HHS Public Access

Author manuscript

Lancet Digit Health. Author manuscript; available in PMC 2021 September 23.

Published in final edited form as:

Lancet Digit Health. 2021 September ; 3(9): e565–e576. doi:10.1016/S2589-7500(21)00104-7.

The use of a next-generation sequencing-derived machine-learning risk-prediction model (OncoCast-MPM) for malignant pleural mesothelioma: a retrospective study

Marjorie G Zauderer,
Thoracic Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Department of Medicine, Weill Cornell Medical College, New York, NY, USA

Axel Martin,
Biostatistics Service, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Jacklynn Egger,
Druckenmiller Center for Lung Cancer Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Hira Rizvi,
Druckenmiller Center for Lung Cancer Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Michael Offin,
Thoracic Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Department of Medicine, Weill Cornell Medical College, New York, NY, USA

Andreas Rimner,
Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Prasad S Adusumilli,
Thoracic Surgery, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Correspondence to: Dr Marjorie G Zauderer, Thoracic Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA zauderem@mskcc.org.

Contributors

MGZ, AM, MGK, and RS were responsible for the conception and design of the study. MGZ, AM, JE, HR, and RS provided administrative support. MGZ, AR, VWR, PSA, JLS, and ML provided study materials or assisted with patient recruitment. MGZ, AM, JE, HR, and RS were responsible for data collection and data assembly. MGZ, AM, JE, HR, MO, MGK, JLS, ML, and RS were responsible for data analysis and data interpretation. All authors had access to all the raw datasets, and were responsible for manuscript writing, final approval of the manuscript, and all aspects of the work. MGZ and RS verified the data. MGZ, RS, and AM had access to all the data. MGZ was responsible for the decision to submit the manuscript.

Data sharing

Individual participant data that underlie the results reported in this Article, after de-identification will be made available with publication upon request to the corresponding author. The source code for the OncoCast analysis is available.

Valerie W Rusch,
Thoracic Surgery, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Mark G Kris,
Thoracic Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Department of Medicine, Weill Cornell Medical College, New York, NY, USA

Jennifer L Sauter,
Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Marc Ladanyi,
Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Ronglai Shen
Biostatistics Service, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Summary

Background—Current risk stratification for patients with malignant pleural mesothelioma based on disease stage and histology is inadequate. For some individuals with early-stage epithelioid tumours, a good prognosis by current guidelines can progress rapidly; for others with advanced sarcomatoid cancers, a poor prognosis can progress slowly. Therefore, we aimed to develop and validate a machine-learning tool—known as OncoCast-MPM—that could create a model for patient prognosis.

Methods—We did a retrospective study looking at malignant pleural mesothelioma tumours using next-generation sequencing from the Memorial Sloan Kettering Cancer Center-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT). We collected clinical, pathological, and routine next-generation sequencing data from consecutive patients with malignant pleural mesothelioma treated at the Memorial Sloan Kettering Cancer Center (New York, NY, USA), as well as the MSK-IMPACT data. Together, these data comprised the MSK-IMPACT cohort. Using OncoCast-MPM, an open-source, web-accessible, machine-learning risk-prediction model, we integrated available data to create risk scores that stratified patients into low-risk and high-risk groups. Risk stratification of the MSK-IMPACT cohort was then validated using publicly available malignant pleural mesothelioma data from The Cancer Genome Atlas (ie, the TCGA cohort).

Findings—Between Feb 15, 2014, and Jan 28, 2019, we collected MSK-IMPACT data from the tumour tissue of 194 patients in the MSK-IMPACT cohort. The median overall survival was higher in the low-risk group than in the high-risk group as determined by OncoCast-MPM (30·8 months [95% CI 22·7–36·2] *vs* 13·9 months [10·7–18·0]; hazard ratio [HR] 3·0 [95% CI 2·0–4·5]; p<0·0001). No single factor or gene alteration drove risk differentiation. OncoCast-MPM was validated against the TCGA cohort, which consisted of 74 patients. The median overall survival was higher in the low-risk group than in the high-risk group (23·6 months [95% CI 15·1–28·4] *vs* 13·6 months [9·8–17·9]; HR 2·3 [95% CI 1·3–3·8]; p=0·0019). Although stage-based

risk stratification was unable to differentiate survival among risk groups at 3 years in the MSK-IMPACT cohort (31% for early-stage disease *vs* 30% for advanced-stage disease; p=0·90), the OncoCast-MPM-derived 3-year survival was significantly higher in the low-risk group than in the high-risk group (40% *vs* 7%; p=0·0052).

Interpretation—OncoCast-MPM generated accurate, individual patient-level risk assessment scores. After prospective validation with the TCGA cohort, OncoCast-MPM might offer new opportunities for enhanced risk stratification of patients with malignant pleural mesothelioma in clinical trials and drug development.

Funding—US National Institutes of Health/National Cancer Institute.

Introduction

Malignant pleural mesothelioma is an uncommon but fatal disease.[1,2] Global estimates suggest that up to 43 000 deaths per year are attributable to mesothelioma. The median overall survival for patients with malignant pleural mesothelioma is 9–17 months for all stages.[2] Two prognostic scoring systems exist: the Cancer and Leukemia Group B (CALGB) index[3] and the European Organisation for Research and Treatment of Cancer (EORTC) index.[4] Both scoring systems were developed retrospectively based on data from clinical trials during a period that predated approved therapies by the US Food and Drug Administration (FDA) for the treatment of malignant pleural mesothelioma, and these therapies are not routinely used because they rely on subjective features (eg, chest pain) and laboratory tests that are not regularly assessed (eg, lactate dehydrogenase).

