un éventail général de patients en meilleure santé. Bien que des travaux supplémentaires soient clairement nécessaires pour faire progresser notre compréhension d'une sélection de patients adaptée pour la chirurgie, pour le moment, le jugement clinique fait un relativement bon travail.

**Conflicts of interest** None declared.

**Editorial responsibility** This submission was handled by Dr. Hilary P. Grocott, Editor-in-Chief, *Canadian Journal of Anesthesia*.

**Financial support** Dr. Wijeyesundera is supported in part by a New Investigator Award from the Canadian Institutes of Health Research and a Merit Award from the Department of Anesthesia at the University of Toronto.

**Conflit d'intérêt** Aucun.

**Responsabilité éditoriale** Cet article a été traité par Dr Hilary P. Grocott, rédacteur en chef, *Journal canadien d'anesthésie*.

**Soutien financier** Dr Wijeyesundera est soutenu en partie par une Bourse salariale de nouveau chercheur des Instituts de recherche en santé du Canada et une Bourse de mérite du département d'anesthésie de l'Université de Toronto.

## References

1. McIsaac D, Lavalley LT, van Walraven C. A retrospective assessment of prognostication in 456,685 patients undergoing elective major non-cardiac surgery. *Can J Anesth* 2017; 64: this issue. DOI: [10.1007/s12630-017-0909-x](https://doi.org/10.1007/s12630-017-0909-x).
2. van Walraven C. The Hospital-patient One-year Mortality Risk score accurately predicted long-term death risk in hospitalized patients. *J Clin Epidemiol* 2014; 67: 1025-34.
3. van Walraven C, McAlister FA, Bakal JA, Hawken S, Donze J. External validation of the Hospital-patient One-year Mortality Risk (HOMR) model for predicting death within 1 year after hospital admission. *CMAJ* 2015; 187: 725-33.
4. van Walraven C, Forster AJ. The HOMR-Now! model accurately predicts 1-year death risk for hospitalized patients on admission. *Am J Med* 2017. DOI:[10.1016/j.amjmed.2017.03.008](https://doi.org/10.1016/j.amjmed.2017.03.008).
5. Wijeyesundera DN. Predicting outcomes: is there utility in risk scores? *Can J Anesth* 2016; 63: 148-58.

6. *Savitz DA*. Interpreting Epidemiologic Evidence: Strategies for Study Design and Analysis. New York, NY: Oxford University Press; 2001.
7. *Leon MB, Smith CR, Mack M, et al*. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med* 2010; 363: 1597-607.
8. *EVAR trial participants*. Endovascular aneurysm repair and outcome in patients unfit for open repair of abdominal aortic aneurysm (EVAR trial 2): randomised controlled trial. *Lancet* 2005; 365: 2187-92.

Canadian Journal of Anaesthesia is a copyright of Springer, 2017. All Rights Reserved.