



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Jorissen, RN;Croxford, M;Jones, IT;Ward, RL;Hawkins, NJ;Gibbs, P;Sieber, OM

**Title:**

Evaluation of the transferability of survival calculators for stage II/III colon cancer across healthcare systems

**Date:**

2019-07-01

**Citation:**

Jorissen, R. N., Croxford, M., Jones, I. T., Ward, R. L., Hawkins, N. J., Gibbs, P. & Sieber, O. M. (2019). Evaluation of the transferability of survival calculators for stage II/III colon cancer across healthcare systems. *International Journal of Cancer*, 145 (1), pp.132-142. <https://doi.org/10.1002/ijc.32100>.

**Persistent Link:**

<https://hdl.handle.net/11343/285347>

Sieber Oliver (Orcid ID: 0000-0001-9480-0786)

## Evaluation of the Transferability of Survival Calculators for Stage II/III Colon Cancer Across Healthcare Systems

**Running Title:** Evaluation of Colon Cancer Prognosis Calculators

Robert N. Jorissen<sup>1,2</sup>, Matthew Croxford<sup>3</sup>, Ian T. Jones<sup>4</sup>, Robyn L. Ward<sup>5</sup>, Nicholas J. Hawkins<sup>6</sup>, Peter Gibbs<sup>1,2,7,8</sup> and Oliver M Sieber<sup>1,2,9,10</sup>

<sup>1</sup>Systems Biology and Personalised Medicine Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia

<sup>2</sup>Department of Medical Biology, The University of Melbourne, Parkville, Victoria, Australia

<sup>3</sup>Department of Surgery, Western Health, Footscray, Victoria, Australia

<sup>4</sup>Department of Surgery, Royal Melbourne Hospital, Parkville, Victoria, Australia

<sup>5</sup>Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

<sup>6</sup>Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/ijc.32100](https://doi.org/10.1002/ijc.32100)

<sup>7</sup>Department of Medical Oncology, Royal Melbourne Hospital, Parkville, Victoria, Australia

<sup>8</sup>Department of Medical Oncology, Western Health, Footscray, Victoria, Australia

<sup>9</sup>Department of Surgery, The University of Melbourne, Parkville, Victoria, Australia

<sup>10</sup>Department of Biochemistry & Molecular Biology, Monash University, Clayton, Victoria, Australia

**Correspondence to:** Oliver Sieber, Systems Biology and Personalised Medicine Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, VIC 3052, Australia. Tel: +61 (0) 3 9345 2885. E-mail: sieber.o@wehi.edu.au.

### **Novelty & Impact Statements**

Survival calculators for colon cancer integrating tumor stage and patient variables are emerging as important tools to assist clinical decision-making. Limited data exist on the transferability of calculators across healthcare systems. Our assessment of five calculators trained on data from either US or clinical trial patients in an Australian community cohort identified deficiencies in calibration for four tools. Our findings

highlight the need for tailoring of cancer survival calculators prior to adoption across healthcare settings.

## ABSTRACT

Adjuvant! Online Inc (A!O), the Memorial Sloan Kettering Cancer Center (MSKCC), MD Anderson (MDA) and Mayo Clinic (MC) provide calculators to predict survival probabilities for patients with resected early-stage colon cancer, trained on data from United States (US) patient cohorts or patients enrolled in international clinical trials. Limited data exist on the transferability of calculators across healthcare systems. Calculator transferability to Australian community practice was evaluated for 1,401 stage II/III patients. Calibration and discrimination were assessed for overall (OS), cancer-specific (CSS) or recurrence-free survival (RFS). The US patient cohort-based calculators, A!O, MSKCC and MDA, significantly overestimated risks of recurrence and death in Australian patients, with 5-year OS, CSS and RFS prediction differences of -6.5% to -9.9%, -9.1% to -14.4% and -3.8% to -6.8%, respectively ( $P < 0.001$ ). Significant heterogeneity in calibration was observed for subgroups by tumor stage and treatment, age, gender, tumor location, ECOG and ASA score. Calibration appeared acceptable for the clinical trial patient-based MC calculator, but restricted tool applicability (stage III patients,  $\geq 12$  examined lymph nodes, receiving adjuvant treatment) limited the sample size. Compared to AJCC 7<sup>th</sup> edition tumor staging, calculators showed improved discrimination for OS, but no improvement for CSS and RFS. In conclusion, deficiencies in calibration limited transferability of US patient cohort-based survival calculators for early-stage colon cancer to the setting of Australian community practice. Our results

demonstrate the utility for multi-feature survival calculators to improve OS predictions but highlight the importance for performance assessment of tools prior to implementation in an external health care setting.