Histology and stage are the current considerations for risk stratification in patients with malignant pleural mesothelioma.[3,5–7] Unfortunately, imaging can underestimate the disease burden,[7] and aggressive types of disease exist within the more indolent epithelioid subtype, which limits the prognostic accuracy of these characteristics.[7,8] Abdel-Rahman applied the staging method used by the eighth American Joint Committee on Cancer (AJCC) to 5382 patients with malignant pleural mesothelioma and determined that prognostic performance was poor, and that better staging systems are needed for patients with this disease.[9] Several studies have tried to further refine histological classification to better understand the heterogeneity of epithelioid disease,[10–15] including Courtiol and colleagues' study[16] in which they used a deep learning-based approach—known as MesoNet—to examine whole-slide digitised images to predict overall survival. MesoNet used regions in the stroma and features related to cellular diversity and inflammation to create its model. Although the findings are striking, overall, the results from these histology-focused studies are inconsistent. Several studies have tried to identify predictive molecular features derived from genetic sequencing, but these molecular features are based on alterations in a single gene or gene family.[17–20] For example, somatic *BAP1* alterations, first identified in 2011,[21] have been associated with prolonged survival,[18] whereas *CDKN2A* deletion is associated with poor survival.[22] In modern analyses, many of these factors have been reconfirmed[6,23,24] and other factors have been identified.[25–31] Unfortunately, the existing indices are unable to integrate clinical characteristics, pathological features, and molecular profiling, and are also unable to generate patient-level prognostication.

We previously tested a machine-learning tool in patients with lung adenocarcinomas; this tool provided superior risk stratification relative to the available prognostic characteristics, including stage and performance status.[32] Refining the ability to prognosticate on the basis of biological features has tremendous potential for patients, and can provide new avenues for research into drivers of poor outcomes.[32] As such, we aimed to develop the machine-learning tool, known as OncoCast-MPM, to create a novel prediction model for patient prognosis that incorporates clinicopathological features and comprehensive molecular profiling data without relying on subjective variables such as performance status and pain, or laboratory values that can fluctuate over short periods of time such as white blood cell count and lactate dehydrogenase.[33,34]

Methods

Participants and procedures

We did a retrospective study looking at malignant pleural mesothelioma tumours using next-generation sequencing from the Memorial Sloan Kettering Cancer Center-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT), which is an institutionally created and operated platform that uses matched normal controls to identify only somatic alterations.[35] The Memorial Sloan Kettering Cancer Center is a tertiary cancer care facility in Manhattan, with most patients coming from the New York City metropolitan area, which includes approximately 20 million people. Previously described oncogenic or possibly oncogenic variants as reported by the OncoKB precision oncology knowledge base were included in this analysis.[36] MSK-IMPACT expanded during the study; the appendix (pp 1–2) lists the genes and the number of samples run in each version. To mitigate the potential influence of genes assessed in only some patients, we included the 341 genes common to all next-generation sequencing panels in our model. We obtained written informed consent from all participating patients. The Memorial Sloan Kettering Cancer Center Institutional Review Board approved a waiver for this retrospective analysis, and the study was done in accordance with the US Common Rule.[37]

Consecutive patients treated at the Memorial Sloan Kettering Cancer Center between Feb 15, 2014, and Jan 28, 2019, with malignant pleural mesothelioma and the MSK-IMPACT data were included. If multiple samples were studied for a given individual patient, the results from the sample obtained closest to the time of diagnosis were used. Medical records were reviewed, and relevant clinical information was extracted: age, sex, date of diagnosis, stage at diagnosis (AJCC, eighth edition), histology, smoking history, self-reported classic occupational asbestos exposure, date of advanced disease (defined as recurrence after a surgical treatment or date of diagnosis with stages IIIB–IV), survival status, and date of death or last follow-up. Overall survival was defined as the time from the date of diagnosis of advanced disease (either stages IIIB–IV or recurrent cancer) until the date of death or last follow-up. Patients were staged in accordance with the American Society of Clinical Oncology Clinical Practice Guidelines for Malignant Pleural Mesothelioma[38] and the International Association for the Study of Lung Cancer staging update[39–42] used for the AJCC (eighth edition) of staging.[43] Histological classification was assigned according to

WHO classification.[44] Only cases with complete data were included in this analysis, which comprised the so-called MSK-IMPACT cohort.

Model development

OncoCast-MPM is an application of the OncoCast algorithm previously used for survival stratification of patients with lung adenocarcinoma.[32] OncoCast is an ensemble learning approach for survival stratification and feature selection based on elastic-net penalised Cox proportional hazard models. This regularised regression method linearly combines the L1 and L2 penalties of the lasso and ridge methods for variable selection as implemented in the R glmnet package.[45] The OncoCast ensemble learning procedure further estimates the variable importance of covariates by compiling the frequency of selection of variables across all models generated. This machine-learning tool repeatedly and randomly splits the entire cohort into training (two-thirds of the cohort) and test (a third of the cohort) sets 200 times to generate an ensemble of classifiers with a varying selection of genes, gene combinations, and other clinicopathological features (age, histology [epithelioid, biphasic, or sarcomatoid], sex, smoking status, stage, and self-reported classic occupational asbestos exposure). The training cohort is used to build the elastic-net model, whereas the testing cohort is used to assess the performance of the model generated.

See **Online** for appendix

The algorithm aggregates prognostic effects across the panel of sequenced cancer genes and other clinicopathological features to derive a risk score for each patient (scaled from 0 to 10). A total of 274 variables were included in the training set. To evaluate prognostic performance, we calculated the concordance probability, which measures the concordance between the risk score and survival. The OncoCast-MPM method and R package used are available online. OncoCast-MPM was developed as a web application and is freely accessible.

For the **OncoCast-MPM method and R package** see https://github.com/AxelitoMartin/OncoCast

For the **OncoCast-MPM web application** see https://tinyurl.com/yd3ujxwv

Risk stratification

The OncoCast-MPM model was used to stratify the MSK-IMPACT cohort of patients into low-risk and high-risk categories and was then validated using publicly available malignant pleural mesothelioma data from The Cancer Genome Atlas (TCGA),[46] a joint effort between the US National Cancer Institute and the National Human Genome Research Institute, which molecularly characterised 10 000 primary cancer and matched normal samples from 33 types of cancer to form the TCGA cohort. Samples for the mesothelioma TCGA cohort were collected from participating institutions from untreated patients. Stringent specimen adequacy criteria were applied, and the histological diagnosis was centrally confirmed.[46]

Details regarding the acquisition and assessment of clinicopathological criteria and whole exome sequencing have been previously described for the TCGA cohort.[46] The risk group dichotomisation threshold was selected by first applying hierarchical clustering to the predicted risk score using the Euclidean distance and Ward's agglomerative method. The resulting dendrogram was then dichotomised, forming two risk groups. The threshold risk score for classifying patients into high-risk versus low-risk groups was 4·18. All model parameters, including the threshold, were trained and locked in the MSK-IMPACT cohort and then applied to the TCGA cohort as validation.

