



## Predicting Survival in Colorectal Liver Metastasis: Time for New Approaches

Georgios Antonios Margonis, MD, PhD<sup>1</sup>, Nikolaos Andreatos, MD<sup>2</sup>, and Murray F. Brennan, MD<sup>1</sup>

<sup>1</sup>Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY; <sup>2</sup>Department of Internal Medicine and Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH

Prognostic models are important tools in forecasting survival outcomes. Prognostication facilitates patient–physician discussions and serves as a critical first step in identifying patient subgroups and potentially individuals with a markedly distinct prognosis. Although such prognostic stratification cannot directly predict treatment benefit, it can help to inform clinical decision-making. In trial design, by directing interventions toward appropriate high-risk groups, prognostic indicators can maximize the likelihood of a clinically meaningful trial. Accurate indices often can allow inclusion of patients most likely to benefit and help to reduce trial cohort size.

One of the first prognostic models for patients with resectable colorectal liver metastasis (CRLM) was developed by Fong et al.<sup>1</sup> in 1999. In a cohort of 1001 patients from the Memorial Sloan Kettering Cancer Center (MSKCC) treated between 1985 and 1998, the authors selected five independent, preoperatively available predictors of survival for incorporation into a clinical risk score (CRS). The weight assigned to each prognostic factor was deliberately the same (although the relative hazards of death varied somewhat between predictors), transforming the model into an easy-to-calculate, 0- to 5-point score. Importantly, the CRS successfully distributed patients across a wide range of outcomes, with 5-year survival rates ranging from 60% for patients with 0 points to 14% for patients with 5 points, thus helping to identify potentially

actionable patient subgroups. The original publication discussed the possible importance of neoadjuvant chemotherapy for patients with high CRS, which currently is standard practice. This description of combined characteristics associated with poor survival can render a model clinically relevant even if the discriminatory ability as reflected by the concordance index (c-index) and area under the curve (AUC) is not high. The CRS remained a valuable clinical tool for two decades.

With the advent of modern chemotherapy and broadened surgical indications for patients with CRLM, a number of reports appropriately began to question the discriminatory power of the CRS, with AUCs ranging from 0.53 to 0.68 in external validation cohorts.<sup>2</sup> Although attempts to develop updated risk scores were made, gains in discriminatory ability were modest at best. A model recommended by Rees et al.<sup>3</sup> reported a promising AUC of 0.8, but that decreased, ranging from 0.59 to 0.66 in the external validation.<sup>2</sup> It slowly became accepted that the combination of biomarkers reflecting the underlying tumor biology and traditional risk factors might increase discriminatory power. Vauthey et al.<sup>4</sup> demonstrated that in addition to a well-established predictive role indicating resistance to anti-epidermal growth factor receptor agents, the presence of KRAS mutations had a prognostic role among patients with CRLM.

Subsequent studies by a group at Johns Hopkins and others confirmed these findings and raised the possibility that KRAS mutation status may be an appropriate candidate for incorporation into clinical risk scores.<sup>5</sup> In 2018, two such prognostic models were reported: the genetic and morphological evaluation (GAME) score, developed at Johns Hopkins and validated at MSKCC, and the modified CRS (m-CRS), developed at MD Anderson and validated in an international cohort.<sup>6, 7</sup> The GAME score used a weighted

approach, whereas the m-CRS retained the same statistical design as the original CRS, assigning 1 point each to the presence of primary tumor lymph node metastasis, KRAS mutation, and largest CRLM diameter (> 5 cm). The GAME score and the m-CRS had slightly better discriminatory power than the original CRS, achieving c-indexes of 0.625 to 0.645 and 0.69, respectively.

In the current study, Paredes et al.<sup>8</sup> built on these two models by developing an alternative clinical score (a-CS), which incorporated KRAS status and preoperatively available clinical risk factors in a cohort of 1406 patients derived from five collaborating institutions in the United States and Europe. The model included 11 prognostic factors and had good discriminatory ability to determine 1-, 3-, and 5-year recurrence-free survival (RFS) for patients with known KRAS status in the validation cohort [AUC: 0.642, 0.660 and 0.667, respectively (for patients with known KRAS status)]. The AUC confidence intervals were narrow, indicating good reliability, and the model was well-calibrated. Importantly, the discriminatory ability of the a-CS surpassed that of both the CRS and the m-CRS. The a-CS differed from these models by incorporating additional prognostic factors, assigning different weights to incorporated predictors and using a bootstrap aggregation-based (“bagging”) approach to the analysis. These features likely underlie the improved discriminatory ability of the model, which because it included a sizeable minority of patients from outside the United States in the study population will hopefully prove to be applicable internationally.