**Keywords:** Colon cancer, survival calculator, calibration, discrimination

## INTRODUCTION

Treatment recommendations for patients with stage II/III colon cancer are based on anticipated prognosis, consideration of treatment-related risks and expected benefit of adjuvant chemotherapy. Standard-of-care for patients with stage III colon cancer is fluoropyrimidine (FP) and oxaliplatin-based chemotherapy, although elderly individuals or those with comorbidities may receive FP alone. FP-based adjuvant chemotherapy is further considered for a subset of stage II patients with high-risk features including T4 stage, high grade, lymphovascular or perineural invasion, inadequately sampled lymph nodes, and obstruction or perforation of the bowel.<sup>1</sup>

In current practice, prognosis for colon cancer is largely estimated from tumor pathological staging, with the American Joint Committee on Cancer (AJCC) staging system, recently updated from the 7<sup>th</sup> to the 8<sup>th</sup> edition, perhaps the most widely used risk classification system.<sup>2, 3</sup> However, multiple other patient characteristics such as age, body mass index (BMI) and performance status affect the risk of recurrence or death after surgery and must be considered based on clinician judgement. To assist the decision-making process for clinicians and patients, calculators that integrate tumor stage and patient variables have been developed to estimate colon cancer outcome, including five web-based tools from United States (US)-based institutions: Adjuvant! Online Inc, the Memorial Sloan Kettering Cancer Center, MD Anderson Cancer Center and Mayo Clinic.

The Adjuvant!Online (A!O) calculator for colon cancer was developed using data from the US Surveillance Epidemiology and End Results (SEER) tumor registry (US National Cancer Institute) and treatment efficacy estimates from the literature ([www.adjuvantonline.com/](http://www.adjuvantonline.com/)). The tool has been online since the early 2000s, although the website is presently unavailable. The A!O tool utilizes data on age, gender, tumor depth (T stage), categorized number of examined and positive nodes, tumor grade and comorbidity. Predicted outcomes include the probabilities of survival, death due to cancer, and death due to other causes at 5 years post-diagnosis, with and without chemotherapy (single agent fluoropyrimidine or combination).

Two calculators are available from the Memorial Sloan Kettering Cancer Center, which estimate recurrence-free survival at 5 and 10 years post-surgery (MSKCC-RFS) and overall-survival at 5 years post-diagnosis (MSKCC-OS), respectively. The MSKCC-RFS calculator was developed using institutional data from 1,320 patients<sup>4</sup> and considers age, gender, T stage, number of lymph nodes examined and positive nodes, tumor location, differentiation, lymphovascular invasion, perineural invasion, pre-surgical carcinoembryonic antigen (CEA) level, and whether the patient will receive chemotherapy. In contrast, the MSKCC-OS calculator was trained on SEER data and only includes patient age, gender, T and N stage, number of lymph nodes examined and positive nodes, and tumor grade and lacks an option for specifying chemotherapy.<sup>5</sup>

The MD Anderson Cancer Center calculator is based on SEER data and estimates (conditional) cancer-specific survival for the period between 1 year and 10 years post-diagnosis.<sup>6</sup> The tool includes tumor stage (AJCC 6<sup>th</sup> edition), patient age category, gender, differentiation (grade), and ethnicity, but does not consider provision of adjuvant chemotherapy.

The Mayo Clinic (MC) calculator was constructed using data from almost 16,000 patients from 12 randomized international clinical trials registered in the Adjuvant Colon Cancer End Points (ACCENT) database.<sup>7</sup> The tool is restricted to patients with stage III colon cancer with  $\geq 12$  examined lymph nodes and receiving adjuvant treatment (fluoropyrimidine single agent or combination). Predictions are provided for 5-year overall survival and 3-year time to recurrence post initiation of chemotherapy. Among the variables considered are patient age, gender, T and N stage, differentiation, number of lymph nodes examined ( $\geq 12$ ) and positive nodes ( $\geq 1$ ), location or multiple occurrence of primary tumors, performance status, race and body mass index.

The above web-based colon cancer survival calculators were either trained on data from US-based patient cohorts or patients enrolled in clinical trials. However, information on the relative utility of these tools and to what extent these are applicable to patients in the context of other healthcare systems remains limited. Population differences in socio-economic and ethnic background, incidence of co-morbidities such as obesity, duration and type of adjuvant and surgical treatment, and intrinsic tumor

characteristics might all affect prognosis and thus limit transferability of colon cancer survival calculators.

A previous evaluation study of the MSKCC-OS calculator reported good calibration in a cohort of 985 Chinese patients, with improved discrimination for overall survival as compared to the AJCC (7<sup>th</sup> edition) tumor staging system.<sup>8</sup> For the A!O calculator, examination of a population-based cohort from British Columbia (n=2,033) and patients enrolled in the North Central Cancer Trials Group (NCCTG) trials NCCTG 94651 and NCCTG 914653 (n=1,729) found reasonable performance for the prediction of outcomes in stage III disease. However, the A!O calculator was found to overestimate recurrence-free and overall survival for patients with stage II disease who received 5-flourouracil, in particular for the population-based cohort.<sup>9</sup> Two small European cohort studies (n=138 to 134) for the A!O or MSKCC calculator demonstrated utility for stratification of patients into risk groups for recurrence-free survival, but no formal assessments of calibration were reported.<sup>10, 11</sup>

A comparative evaluation of the transferability of the A!O, MSKCC, MDA and MC colon cancer survival calculators, trained on either US or clinical trial patient data, to a non-US community hospital setting is currently lacking. Here we examined the transferability of these tools to the Australian healthcare system utilizing a cohort of 1,401 community-based patients with resected stage II/III colon cancer. Specifically, calculators were evaluated for their ability to predict outcomes in groups of patients

(calibration) and to distinguish individuals who will experience different outcomes (discrimination), comparing discriminatory accuracy with the AJCC (7<sup>th</sup> edition) tumor staging system.