Statistical analysis

The clinical and pathological features as well as the gene alteration frequencies of the MSK-IMPACT and TCGA cohorts were compared using the χ^2 test in R. Left-truncation was used to adjust for the potential survival bias introduced by patients whose tumours were sequenced substantially after their initial diagnosis;[32] these patients were considered so-called immortal from their initial time of diagnosis to the time of referral for MSK-IMPACT sequencing, introducing survival bias if not adjusted.[32] We used the method discussed in Kalbfleisch and Prentice,[47] which incorporates left-truncated and right-censored data to construct the Surv function in the R penalised package for lasso-penalised Cox regression. Kaplan-Meier survival curves with log-rank testing was used to analyse overall survival. Concordance probability estimate was used to determine the discriminative capability of the OncoCast model.[48] This estimate has the same interpretation as the C index; however, the estimate is more robust in the presence of censoring.

We did all statistical analyses using R (version 3.6.3).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

From Jan 1, 2014, MSK-IMPACT testing was offered to all patients with malignant pleural mesothelioma who were receiving ongoing treatment at the Memorial Sloan Kettering Cancer Center. Not all approached patients consented and not all patients who consented had sufficient archival tissue available. Ultimately, between Feb 15, 2014, and Jan 28, 2019, we collected MSK-IMPACT data from the tumour tissue of 194 patients in the MSK-IMPACT cohort. The median age was 70 years (IQR 63–74). In the MSK-IMPACT cohort, 148 (76%) of 194 patients were male and 163 (84%) had epithelioid histology, and tumour mutation burden was generally low (table). 129 (66%) of 194 patients had stages I–IIIA disease, and 136 (70%) of 194 patients' tumours were sampled and sequenced within 6 months of the diagnosis of advanced disease. Common alterations identified in this cohort were *BAP1* (including deletions), *NF2*, *TP53*, *SETD2*, and *LATS2*. 116 (60%) of 194 patients had tissue available for assessment of *BAP1* expression by immunohistochemistry. 17 (24%) of 72 patients with *BAP1* loss by immunohistochemistry did not have an alteration identified on the sequencing platform. However, in contrast to the TCGA cohort,

no relationship was identified between *BAP1* and *SETD2* alterations (data not shown). Commercially available next-generation sequencing data with routine patient characteristics and pathological features were entered into OncoCast-MPM, which generated predictive risk scores (scaled from 0 to 10) for individual patients (appendix p 4). Higher scores indicated a worse prognosis (ie, shorter expected survival).

A wide spread of risk scores was observed within the MSK-IMPACT cohort, with an almost bimodal-appearing distribution (appendix p 5). Hierarchical clustering with a cutoff at the 73rd percentile (risk score of 4·18) was used to stratify patients into two risk groups. Of the 194 patients in the MSK-IMPACT cohort, 52 (27%) were stratified into the high-risk group and 142 (73%) into the low-risk group. OncoCast-MPM prognostic risk scores were also calculated for the 74 patients with malignant pleural mesothelioma profiled in the TCGA cohort (appendix p 5). Although the distribution of scores in the TCGA cohort appears different from the MSK-IMPACT cohort, the cutoff risk score of 4·18 also successfully separated the TCGA cohort into low-risk and high-risk groups.

274 variables were inputted and available for inclusion in the OncoCast-MPM model (appendix pp 2–3). The median number of prognostic features selected was 91 (IQR 27–194). For all of the 274 variables included in OncoCast-MPM, we used univariate Cox's proportional hazard models in which the effect size and p value of each feature were recorded. The weights for individual features and their significance are shown in the volcano plot of the univariate Cox proportional hazard coefficients (appendix p 6). Highly influential unfavourable features in this univariate analysis included *CDKN2A* and *CDKN2B* deletions, as well as mutations of *TP53*, *TERT*, *GNAS*, and *DICER1*. Highly influential favourable features included epithelioid histology and *BAP1* or *PBRM1* mutations.

To examine the importance of features within the integrated OncoCast-MPM model, we did a multivariate analysis with repeated sampling across all cross-validation in the training model (appendix p 6). Selection frequency served as a surrogate for the importance of each variable. Highly favourable features included *BAP1* mutations, *PBRM1* mutations, epithelioid histology, a history of smoking or tobacco use, and reported classic occupational asbestos exposure, whereas highly unfavourable features included male sex, deletion of *CDKN2A* and *CDKN2B*, mutations of *TP53* and *TERT*, age, advanced-stage disease, and biphasic histology.

OncoCast-MPM integrated these features to provide individualised risk scores for each patient within the cohort. Patients with risk scores below 4·18 were considered low risk. The median overall survival was 30·8 months (95% CI 22·7–36·2) for the low-risk group compared with 13·9 months (10·7–18·0) for the high-risk group (hazard ratio [HR] 3·0 [95% CI 2·0–4·5]; p<0·0001; figure 1A). At each of 200 iterations, the OncoCast-MPM model was internally cross-validated by using the train model to calculate the concordance probability estimate on the omitted samples, resulting in a cross-validated estimate of 0·67 (IQR 0·64–0·69). To explain the survival among the high-risk group, we examined patterns of treatment with known active agents (pemetrexed-based cytotoxic therapy and checkpoint inhibitors), which were similar between the low-risk and high-risk groups (data not shown). Furthermore, the high-risk group was composed as follows: 27 (52%) of 52 had early-stage

disease, 17 (33%) had surgery, and 29 (56%) had epithelioid, 17 (33%) biphasic, or six (12%) sarcomatoid histology.