Notably, the m-CRS was originally created on the basis of risk factors for poor overall survival (OS), rather than RFS.<sup>7</sup> Although Brudvik et al.<sup>7</sup> demonstrated that the m-CRS is an adequate predictor of RFS in addition to OS, the correlation between these two survival metrics is known to be imperfect. Further analysis is required to establish whether the prognostic advantage demonstrated by the a-CS over the m-CRS in terms of RFS is also applicable to OS, a finding that would enhance the impact of the current study.

Compared with the CRS and m-CRS, the current model incorporated the following additional prognostic factors: age, sex, primary tumor location, primary tumor T stage, and receipt of pre-hepatectomy chemotherapy. Pre-hepatectomy chemotherapy and primary tumor location had the largest weight by far and likely contributed most to the model’s improved performance. Carcinoembryonic antigen (CEA) and disease free interval (DFI) (with the same cutoffs used in the CRS) also had considerable prognostic impact, whereas age and sex contributed only marginally. The strong association of pre-hepatectomy chemotherapy with shorter RFS has been reported previously by Margonis et al.<sup>9</sup> and others and, as noted by Paredes et al.,<sup>8</sup> is likely secondary to the considerable underlying disease burden that leads to patients treated initially with chemotherapy.<sup>10</sup>

Nonetheless, the retrospective design of the study and the lack of detailed information on chemotherapy indications render the interpretation of this association a matter of conjecture.

In recent years, primary tumor location has gradually emerged as an important prognostic and is thought to serve as a proxy for underlying tumor biology. This was not known at the time the CRS was developed.<sup>11</sup> The inclusion of KRAS mutation status resulted in a relatively modest difference in the AUC. Although this is consistent with recent data suggesting that the prognostic importance of KRAS mutation status may have been overestimated, the impact of other RAS pathway mutations was not assessed.<sup>9, 12</sup> Given that the complexity of tumor biology is inadequately captured by a single gene-based biomarker, these findings are not surprising. The use of comprehensive gene panels to simultaneously capture mutations in multiple genetic loci of possible prognostic significance (e.g., BRAF) likely is a necessary step for the development of more robust prognostic biomarkers. The rapid proliferation of next-generation sequencing technologies should render this a possibility.

Weighting prognostic factors by their respective prognostic impact is intuitively sensible and likely to contribute to prognostic accuracy. The bootstrap aggregation algorithm used by the authors is an ensemble method that trains multiple different models from a single training dataset and averages their predictions. This approach generally produces more accurate predictions than the individual models, thus improving the stability of high-variance models. Although logistic regression does not traditionally belong to the latter (e.g., decision trees are high-variance machine-learning algorithms), “bagging” likely contributed to the narrow confidence intervals observed. The model’s superior discriminatory ability probably stemmed largely from the incorporation of multiple prognostic factors and the use of appropriate prognostic weighting.

Paredes et al.<sup>8</sup> successfully developed a model with better discriminatory ability than both the CRS and the more recent m-CRS by successfully incorporating multiple predictors according to prognostic weight and maximizing the robustness of traditional regression via innovative use of bootstrap aggregation. Whereas these results demonstrate the potential of well-known prognostic factors combined with new analytic approaches, they also confirm the current limitations in forecasting survival among patients with CRLM. Improving AUC substantially beyond the 0.7 range for patients with CRLM likely will require the discovery of new and novel prognostic biomarkers as well as the inclusion of interactions between known prognostic factors.<sup>12</sup> For example, it has recently been reported that the poor prognostic impact of a right-sided primary tumor among patients with CRLM may manifest

only in the absence of KRAS mutations.<sup>11</sup> Assessing each factor in isolation would have failed to detect this interplay, thus emphasizing the need to recognize that combination indices are important and additive to their individual components.

Artificial intelligence-based methods are ideal for capturing such interactions, which the “bagging” technique does not address. Other approaches to modeling complex interactions between prognostic variables with the aid of Bayesian methodology have received little emphasis to date and deserve further study.<sup>13</sup> Assuming that interactions are difficult and impractical to capture, a different strategy would be to try limiting their impact by separating patients into smaller biologically and/or clinically homogeneous groups within which the effect of prognostic factors is more likely to be uniform. This approach has been used successfully for soft tissue sarcoma, in which the original nomograms developed nearly 20 years ago with “all-comers” have been subdivided gradually into site-, histology- and post-recurrence-specific versions, with considerable improvement in prognostic accuracy.<sup>14–16</sup> Nonetheless, this approach remains imprecise for the individual patient because although the effect of interactions is reduced, it can never be eliminated entirely, and the nomograms fail to account for it.