## **MATERIALS AND METHODS**

### **Study population**

We analyzed data from 1,401 patients with resected stage II or III colon adenocarcinoma from two Melbourne Health Services, Royal Melbourne Hospital and Western Health, recruited between January 1, 1995 and December 31, 2011 within the Australian Comprehensive Cancer Outcomes and Research Database (ACCORD), accessed via the BioGrid Australia platform (<https://www.biogrid.org.au/>). Patients who received neoadjuvant or radiation treatment or presented with early-onset colon cancer (<40 years of age) or familial colorectal cancer syndromes were excluded. ACCORD is a prospectively collected national database with point-of-care recording of patient demographics, surgery, pathology, treatment, recurrence, and follow-up data. Follow-up post curative-intent surgery is conducted according to the Australian National Health and Medical Research Council guidelines (clinical examination and CEA levels every 3-6 months for 2 years, and every 6 months, to yearly thereafter; with computed tomography (CT) scan at 12, 24, and 36 months; and initial colonoscopy 12 months

postresection and every 3-5 years thereafter). Data obtained from ACCORD included patient age, gender, date of diagnosis, tumor stage, T stage, number of nodes examined and positive nodes, tumor location, differentiation, lymphovascular invasion, preoperative CEA level, American Society of Anesthesiologists (ASA) score, Eastern Cooperative Oncology Group (ECOG) Performance Status, body mass index, adjuvant treatment and regimen received (Supplementary Table 1). All patients gave informed consent, and this study was approved by the medical ethics committees of all sites.

### **Web-based colon cancer survival calculators**

Five colon cancer survival calculators were evaluated: A!O, MSKCC-OS and MSKCC-RFS, MDA and MC tools (Supplementary Table 2). For the A!O calculator, two model variants were assessed: one in which the comorbidity status variable [“perfect health”, “average for age” “minor problems”, “major problems (+10 years)”, “major problems (+20 years)”, “major problems (+30 years)”] was defined based on the observed distributions of patient age, ECOG status and ASA scores (A!O model 1; Supplementary Methods); another in which every patient was assigned a comorbidity status of “average for age” (A!O model 2). Use of the MSKCC-RFS predictor required imputation of data for perineural invasion making use of an independent Australian community cohort (n=611) from the Molecular and Cellular Oncology (MCO) study tumor collection (R.L.W. and N.J.H; <https://researchdata.andso.org.au/mco-study->

tumour-collection/17113) as this was not recorded in our cohort (Supplementary Methods, Supplementary Table 3).

### **Statistical analysis**

Statistical analysis was performed using the R statistical computing environment (version 3.2.5) making use of the “survival”, “Hmisc” and “rms” libraries (<https://www.R-project.org/>). Predictions for colon cancer survival calculators were made from time of diagnosis for the A!O, MSKCC-OS and MDA tools, time from curative surgery for the MSKCC-RFS calculators and time from provision of chemotherapy for the MC calculator as stipulated by the respective tools (Table 1). We assessed 5-year survival percentages with appropriate censoring of data for all calculators, with the exception of recurrence-free survival for the MC calculator which was for 3-year survival.

We assessed the calibration of colon cancer calculators by comparing observed Kaplan-Meier survival percentages for a given set of patients against the mean of the predicted percentages for that set of patients. Statistical testing of differences in predicted from observed survival percentages was performed using a Z test, in which the observed survival percentage and its standard error were obtained from Kaplan-Meier analysis, and the standard error from the mean of the predictions was estimated from 10,000 bootstrap samples. Extension of this procedure (using the square root of the sum of the squares of estimated standard errors) was used to test for differences in calibration in between two subgroups of patients. Mean observed *versus* predicted

survival were plotted for quintiles to assess whether calibration varied across the range of predictions and assessed statistically using the Nam-D'Agostino test.<sup>12</sup>

The ability of the colon cancer survival calculators to correctly rank patients with different outcomes, termed discrimination, was quantified using Harrell's concordance index (c-index).<sup>13</sup> Standard errors were obtained from 10,000 bootstrap samples and used to calculate 95% confidence intervals (95% CI) as mean  $\pm$  1.96 x standard error. A concordance index of 0.5 represents random chance, whereas a concordance index of 1.0 represents perfect pairwise ranking from the prediction model. Discriminatory ability of calculators was compared against the AJCC 7<sup>th</sup> edition tumor staging system using a Z test, with standard errors obtained from 10,000 bootstrap samples.

## **RESULTS**

### **Applicability of colon cancer survival calculators in Australian community practice**

Between January 1, 1995 and December 31, 2011, a total of 1,401 Australian patients who underwent curative intent surgery for stage II/III colon cancer were recruited and prospectively followed-up at the participating hospitals. Following surgery, 649 of 1,397 (46.5%) patients with recorded treatment information received adjuvant FP-based chemotherapy. The median follow-up time was 4.8 years (IQR = 2.6 to 6.9 years). Of

1,013 patients with complete 5-year follow-up, 273 (26.9%) patients had a recurrence, 178 (17.6%) patients died from cancer, and 372 (36.7%) patients died from any cause. Supplementary Table 1 summarizes the clinical characteristics and adjuvant treatment details of the cohort.