The publicly available TCGA cohort was used for validation and was similar to the MSK-IMPACT cohort (table). The TCGA cohort was a comprehensive integrated genomic study of 74 patients with malignant pleural mesotheliomas.[46] Similar proportions of disease stage and histology were observed among the two cohorts. Although all of the TCGA samples were surgical specimens, those in the MSK-IMPACT cohort consisted of a mix of 65 (34%) of 194 resection samples and 129 (66%) biopsy samples. A significant difference was observed in the median age of the cohorts, with the TCGA cohort's median age being lower than that of the MSK-IMPACT cohort (64 years [IQR 56–69] *vs* 70 years [63–74]; p<0·0001). Frequencies of some genetic alterations also differed between the TCGA and MSK-IMPACT cohorts, with *BAP1* alterations being more common in the MSK-IMPACT cohort than in the TCGA cohort (93 [48%] of 194 *vs* 17 [23%] of 74; appendix p 7).

The OncoCast-MPM risk model was applied to malignant pleural mesothelioma data from the TCGA cohort. This risk model identified 33 (45%) of 74 high-risk patients and 41 (55%) low-risk patients. The median overall survival was 23·6 months (95% CI 15·1–28·4) for the low-risk group and 13·6 months (9·8–17·9) for the high-risk group (HR 2·3 [95% CI 1·3–3·8]; p=0·0019; figure 1B).

We sought to compare our model with the current risk stratification method. In the clinical setting, patients with malignant pleural mesothelioma are stratified by disease stage (early [IA–IIIA] or advanced [IIIB–IV]). In the MSK-IMPACT cohort, the median overall survival was 27·0 months (95% CI 21·2–32·5) for the early-stage disease group versus 18·3 months (13·5–25·3) for the advanced-stage disease group. No significant difference was observed in the Kaplan-Meier survival curve when stratifying by stage (HR 1·4 [95% CI 1·0–2·1]; p=0·064; figure 1C).

Landmark survival analyses were significantly divergent at 1 year for the OncoCast-MPM MSK-IMPACT cohort (90% for the low-risk group *vs* 59% for the high-risk group; p=0·0030); and somewhat divergent—although not significant—for the OncoCast-MPM TCGA cohort (77% for the low-risk group *vs* 57% for the high-risk group; p=0·099) and stratifying by stage in the MSK-IMPACT cohort (86% for the early-stage disease *vs* 71% for the advanced-stage disease; p=0·20; figure 1D). However, although the divergence was maintained at 3 years for OncoCast-MPM-derived risk groupings in both the MSK-IMPACT cohort (40% for the low-risk group *vs* 7% for the high-risk group; p=0·0052) and TCGA cohort (28% for the low-risk group *vs* 4% for the high-risk group; p=0·062), disease stage did not define groups with different overall survival at 3 years in the MSK-IMPACT cohort (31% for early-stage disease *vs* 30% for advanced-stage disease; p=90).

The prevalence of commonly mutated genes in the MSK-IMPACT cohort was also examined as a function of stratification by the OncoCast-MPM risk group and disease stage (figure 2). The OncoCast-MPM identified high-risk and low-risk groups by considering as many as 239 variables per patient. Several genetic alterations segregated by risk in the OncoCast-MPM model were observed. *CDKN2B* deletion, *CDKN2A* deletion, and alterations in *TERT*, *NF2*,

TP53, and *LATS2* were associated with high risk, whereas *SETD2* and *BAP1* alterations were associated with low risk. Applying the same analysis to stage stratification, we only found one gene with a significant association: *TP53* was associated with advanced-stage disease. No genetic alterations were associated with early-stage disease.

To emphasise the use of combining clinical, genomic, and histological features into a single analysis, OncoPrint plots for patients in low-risk and high-risk groups from the MSK-IMPACT cohort were examined (figure 3A). *CDKN2A* (40 [77%] of 52) and *CDKN2B* (40 [77%]) deletions, as well as *NF2* mutations (25 [48%]), were enriched in the high-risk group. However, *CDKN2A* (21 [15%] of 142) and *CDKN2B* (14 [10%]) deletions, as well as *NF2* mutations (24 [17%]), were still relatively common in the low-risk group. Similarly, although *BAP1* mutations were enriched in the low-risk group (60 [42%] of 142), many patients in the high-risk group also had these mutations (11 [21%] of 52). In addition, although *BAP1* mutations were a strong driver of the low-risk group, 82 (58%) of 142 patients in this group did not have tumours with *BAP1* mutations. With respect to histology, although epithelioid disease predominated in the low-risk group (133 [94%] of 142), 29 (56%) of 52 patients in the high-risk group also had epithelioid histology. Furthermore, among the patients with a sarcomatoid histology in the MSK-IMPACT cohort, three (2%) of 142 were in the low-risk group, whereas six (12%) of 52 were in the high-risk group. To ensure that differences were not related to disparities in therapy, we examined treatment patterns by risk group. Among the low-risk group, 48 (34%) of 142 patients received checkpoint inhibitor treatment (four received dual checkpoint inhibition and 49 received single-agent therapy), and 70 (49%) participated in a therapeutic clinical trial. Treatment patterns were similar with respect to chemotherapy, with 46 (88%) of 52 patients in the high-risk group and 118 (83%) of 142 patients in the low-risk group receiving platinum-based chemotherapy (p=0·50). For the high-risk group, 24 (46%) of 52 patients received immunotherapy (two received dual checkpoint inhibition and 22 received single-agent therapy); although numerically distinct, this difference was not significant (p=0·13). Likewise, although fewer low-risk patients (72 [51%] of 142) than high-risk patients (19 [37%] of 52) participated in a therapeutic clinical trial, the difference was not significant (p=0·10).