Future models with improved discriminatory ability may not automatically be of clinical interest unless they successfully identify discrete patients or patient subgroups that could potentially inform management and clinical trial design as was originally the intent of the CRS. Although the current model by Paredes et al.<sup>8</sup> allows us to anticipate the likelihood of recurrence with improved accuracy, predicting the site and distribution of recurrent disease may help to guide management further. Specifically, the ability to identify patients that may be technically but not “biologically” resectable due to the high likelihood of multifocal extrahepatic recurrence as well as patients at high-risk for isolated intrahepatic recurrence who may be candidates for aggressive re-resection or liver-directed treatment (e.g., hepatic arterial infusion therapy) would add a predictive (rather than a merely prognostic) dimension to the a-CS and should be considered in future updated versions of the model.

It has taken two decades to get this far. We might hope that with improved identification of relevant biologic markers and improved artificial intelligence techniques we shall make greater progress for the individual patient in the next decade.<sup>17</sup>

## REFERENCES

1. Fong Y, Fortner J, Sun RL, Brennan MF, Blumgart LH. Clinical score for predicting recurrence after hepatic resection for

- metastatic colorectal cancer: analysis of 1001 consecutive cases. *Ann Surg.* 1999;230:309–18. (discussion 18–21).
2. Mahar AL, Compton C, Halabi S, Hess KR, Weiser MR, Groome PA. Personalizing prognosis in colorectal cancer: a systematic review of the quality and nature of clinical prognostic tools for survival outcomes. *J Surg Oncol.* 2017;116:969–82.
3. Rees M, Tekkis PP, Welsh FK, O'Rourke T, John TG. Evaluation of long-term survival after hepatic resection for metastatic colorectal cancer: a multifactorial model of 929 patients. *Ann Surg.* 2008;247:125–35.
4. Vauthey JN, Zimmiti G, Kopetz SE, Shindoh J, Chen SS, Andreou A, et al. RAS mutation status predicts survival and patterns of recurrence in patients undergoing hepatectomy for colorectal liver metastases. *Ann Surg.* 2013;258:619–26. (discussion 26–7).
5. Margonis GA, Spolverato G, Kim Y, Karagkounis G, Choti MA, Pawlik TM. Effect of KRAS mutation on long-term outcomes of patients undergoing hepatic resection for colorectal liver metastases. *Ann Surg Oncol.* 2015;22:4158–65.
6. Margonis GA, Sasaki K, Gholami S, Kim Y, Andreatos N, Rezaee N, et al. Genetic and morphological evaluation (GAME) score for patients with colorectal liver metastases. *Br J Surg.* 2018;105:1210–20.
7. Brudvik KW, Jones RP, Giuliani F, Shindoh J, Passot G, Chung MH, et al. RAS mutation clinical risk score to predict survival after resection of colorectal liver metastases. *Ann Surg.* 2019;269:120–6.
8. Paredes A, Hyer, J M, Tsilimigras, D. A novel machine-learning approach to predict recurrence following resection of colorectal liver metastases. *Ann Surg Oncol.* 2020. (in press).
9. Margonis GA, Buettner S, Andreatos N, Kim Y, Wagner D, Sasaki K, et al. Association of BRAF mutations with survival and recurrence in surgically treated patients with metastatic colorectal liver cancer. *JAMA Surg.* 2018;153:e180996.
10. Ito H, Are C, Gonen M, D'Angelica M, Dematteo RP, Kemeny NE, et al. Effect of postoperative morbidity on long-term survival after hepatic resection for metastatic colorectal cancer. *Ann Surg.* 2008;247:994–1002.
11. Margonis GA, Amini N, Buettner S, Kim Y, Wang J, Andreatos N, et al. The prognostic impact of primary tumor site differs according to the KRAS mutational status: a study by the International genetic consortium for colorectal liver metastasis. *Ann Surg.* 2019. <https://doi.org/10.1097/SLA.0000000000003504>.
12. Margonis GA, Sasaki K, Kim Y, Samaha M, Buettner S, Amini N, et al. Tumor biology rather than surgical technique dictates prognosis in colorectal cancer liver metastases. *J Gastrointest Surg.* 2016;20:1821–9.
13. Forsberg JA, Healey JH, Brennan MF. A probabilistic analysis of completely excised high-grade soft tissue sarcomas of the extremity: an application of a Bayesian belief network. *Ann Surg Oncol.* 2012;19:2992–3001.
14. Kattan MW, Leung DH, Brennan MF. Postoperative nomogram for 12-year sarcoma-specific death. *J Clin Oncol.* 2002;20:791–6.
15. Dalal KM, Kattan MW, Antonescu CR, Brennan MF, Singer S. Subtype specific prognostic nomogram for patients with primary liposarcoma of the retroperitoneum, extremity, or trunk. *Ann Surg.* 2006;244:381–91.
16. Kattan MW, Heller G, Brennan MF. A competing-risks nomogram for sarcoma-specific death following local recurrence. *Stat Med.* 2003;22:3515–25.
17. Bertsimas D, Dunn J. Optimal classification trees. *Mach Learn.* 2017;106:1039–82.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.