To evaluate the general applicability of the A!O, MSKCC, MDA and MC colon cancer survival calculators in the setting of Australian community practice, we examined the calculators requirements for input parameters, considered treatment scenarios and predicted outcomes (Table 1). The A!O, MSKCC and MDA calculators were broadly applicable to the entire population of stage II/III patients, while the MC calculator was limited to stage III patients with  $\geq 12$  examined lymph nodes and receiving adjuvant treatment. The calculator with the most comprehensive consideration of treatment scenarios and outcome measures was A!O, providing probabilities for 5-year recurrence-free survival (RFS), cancer-specific survival (CSS) and overall survival (OS) for the settings of no adjuvant treatment and FP-based adjuvant chemotherapy with or without oxaliplatin. Measures of uncertainty for survival predictions were only included in the MSKCC-OS and MC calculators.

Colon cancer survival calculators varied widely in the use and parameterization of clinical input characteristics (Table 1). Parameters included in all calculators comprised patient age, gender, T stage and tumor differentiation/grade. Numbers of lymph nodes examined and positive nodes were considered with variable cut-offs in all

tools except for the MDA calculator. Highly variable consideration was given to other characteristics including N stage, tumor site, lymphovascular and perineural invasion, race, comorbidity, body mass index, ECOG status and CEA level.

With few exceptions (such as for tumor grade in the A!O calculator) the prognosis calculators required complete data for input parameters which limited their utility in the setting of community practice data collection. In our cohort, predictions could be made for 80.5% (1128/1401), 96.0% (1345/1401), 96.6% (1353/1401) and 96.6% (1354/1401) of patients for the A!O model 1, A!O model 2, MSKCC-OS and MDA calculators, respectively (Table 2). The MSKCC-RFS predictor required details on pre-operative CEA levels and ECOG status which was only available for a subset of patients in our cohort; this limited our predictions to 50.8% (712/1401) of patients. Additionally, this calculator required perineural invasion (PNI) information. As this data was not recorded in our cohort, we used imputed data based on the Australian MCO study patient cohort (n=611) and adjusted our subsequent analyses accordingly (Supplementary Methods). Predictions for the stage III patient-focused MC predictor could be made for 59.1% (212/359) of stage III patients with  $\geq 12$  examined lymph nodes and receiving adjuvant treatment. The proportions of patients for whom predictions could be made in our cohort cannot be regarded as a benchmark of calculator applicability as data availability will be dependent on the local context. However, these observations nonetheless illustrate the limitation of calculators requiring complete input data.

## Calibration of colon cancer survival calculators in Australian community patients

To evaluate performance of the A!O, MSKCC, MDA and MC colon cancer survival calculators, trained on either US or clinical trial patient data, in the Australian community setting, we first compared predicted against observed survival outcomes for our cohort (calibration).

For patients with stage II/III colon cancer, the 5-year OS percentages predicted by the A!O model 1 and MSKCC calculators were significantly worse than the observed percentages by -9.9% (95% CI = -13.0% to -6.8%) and -6.5% (95% CI = -9.3% to -3.7%), respectively ( $P < 0.001$  for both comparisons) (Table 2). For 5-year CSS, the predicted outcomes were also poorer, with predictions worse by -9.8% (95% CI = -12.4% to -7.1%) for A!O model 1 and -14.4% (95% CI = -16.7% to -12.0%) for the MSKCC tool ( $P < 0.001$  for both comparisons). Overestimation of risk was further observed for 5-year RFS for both models (A!O model 1: -4.0%, 95% CI = -7.0% to -1.1%; MSKCC: -6.8%, 95% CI = -10.4% to -3.1%;  $P < 0.001$  for both comparisons). The A!O model 2 produced similar results to A!O model 1 (Table 2). The tendency to overestimate risk was maintained when considering stage II and III patients separately, although this did not reach formal statistical significance for RFS predictions among stage II patients for the MSKCC and A!O model 2 calculators (Table 2). For the A!O models, the differences between predicted and observed CSS and RFS percentages were more pronounced for patients with poor survival (Supplementary Figure 1). Overall

similar trends were observed for the independent Australian MCO study patient cohort (N=587; Supplementary Table 4) which could be evaluated for a subset of three predictors (A!O model 2, MSKCC-OS and MDA) given the limited available clinicopathological data; the only notable difference was for the MSKCC-OS calculator which showed good calibration for stage III patients of the MCO cohort (Supplementary Table 5).

Marked differences with respect to calibration were identified when considering patient subgroups as defined by tumor stage and treatment allocation (Table 3). For OS outcomes, overestimation of risk for the A!O models and MSKCC calculators was prominent for stage II patients treated with surgery alone and stage III patients receiving adjuvant therapy (-7.1% to -12.1%;  $P < 0.001$  for all comparisons), while predictions were similar to observed outcomes for high-risk stage II patients receiving adjuvant therapy and stage III patients unfit for or declining adjuvant treatment. For CSS outcomes, A!O models and MDA calculators overestimated risk in most scenarios (-5.9% to -22.4%;  $P < 0.001$  for all comparisons) except for the A!O models for high-risk stage II patients receiving adjuvant treatment (Table 3). For RFS outcomes, overestimation of risk for the A!O models was observed for stage II and III patients treated with surgery alone (-3.5% to -26.8%), while for the MSKCC calculator this was observed for stage III patients irrespective of adjuvant treatment (-10.2% to -15.6%) ( $P < 0.020$  for all comparisons).