OncoPrint plots for the TCGA cohort showed a similar pattern to those of the MSK-IMPACT cohort (figure 3B). As with the MSK-IMPACT cohort, certain features were enriched in a specific risk group but still substantially present in the other risk group. In the TCGA cohort, *CDKN2A* (23 [70%] of 33) and *CDKN2B* (22 [67%]) deletions were enriched in the high-risk group as in the MSK-IMPACT cohort. However, *CDKN2A* (ten [24%] of 41) and *CDKN2B* (nine [22%]) deletions were still abundant in the low-risk group. Similar to the MSK-IMPACT cohort, *BAP1* mutations in the TCGA cohort were more common in the low-risk group (14 [34%] of 41) but were still observed in the high-risk group (three [9%] of 33).

The percentage of patients who would be reassigned from low-risk prognostic categorisation based on stage to high risk based on OncoCast-MPM and those who would be reassigned from high-risk prognostic categorisation based on stage to low risk based on OncoCast-MPM was examined (figure 4A). Among those with early-stage disease, 27 (21%) of 129

patients were categorised as high risk by OncoCast-MPM, whereas 40 (62%) of 65 patients with advanced-stage disease were categorised as low risk by OncoCast-MPM. Overall, 67 (35%) of 194 patients were potentially mischaracterised by disease stage. Exploring this potential misclassification by stage further, we examined whether a significant survival difference existed based on the OncoCast-MPM risk group within a stage category. Among patients with early-stage disease, the median overall survival of the low-risk group was 33·2 months (95% CI 27·0–36·2) and the high-risk group was 15·0 months (10·3–18·3), with a significant survival difference observed between both risk groups (HR 3·9 [95% CI 2·3–6·9; p<0·0001; figure 4B). Similarly, among those with advanced-stage disease, the median overall survival of the low-risk group was 21·1 months (95% CI 16·0–41·1) and the high-risk group was 13·5 months (9·2–23·4), with a significant survival difference also observed between both risk groups (HR 2·1 [95% CI 1·1–4·0; p=0·022; figure 4C).

Discussion

Stage-based stratification in malignant pleural mesothelioma has substantial limitations for research and practising clinicians.[49] OncoCast-MPM represents the first machine-learning-generated prognostic tool that combines clinical characteristics, pathological features, and molecular profiling from standard next-generation sequencing testing for malignant pleural mesothelioma. Not only was our use of OncoCast-MPM successful in generating a prediction model that stratified our patient cohort into two risk groups with a more than two times difference in survival, it was considerably more accurate than the current stage and histology stratification model, particularly at later timepoints. OncoCast-MPM was also able to stratify patients with advanced-stage or early-stage malignant pleural mesothelioma into risk groups with significant survival differences. Although the concordance probability estimate did not reach 0·80, an estimate of 0·67 was reached, and compares favourably with the eighth edition of the malignant pleural mesothelioma AJCC staging system, which has a C index of 0·54.[9] Unlike the EORTC[4] and CALGB[3] scoring systems, our model does not require physician assessment of performance status, patient-reported symptoms, or a single datapoint for laboratory values that can fluctuate. When compared with the EORTC scoring system[4] using the same cutoff values in both models, OncoCast-MPM showed a higher discriminatory effect of different risk groups. Importantly, our cohort represents the typical population with malignant pleural mesothelioma, and our model was validated in an independent dataset (ie, the TCGA cohort) despite differences in clinical characteristics, pathological features, and frequencies of key genomic alterations. This study design suggests that OncoCast-MPM is robust and generalisable. Prospective independent validation is the next important step to confirm the validity and reproducibility of our model.

OncoCast-MPM powerfully separated patients for survival, even as early as 1 year, and was not dependent on any single feature or gene alteration. For example, although *CDKN2A* loss identifies patients with poor survival, 17 (12%) of 142 patients in the low-risk group also had *CDKN2A* loss and would be inappropriately risk stratified for a poor outcome if only *CDKN2A* status was used. Furthermore, although *BAP1* alteration identifies patients with improved survival, ten (19%) of 52 patients in the high-risk group also had *BAP1* alteration and would be inappropriately risk stratified for a good outcome if only *BAP1* status was used. Similarly, although nearly all of the low-risk group had epithelioid histology, more

than half of the high-risk group also had epithelioid histology. OncoCast-MPM generated risk scores for individual patients with more powerful separation of high risk and low risk than individual gene or feature analyses; our use of a continuous risk score provides tremendous granularity for understanding heterogeneity in clinical outcomes.

To the best of our knowledge, this is the first integrated risk-prediction model in malignant pleural mesothelioma that requires no special testing. Any commercially available next-generation sequencing and routinely available clinical characteristics and histological classification (epithelioid, biphasic, or sarcomatoid) are sufficient to generate risk scores. Thus, OncoCast-MPM can be applied to any patient anywhere in the world for free.

For the **OncoCast-MPM web application** see https://tinyurl.com/yd3ujxwv

The paucity of high-quality prognostic research validated in an independent dataset leaves clinicians treating patients with malignant pleural mesothelioma few insights for stratification and therapeutic adaptability. Management for several malignancies have progressed because of the discovery of targetable oncogenes; malignant pleural mesothelioma, however, lags behind in gaining benefits from this transformation in tumour-sequencing capability. Indeed, malignant pleural mesothelioma is characterised by an abundance of tumour suppressor alterations for which a putative intervention remains unknown.[50] In the absence of highly effective therapies with valid biomarkers in a very heterogeneous disease, efforts to provide accurate and personalised prognoses become essential to provide individualised patient care, prioritise high-risk patients for clinical trials, evaluate real-world data, and generate accurately matched historical control groups. OncoCast-MPM might fill this unmet need.

Several important potential limitations to OncoCast-MPM exist. First, our model has not been prospectively validated, which would be necessary to ensure that this retrospective analysis is not confounded by variables beyond the scope of the model. Second, as molecular testing expands to cover more genes and panels, variations between testing platforms might increase, and the 341 genes common to all of our cases to derive the model might not be universally available for all patients. However, the most commonly altered genes in malignant pleural mesothelioma are included in the 341 genes used in the MSK-IMPACT panel and on the US FDA-approved commercially available tumour testing panels.[51] Additional genes are likely to have low frequencies of alterations and would therefore possibly have little effect on survival. To account for the ongoing evolution of molecular testing, this model will require periodic updates as gene panels expand. Additionally, although increasingly available via commercial platforms and reimbursable in the USA given several tumour agnostic drug approvals, next-generation sequencing might not be readily available everywhere. It is also noteworthy that OncoCast-MPM is treatment agnostic, so it does not consider treatment regimens. Finally, the cases used for creation and validation of our model might not be reflective of most patients with mesothelioma seen around the world. Wide prospective validation will be an essential next step.