In contrast, the stage III patient-focused MC calculator showed good calibration for 5-year OS and 3-year RFS, with predicted and observed outcomes within 3.2% (Table 2). Calibration for the A!O models and MSKCC calculators applied to this patient subgroups remained poor for OS, exhibiting pessimism, although the A!O models showed acceptable calibration for RFS (Supplementary Table 6; Supplementary Figure 2).

### **Calibration of colon cancer survival calculators in patient subgroups**

We further examined whether calculator calibration varied over the recruitment period and among patient subgroups by age, ASA score, ECOG status, gender and tumor location (Figure 1, Supplementary Table 7). Overestimation of risk for colon cancer calculators was similar for patients diagnosed before or after 2006, when FOLFOX treatment became routinely available in Australia (listing by the Pharmaceutical Benefits Scheme was effective 1 December 2005; <http://www.pbs.gov.au>), and similarly no discernible differences were found for patients diagnosed before or after 2000. However, differences in calibration were apparent for the other patient characteristics.

For OS, the MSKCC calculator which does not consider comorbidities showed greater overestimation of risk for patients with lower ECOG status and ASA score, and similar results were found for A!O model 2 in which comorbidity for all patients was entered as “average for age” (7.8% to 11.3%,  $P < 0.028$  for all comparisons). For CSS, the A!O model 1 showed evidence of differential calibration, with greater overestimation

of risk for higher ECOG status (-6.2%,  $P = 0.047$ ). For RFS, the A!O model 1 showed differences in calibration between patient subgroup for age, ASA score and ECOG status, with a greater overestimation of risk for the elderly and individuals with high ASA score and ECOG status; a difference by age was also observed for A!O model 2 (-8.6% to -13.6%,  $P < 0.012$  for all comparisons). The MSKCC calculator showed a difference in calibration for gender and tumor location, with larger overestimation of risk for female patients and left-sided cancers (9.1% and 7.6%, respectively,  $P < 0.031$  for both comparisons).

No significant differences in calibration were observed for the stage III patient-focused MC calculator for OS or RFS, although group sizes were limited (Figure 1, Supplementary Table 7).

### **Comparison of discrimination of colon cancer survival calculators against the AJCC 7<sup>th</sup> edition tumor staging system**

We next examined the ability of the calculators to correctly rank pairs of patients by outcome (discrimination), as measured via the concordance index (c-index), in comparison to the AJCC 7<sup>th</sup> edition tumor staging system (AJCC7).

For patients with stage II/III colon cancer, the A!O model 1 and MSKCC calculators showed significantly improved discrimination for OS as compared to AJCC7 staging ( $P < 0.001$ ; Figure 2, Supplementary Table 8). Discriminatory ability for both

calculators was good with a c-index of 74.0% (95% CI = 71.2% - 76.8%) for the A!O model 1 and 71.7 (95% CI = 69.1 -74.3) for the MSKCC calculator. Superior discrimination for OS was maintained when considering stage II and III patients separately. For CSS, the A!O model 1 maintained improved discriminatory ability for stage III patients (c-index = 70.9%, 95% CI = 65.2% - 76.5%), whilst the improvement in discrimination from the MDA calculator from AJCC7 staging did not reach statistical significance. Results for A!O model 2 mirrored findings from A!O model 1.

No significant improvement in discrimination was observed for the stage III patient-focused MC calculator for OS or RFS, although group sizes were limited (Figure 2, Supplementary Table 9).

## **DISCUSSION**

The present study evaluated the transferability to the Australian community setting of five web-based survival calculators for patients with stage II/III colon cancer, A!O, MSKCC-OS, MSKCC-RFS, MDA and MC, each developed using data from either US or clinical trial patient populations. Calculators varied considerably in the inclusion and parameterization of clinical input characteristics, and all were found to exhibit limitations in applicability due to requirements for complete data, restriction to patient subgroups, and/or lack of specification of adjuvant treatment. Only two tools, MSKCC-OS and MC,

considered uncertainty of survival predictions. Our assessment of calculators for 1,401 stage II/III patients showed that while all tools exhibited similar or improved discrimination as compared to AJCC7 staging, all but the stage III patient-focused MC calculator displayed deficiencies in calibration in the Australian healthcare setting.

Calculators applicable across stage II/III patients (A!O, MSKCC and MDA) significantly overestimated the risk of death, and to a lesser extent the risk of recurrence, in our Australian community cohort. In addition, we observed substantial heterogeneity with respect to calibration by patient stage and treatment. For OS outcomes, the A!O and MSKCC calculators overestimated the risk of death for stage II patients treated with surgery alone and stage III patients receiving adjuvant therapy, while predictions closely matched observed outcomes for high-risk stage II patients receiving adjuvant therapy and stage III patients unfit for or declining adjuvant treatment. For CSS outcomes, A!O models and MDA calculators overestimated risk in most scenarios except for the A!O models for high-risk stage II patients receiving adjuvant treatment. In addition, there was evidence for differences in calibration between patient subgroups, including by age, gender, tumor location, ECOG status and ASA score.