The OncoCast-MPM risk score provides a new tool to accurately estimate the prognosis of patients with malignant pleural mesothelioma. As suggested by the recent National

Cancer Institute, International Association for the Study of Lung Cancer, and Mesothelioma Applied Research Foundation joint position paper,[49] OncoCast-MPM includes multiple clinical and translational correlatives that are readily obtained on all patients with malignant pleural mesothelioma. Our assessment can be applied to clinical trials and real-world datasets. By incorporating genomic data with conventional clinical characteristics and pathological features, we have shown an ability to significantly improve the description of patient populations. These more precise individualised prognoses will facilitate better use of historical control groups and real-world datasets, as well as help to risk stratify patients enrolling onto clinical trials. This refinement of the ability to prognosticate based on biological features also yields tremendous value to patient care and will help to fuel a new cycle of refined research that can be aimed towards drivers of poor outcomes as well as prioritising patients with poor prognosis for clinical trial enrolment. Unlike previously proposed prognostic scoring systems, the discriminatory ability of our model is validated, powerful, and present at both 1-year and 3-year landmarks; does not require data from special testing; and is freely publicly available for further independent and prospective validation. Ultimately, we hope OncoCast-MPM will facilitate the development of better, targeted therapeutics in malignant pleural mesothelioma and improve outcomes similar to the gains observed in other oncogenic cancers.

For the **source code for the OncoCast analysis** see https://github.com/AxelitoMartin/OncoCast

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded and supported in part by the National Institutes of Health/National Cancer Institute (grant P30 CA008748).

Declaration of interests

MGZ has received consulting fees from Takeda, GlaxoSmithKline, Epizyme, Aldeyra Therapeutics, Novocure, and Atara; honoraria from Research to Practice, Medical Learning Institute, and OncLive; and grants from the National Institutes of Health/National Cancer Institute. MGZ serves as chair of the Board of Directors of the Mesothelioma Applied Research Foundation and is an employee of Memorial Sloan Kettering. Memorial Sloan Kettering receives research funding from the US Department of Defense, the National Institutes of Health, GlaxoSmithKline, Epizyme, Polaris, Sellas Life Sciences, Bristol Myers Squibb, Millenium, Curis, and Roche for research done by MGZ. Memorial Sloan Kettering had an institutional agreement with IBM for Watson for Oncology and receives royalties from IBM. MO reports honoraria from PharmaMar, Novartis, and Targeted Oncology outside the submitted work. AR reports grants from Varian Medical Systems, Boehringer Ingelheim, and Pfizer; grants and personal fees from AstraZeneca and Merck; personal fees from Research to Practice, Cybrexa, and More Health; and non-financial support from Philips/Elekta, outside the submitted work. PSA's laboratory work is supported by grants from the National Institutes of Health (P30 CA008748, R01 CA236615-01, R01 CA235667, R21 CA213139, and T32 CA009501), the US Department of Defense (CA170630 and CA180889), the Memorial Sloan Kettering Technology Development Fund, the Miner Fund for Mesothelioma Research, Mr William H Goodwin and Alice Goodwin, the Commonwealth Foundation for Cancer Research, and the Experimental Therapeutics Center of Memorial Sloan Kettering Cancer Center. PSA reports research funding from ATARA Biotherapeutics; scientific advisory board or consultant roles with ATARA Biotherapeutics, Carisma Therapeutics, Imugene, and Takeda Therapeutics; patents, royalties, and intellectual property on mesothelin-targeted CAR; and pending patent applications on other T-cell therapies. VWR receives grant support for institutional clinical trials from Genelux and Genentech, and receives travel reimbursement for robotic mentoring from Intuitive Surgical. VWR is also co-chair for the Thoracic Malignancy Steering Committee and National Cancer Institute; and a member of the Data Safety

Monitoring Committee, MARS2 Trial (UK). MGK receives personal fees from AstraZeneca, Pfizer, Regeneron, and Daiichi-Sankyo; received honoraria for participation in educational programmes from WebMD, OncLive, Physicians Education Resources, Prime Oncology, Intellisphere, Creative Educational Concepts, Peerview, i3 Health, Paradigm Medical Communications, AXIS, Carvive Systems, AstraZeneca, and Research to Practice; and received travel support from AstraZeneca, Pfizer, Regeneron, and Genentech. MGK is an employee of Memorial Sloan Kettering. Memorial Sloan Kettering has received research funding from the National Cancer Institute, The Lung Cancer Research Foundation, Genentech Roche, and PUMA Biotechnology for research done by MGK. Memorial Sloan Kettering has licensed testing for EGFR T790M to MolecularMD. JLS reports stock ownership in the following companies: Allergan, Chemed, Johnson & Johnson, Merck, Pfizer, and Thermo Fischer Scientific. ML reports honoraria for ad-hoc advisory board participation from Merck, AstraZeneca, Bristol Myers Squibb, Takeda, Bayer, and Lilly Oncology; and research support from LOXO Oncology, Merus, and Helsinn Therapeutics. All other authors declare no competing interests.