Overestimation of the risk of death for the A!O calculator has previously been reported by Papamichael *et al.* in an evaluation of 2,967 elderly patients (age 70+) with stage III colon cancer enrolled into multiple clinical trials captured by the ACCENT

database when assuming “average for age” co-morbidities, while closely matching observed and predicted outcomes were observed for recurrence.<sup>14</sup> However, when assuming “minor” co-morbidities Papamichael *et al.* found that the calculator yielded matching predictions for risk of death but underestimated the risk of recurrence, highlighting the challenge of assessing the A!O tool when individual patient co-morbidity assessments are not available. In contrast, when considering patient subgroups by tumor stage and treatment, Gill *et al.* found the A!O calculator showed reasonable calibration for stage III disease assuming either “minor problems” or “average for age” for comorbidity status in two cohorts, 2,033 stage II/III patients from the British Columbia Colorectal Cancer Outcomes Unit database and 1,729 patients who were enrolled in the NCCTG 894651 and NCCTG 914653 adjuvant therapy trials. However, the A!O calculator was found to overestimate recurrence free and overall survival for patients with stage II disease who received 5-FU, in particular for the population-based cohort.<sup>9</sup> For the MSKCC-OS calculator, Liu *et al.* observed good calibration in a cohort of 985 Chinese patients, but no subgroup analysis by tumor stage and treatment was performed.<sup>8</sup>

In our Australian cohort, only the MC calculator appeared to show acceptable calibration for OS and RFS, although sample size was limited due to restricted applicability of the tool to stage III patients with  $\geq 12$  examined lymph nodes and receiving adjuvant treatment. An external validation of the MC calculator was reported

as part of the original model development with good calibration, but this was for another international clinical trial cohort (NSABP trial C-08).<sup>7</sup>

Multiple reasons may account for the miscalibration of US patient-based colon cancer survival calculators in the Australian community setting. Demographic disparities exist between the US and Australian population including socio-economic background, ethnic differences, access and attitudes to healthcare which may influence baseline survival. A differential impact of these factors depending on the clinical scenario may further explain the observed differences in the extent of miscalibration between treatment subgroups by tumor stage. The A!O, MSKCC-OS and MDA calculators were trained using data from the SEER database which over-represents certain ethnic groups compared to the Australian population, such as African, Native American/Hawaiian and some South East Asian groups<sup>15</sup> who have documented differences in baseline CRC outcomes.<sup>16, 17</sup> Australians have an overall life-expectancy of 3-4 years greater than the US population,<sup>18</sup> and Australia ranked higher (third) than the United States (ninth) for CRC survival among a group of 23 OCED countries.<sup>19</sup> Some data fields captured in the SEER database are known to be problematic; for example, date of death is considered robust, while cause of death is considered of poor quality.<sup>20</sup> Relapse data are not available in the SEER database and patients receiving adjuvant therapy are not identified, requiring calculators to make assumptions based on external evidence and respective mathematical adjustments. The MSKCC-RFS calculator was developed

using a relatively small institutional cohort of 1,320 patients which may be associated with a local bias. Our findings highlight the need to retrain survival calculators using local population data to minimize miscalibration.

The discriminatory performance of the survival calculators was generally better than the corresponding discrimination from the AJCC7 classification scheme, highlighting the potential of the multi-variable models for prediction over routinely used category-based ranking schemes. As anticipated for models including patient demographic features and/or treatment in addition to traditional staging, the most pronounced improvement in prediction was evident for OS outcomes which are significantly impacted by comorbidities. For CSS and RFS outcomes, the discrimination from the more sophisticated prediction models were similar or marginally better than the AJCC7 scheme, with a greater difference for stage III as compared to stage II patients. In their study using Chinese patient data, Liu *et al.* observed the MSKCC-OS calculator to give superior discriminatory performance than AJCC7 staging.<sup>8</sup> Renfro *et al.* also reported better performance for the stage III patient-focused MC calculator as compared to AJCC7 staging in their study,<sup>7</sup> but we did not observe significant improvement for our cohort although the group size was limited for this patient subset.

While discriminatory performance of survival calculators was greater than AJCC7 staging for OS, accuracy for individual patient predictions nonetheless remained limited, highlighting the need for further predictor development. At the patient level, calculators

correctly predicted outcome (scored as correct if the patients predicted probability of 5-year OS was >50% and the patient was alive, and accordingly for death probability) in 72%, 71% and 71% of cases for the A!O (1), A!O (2) and MSKCC-OS tools, respectively. For the stage III-patient focused MC calculator, 69% of patient outcomes were correctly predicted. Extended use of clinical information, such as details of treatment for recurrence may improve model performance. Furthermore, incorporation of molecular prognostic indicators such as microsatellite instability, *KRAS/BRAF* mutation and chromosome instability, or stroma and immune markers such as the Immunoscore,<sup>21-25</sup> has potential for refining predictions.