References

1. Noone AM, Howlader N, Krapcho M, et al. SEER cancer statistics review, 1975–2015. Bethesda: National Cancer Institute, 2018.
2. Tsao AS, Wistuba I, Roth JA, Kindler HL. Malignant pleural mesothelioma. J Clin Oncol 2009; 27: 2081–90. [PubMed: 19255316]
3. Herndon JE, Green MR, Chahinian AP, Corson JM, Suzuki Y, Vogelzang NJ. Factors predictive of survival among 337 patients with mesothelioma treated between 1984 and 1994 by the Cancer and Leukemia Group B. Chest 1998; 113: 723–31. [PubMed: 9515850]
4. Curran D, Sahmoud T, Therasse P, van Meerbeeck J, Postmus PE, Giaccone G. Prognostic factors in patients with pleural mesothelioma: the European Organization for Research and Treatment of Cancer experience. J Clin Oncol 1998; 16: 145–52. [PubMed: 9440736]
5. Davidson B Prognostic factors in malignant pleural mesothelioma. Hum Pathol 2015; 46: 789–804. [PubMed: 25824607]
6. Pass HI, Giroux D, Kennedy C, et al. Supplementary prognostic variables for pleural mesothelioma: a report from the IASLC staging committee. J Thorac Oncol 2014; 9: 856–64. [PubMed: 24807157]
7. Kadota K, Kachala SS, Nitadori J, et al. High SUVmax on FDG-PET indicates pleomorphic subtype in epithelioid malignant pleural mesothelioma: supportive evidence to reclassify pleomorphic as nonepithelioid histology. J Thorac Oncol 2012; 7: 1192–97. [PubMed: 22617244]
8. Lim JH, Choi JY, Im Y, et al. Prognostic value of SUVmax on ^{18}F-fluorodeoxyglucose PET/CT scan in patients with malignant pleural mesothelioma. PLoS One 2020; 15: e0229299. [PubMed: 32069313]
9. Abdel-Rahman O Challenging a dogma; AJCC 8th staging system is not sufficient to predict outcomes of patients with malignant pleural mesothelioma. Lung Cancer 2017; 113: 128–33. [PubMed: 29110839]
10. Bilecz A, Stockhammer P, Theegarten D, et al. Comparative analysis of prognostic histopathologic parameters in subtypes of epithelioid pleural mesothelioma. Histopathology 2020; 77: 55–66. [PubMed: 32170970]
11. Kadota K, Suzuki K, Colovos C, et al. A nuclear grading system is a strong predictor of survival in epitheloid diffuse malignant pleural mesothelioma. Mod Pathol 2012; 25: 260–71. [PubMed: 21983936]
12. Kadota K, Suzuki K, Sima CS, Rusch VW, Adusumilli PS, Travis WD. Pleomorphic epithelioid diffuse malignant pleural mesothelioma: a clinicopathological review and conceptual proposal to reclassify as biphasic or sarcomatoid mesothelioma. J Thorac Oncol 2011; 6: 896–904. [PubMed: 21358344]
13. Paajanen J, Laaksonen S, Kettunen E, et al. Histopathological features of epithelioid malignant pleural mesotheliomas in patients with extended survival. Hum Pathol 2020; 98: 110–19. [PubMed: 32142836]
14. Rosen LE, Karrison T, Ananthanarayanan V, et al. Nuclear grade and necrosis predict prognosis in malignant epithelioid pleural mesothelioma: a multi-institutional study. Mod Pathol 2018; 31: 598–606. [PubMed: 29327706]

15. Takuwa T, Hashimoto M, Kuroda A, et al. Poor prognostic factors in patients with malignant pleural mesothelioma classified as pathological stage IB according to the eighth edition TNM classification. Ann Surg Oncol 2018; 25: 1572–79. [PubMed: 29616420]
16. Courtiol P, Maussion C, Moarii M, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nat Med 2019; 25: 1519–25. [PubMed: 31591589]
17. van der Bij S, Koffijberg H, Burgers JA, et al. Prognosis and prognostic factors of patients with mesothelioma: a population-based study. Br J Cancer 2012; 107: 161–64. [PubMed: 22644294]
18. Meerang M, Berard K, Friess M, et al. Low Merlin expression and high Survivin labeling index are indicators for poor prognosis in patients with malignant pleural mesothelioma. Mol Oncol 2016; 10: 1255–65. [PubMed: 27378628]
19. Illei PB, Rusch VW, Zakowski MF, Ladanyi M. Homozygous deletion of *CDKN2A* and codeletion of the methylthioadenosine phosphorylase gene in the majority of pleural mesotheliomas. Clin Cancer Res 2003; 9: 2108–13. [PubMed: 12796375]
20. Walsh M Beyond statistical significance. Appl Nurs Res 1988; 1: 101–03. [PubMed: 3239978]
21. Bott M, Brevet M, Taylor BS, et al. The nuclear deubiquitinase *BAP1* is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. Nat Genet 2011; 43: 668–72. [PubMed: 21642991]
22. López-Ríos F, Chuai S, Flores R, et al. Global gene expression profiling of pleural mesotheliomas: overexpression of aurora kinases and *P16*/*CDKN2A* deletion as prognostic factors and critical evaluation of microarray-based prognostic prediction. Cancer Res 2006; 66: 2970–79. [PubMed: 16540645]
23. Bille A, Krug LM, Woo KM, Rusch VW, Zauderer MG. Contemporary analysis of prognostic factors in patients with unresectable malignant pleural mesothelioma. J Thorac Oncol 2016; 11: 249–55. [PubMed: 26845118]
24. Brims FJ, Meniawy TM, Duffus I, et al. A novel clinical prediction model for prognosis in malignant pleural mesothelioma using decision tree analysis. J Thorac Oncol 2016; 11: 573–82. [PubMed: 26776867]
25. Amin W, Linkov F, Landsittel DP, et al. Factors influencing malignant mesothelioma survival: a retrospective review of the National Mesothelioma Virtual Bank cohort. F1000Res 2018; 7: 1184. [PubMed: 30410729]
26. Doi H, Kuribayashi K, Kitajima K, Yamakado K, Kijima T. Development of a novel prognostic risk classification system for malignant pleural mesothelioma. Clin Lung Cancer 2020; 21: 66–74.e2. [PubMed: 31551134]
27. Fuchs TL, Chou A, Sioson L, Sheen A, Gill AJ. Stromal tumour-infiltrating lymphocytes (TILs) assessed using the ITWG system do not predict overall survival in a cohort of 337 cases of mesothelioma. Histopathology 2020; 76: 1095–101. [PubMed: 32215942]
28. Muller S, Victoria Lai W, Adusumilli PS, et al. V-domain Ig-containing suppressor of T-cell activation (VISTA), a potentially targetable immune checkpoint molecule, is highly expressed in epithelioid malignant pleural mesothelioma. Mod Pathol 2020; 33: 303–11. [PubMed: 31537897]
29. Salaroglio IC, Kopecka J, Napoli F, et al. Potential diagnostic and prognostic role of microenvironment in malignant pleural mesothelioma. J Thorac Oncol 2019; 14: 1458–71. [PubMed: 31078776]
30. Van Gerwen M, Alpert N, Wolf A, et al. Prognostic factors of survival in patients with malignant pleural mesothelioma: an analysis of the National Cancer Database. Carcinogenesis 2019; 40: 529–36. [PubMed: 30649229]
31. Mansfield AS, Peikert T, Smadbeck JB, et al. Neoantigenic potential of complex chromosomal rearrangements in mesothelioma. J Thorac Oncol 2019; 14: 276–87. [PubMed: 30316012]
32. Shen R, Martin A, Ni A, et al. Harnessing clinical sequencing data for survival stratification of patients with metastatic lung adenocarcinomas. JCO Precis Oncol 2019; 3: PO.18.00307.
33. Martin A, Shen R. AxelitoMartin/OncoCast: OncoCast internal application update (version v1.1.1). 2020. https://zenodo.org/record/4311821#.YNCaoGhKiUk (accessed July 6, 2021).
34. Martin A, Whiting K, Arora A, Hannum M. AxelitoMartin/gnomeR: initial gnomeR release (version 1.0.0). 2020. https://zenodo.org/record/4171636#.YNCa82hKiUk (accessed July 6, 2021).