A strength of our study is the multi-institutional nature of the cohort, originating from two Melbourne Health Services with prospective point-of-care recording of patient clinical and treatment data and follow-up according to national guidelines. Detailed collection of patient performance status and geriatric information further enabled us to assess a comorbidity classification that was reflective of routine practice. Patients who were partly treated at Melbourne hospitals were mainly referred from regional hospitals. As a result, the population includes a wide representation of the community, reducing selection bias inherent to single-institute cohorts. Overestimation of risk was further replicated in an independent Australian cohort which could be examined for a subset of the survival calculators.

This study also has some limitations. For analysis of CSS, cause of death was extracted from death certificates of patients which is not always accurate. For a subset of 43% of patients, enrolment was prior to 2006 when adoption of combination chemotherapy regimens occurred for both adjuvant and palliative treatment of colon cancer. Surgical techniques and guidelines have also evolved over the study period. However, we observed only minor differences in model calibration between patients enrolled pre-and post-2006. A challenge with evaluating the A!O model is the lack of standardized definition of comorbidity status. We based our definitions on the observed distributions of age, ASA score and ECOG status, but also evaluated a model where all patients were treated as “average for age” with similar results. Notably, the A!O calculator is presently undergoing updating and not available online. Another limitation of this study is that our cohort would have included some treatment heterogeneity; with increasing age, patients are less often treated according to treatment guidelines including surgical procedures and chemotherapy regimens. There was a proportion of missing data for our cohort, which precluded application of calculators to respective patient subsets. Within each Dukes' stage, none of the differences in survival (overall, recurrence-free and cancer-specific) between the patient groups with and without predictions (due to incomplete data) reached statistical significance, indicating little selection bias due to missing data. For the MSKCC-RFS calculator, perineural invasion had to be imputed based on feature mapping against an independent cohort; although

Brier scores of 0.116 and 0.117 were obtained for cross-validation in two independent runs, this remains a limitation. Our comparison of calculator discrimination was against the AJCC 7<sup>th</sup> edition staging system, as the recently introduced AJCC 8<sup>th</sup> edition staging system could not be retrospectively applied to our cohort.

In conclusion, our assessment of transferability of web-based colon cancer survival calculators trained on data from either US or clinical trial patients in a community-based Australian cohort identified deficiencies in calibration for four out of five evaluated tools (A!O, MSKCC-OS, MSKCC-RFS and MDA). Calibration appeared acceptable for the MC calculator, although sample size was small due to tool applicability restricted to stage III patients with  $\geq 12$  examined lymph nodes and receiving adjuvant treatment. The discriminatory performance of the prognosis predictors was generally better than the corresponding discrimination from the AJCC7 classification schemes, in particular for OS, highlighting the potential for risk assessment and guiding interventions after surgery. Our findings highlight the need for validation and tailoring of cancer survival calculators prior to implementation across healthcare systems.

#### **CONFLICT OF INTEREST**

The authors declare no conflict of interest.

## **FUNDING SUPPORT**

This work was supported by grant #1120882, awarded through the Priority-driven Collaborative Cancer Research Scheme and funded by Cancer Australia, and the Victorian Government's Operational Infrastructure Support Program. O.M.S. is a NHMRC Senior Research Fellow (APP1136119).

## **AUTHORS' CONTRIBUTIONS**

Study design: R.N.J. and O.M.S.; collection of patient clinicopathological data: M.C., I.T.J., R.L.W., N.J.H and P.G.; data assembly, acquisition of survival predictions and data analysis: R.N.J. and O.M.S.; manuscript preparation and approval: R.N.J, M.C., I.T.J., R.L.W., N.J.H, P.G. and O.M.S.

## **ACKNOWLEDGEMENTS**

We thank all patients who participated in this study and colleagues who undertook sample and clinical data collection. We acknowledge the Victorian Cancer Biobank for the provision of specimens and BioGrid Australia for providing clinical data.



## REFERENCES

1. Engstrom PF, Arnoletti JP, Benson AB, 3rd, et al. National Comprehensive Cancer N. NCCN Clinical Practice Guidelines in Oncology: colon cancer. *J Natl Compr Canc Netw* 2009;7: 778-831.
2. Edge SB, Byrd DR, Compton CC, et al., eds. *AJCC cancer staging manual*, 7th ed. New York: Springer, 2010.
3. Amin MB, Edge S, Greene F, et al., eds. *AJCC cancer staging manual*, 8th ed. New York: Springer, 2017.
4. Weiser MR, Landmann RG, Kattan MW, et al. Individualized prediction of colon cancer recurrence using a nomogram. *J Clin Oncol* 2008;26: 380-5.
5. Weiser MR, Gonen M, Chou JF, et al. Predicting survival after curative colectomy for cancer: individualizing colon cancer staging. *J Clin Oncol* 2011;29: 4796-802.
6. Chang GJ, Hu CY, Eng C, et al. Practical application of a calculator for conditional survival in colon cancer. *J Clin Oncol* 2009;27: 5938-43.
7. Renfro LA, Grothey A, Xue Y, et al. Adjuvant Colon Cancer Endpoints G. ACCENT-based web calculators to predict recurrence and overall survival in stage III colon cancer. *J Natl Cancer Inst* 2014;106.
8. Liu M, Qu H, Bu Z, et al. Validation of the Memorial Sloan-Kettering Cancer Center Nomogram to Predict Overall Survival After Curative Colectomy in a Chinese Colon Cancer Population. *Ann Surg Oncol* 2015;22: 3881-7.

9. Gill S, Loprinzi C, Kennecke H, et al. Prognostic web-based models for stage II and III colon cancer: A population and clinical trials-based validation of numeracy and adjuvant! online. *Cancer* 2011;117: 4155-65.
10. Kazem MA, Khan AU, Selvasekar CR. Validation of nomogram for disease free survival for colon cancer in UK population: A prospective cohort study. *Int J Surg* 2016;27: 58-65.
11. Collins IM, Kelleher F, Stuart C, et al. Clinical decision aids in colon cancer: a comparison of two predictive nomograms. *Clin Colorectal Cancer* 2012;11: 138-42.
12. D'Agostino RB, Nam BH. Evaluation of the performance of survival analysis models: discrimination and calibration measures. In: Balakrishnan N, Rao CR, eds. *Advances in survival analysis*. Amsterdam: Elsevier North-Holland, 2004: 1-25.
13. Harrell FE, Jr., Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. *JAMA* 1982;247: 2543-6.
14. Papamichael D, Renfro LA, Matthaiou C, et al. Adjuvant Colon Cancer Endpoints G. Validity of Adjuvant! Online in older patients with stage III colon cancer based on 2967 patients from the ACCENT database. *J Geriatr Oncol* 2016;7: 422-9.

15. Central Intelligence Agency. *The world factbook*. Washington, DC: Central Intelligence Agency, 2017. <https://www.cia.gov/library/publications/resources/the-world-factbook/index.html>
16. Doubeni CA, Field TS, Buist DS, et al. Racial differences in tumor stage and survival for colorectal cancer in an insured population. *Cancer* 2007;109: 612-20.
17. Phatak UR, Kao LS, Millas SG, et al. Interaction between age and race alters predicted survival in colorectal cancer. *Ann Surg Oncol* 2013;20: 3363-9.
18. World Health Organization. World health statistics 2018: monitoring health for the SDGs, sustainable development goals Geneva: World Health Organization, 2018.
19. OECD. *Health at a Glance 2015*. OECD Publishing, Paris. 2015. [http://dx.doi.org/10.1787/health\\_glance-2015-en](http://dx.doi.org/10.1787/health_glance-2015-en).
20. Warren JL, Harlan LC, Fahey A, et al. Utility of the SEER-Medicare data to identify chemotherapy use. *Med Care* 2002;40: IV-55-61.
21. Sinicrope FA, Shi Q, Smyrk TC, et al. Molecular markers identify subtypes of stage III colon cancer associated with patient outcomes. *Gastroenterology* 2015;148: 88-99.
22. Mouradov D, Domingo E, Gibbs P, et al. Survival in stage II/III colorectal cancer is independently predicted by chromosomal and microsatellite instability, but not by specific driver mutations. *Am J Gastroenterol* 2013;108: 1785-93.

23. Williams DS, Mouradov D, Jorissen RN, et al. Lymphocytic response to tumour and deficient DNA mismatch repair identify subtypes of stage II/III colorectal cancer associated with patient outcomes. *Gut* 2018.
24. Hutchins GGA, Treanor D, Wright A, et al. Intratumoral stromal morphometry predicts disease recurrence but not response to 5-fluorouracil-results from the QUASAR trial of colorectal cancer. *Histopathology* 2018;72: 391-404.
25. Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet* 2018;391: 2128-39.

## FIGURE LEGENDS

**Figure 1.** Calibration of survival calculators in an Australian community cohort of patients with stage II/III colon cancer for subgroups by age, ASA score, ECOG status, gender (female relative to male), primary tumor location (right-sided relative to left-sided) and year of diagnosis. The calibration difference is calculated as predicted minus observed survival percentages; horizontal lines indicate 95% confidence intervals. Calibration is for 3-year survival for the Mayo Clinic recurrence-free survival calculator, and for 5-year survival for all other calculators. N1 and N2 indicate the subgroup sizes for the feature of interest (e.g. Age  $\geq$  80 years) and the corresponding comparison group (e.g. Age  $<$  80 years), respectively. OS, overall survival; RFS, recurrence-free survival; CSS, cancer-specific survival; A!O (1), Adjuvant! Online model 1; A!O (2), Adjuvant! Online model 2; MSKCC, Memorial Sloan Kettering Cancer Center; MDA, MD Anderson Cancer Center; MC, Mayo Clinic.

**Figure 2.** Discrimination of survival calculators (black) in an Australian community cohort of patients with stage II/III colon cancer, compared with the AJCC 7<sup>th</sup> edition tumor staging system (grey). Point estimates for c-indices are shown as circles and horizontal lines indicate 95% confidence intervals. OS, overall survival; RFS,

recurrence-free survival; CSS, cancer-specific survival; A!O (1), Adjuvant! Online model 1; A!O (2), Adjuvant! Online model 2; MSKCC, Memorial Sloan Kettering Cancer Center; MDA, MD Anderson Cancer Center; MC, Mayo Clinic.