35. Cheng DT, Mitchell TN, Zehir A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J Mol Diagn 2015; 17: 251–64. [PubMed: 25801821]
36. Chakravarty D, Gao J, Phillips SM, et al. OncoKB: a precision oncology knowledge base. JCO Precis Oncol 2017; 2017: PO.17.00011.
37. Office for Human Research Protections. US Department of Health & Human Services. 45 CFR 46. 2018. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html (accessed July 6, 2021).
38. Kindler HL, Ismaila N, Armato SG 3rd, et al. Treatment of malignant pleural mesothelioma: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 2018; 36: 1343–73. [PubMed: 29346042]
39. Pass H, Giroux D, Kennedy C, et al. The IASLC Mesothelioma Staging Project: improving staging of a rare disease through international participation. J Thorac Oncol 2016; 11: 2082–88. [PubMed: 27670823]
40. Nowak AK, Chansky K, Rice DC, et al. The IASLC Mesothelioma Staging Project: proposals for revisions of the T descriptors in the forthcoming eighth edition of the TNM classification for pleural mesothelioma. J Thorac Oncol 2016; 11: 2089–99. [PubMed: 27687963]
41. Rice D, Chansky K, Nowak A, et al. The IASLC Mesothelioma Staging Project: proposals for revisions of the N descriptors in the forthcoming eighth edition of the TNM classification for pleural mesothelioma. J Thorac Oncol 2016; 11: 2100–11. [PubMed: 27687964]
42. Rusch VW, Chansky K, Kindler HL, et al. The IASLC Mesothelioma Staging Project: proposals for the M descriptors and for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for mesothelioma. J Thorac Oncol 2016; 11: 2112–19. [PubMed: 27687962]
43. Amin MB, Edge S, Greene F, et al., eds. AJCC cancer staging manual, 8th edn. New York, NY: Springer International Publishing, 2017.
44. International Agency for Research on Cancer. Diffuse malignant mesothelioma. In: Travis WD, Brambilla E, Burke A, et al., eds. World Health Organization Classification of tumours of the lung, pleura, thymus and heart, 4th edn. Lyon: International Agency for Research on Cancer, 2015: 156–69.
45. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33: 1–22. [PubMed: 20808728]
46. Hmeljak J, Sanchez-Vega F, Hoadley KA, et al. Integrative molecular characterization of malignant pleural mesothelioma. Cancer Discov 2018; 8: 1548–65. [PubMed: 30322867]
47. Kalbfleisch JD, Prentice RL The statistical analysis of failure time data. Hoboken, NJ: John Wiley & Sons, 2002.
48. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika 2005; 92: 965–70.
49. Tsao AS, Lindwasser OW, Adjei AA, et al. Current and future management of malignant mesothelioma: a consensus report from the National Cancer Institute Thoracic Malignancy Steering Committee, International Association for the Study of Lung Cancer, and Mesothelioma Applied Research Foundation. J Thorac Oncol 2018; 13: 1655–67. [PubMed: 30266660]
50. Yap TA, Aerts JG, Popat S, Fennell DA. Novel insights into mesothelioma biology and implications for therapy. Nat Rev Cancer 2017; 17: 475–88. [PubMed: 28740119]
51. US Food & Drug Administration. Nucleic acid based tests. 2021. https://www.fda.gov/medical-devices/in-vitro-diagnostics/nucleic-acid-based-tests (accessed April 23, 2021).

Research in context

Evidence before this study

We searched PubMed for research articles published between Jan 1, 1980, and April 15, 2021, using the following search terms: “malignant pleural mesothelioma”, “risk prediction”, and “model”. Many reports exist for prognostically distinct malignant pleural mesothelioma molecular subgroups. Because these studies use features that are not standardly available, none of these potential prognostic tools are widely used. Stage and histology remain the core elements used for prognostication, and there remains much room for improvement. An effort exists to challenge the current guidelines around staging by showing the poor prognostic value of the current eighth edition of the American Joint Committee on Cancer (AJCC) staging system. To help address the absence of reliable prognostic tools for this disease, the National Cancer Institute Thoracic Malignancy Steering Committee, International Association for the Study of Lung Cancer, and Mesothelioma Applied Research Foundation published a consensus paper in 2018 that recognises the substantial limitations of clinical staging and asserts that investigation of prognostic markers should be incorporated into any and all screening and therapeutic protocols.

Added value of this study

To the best of our knowledge, this study is the first externally validated prediction model for malignant pleural mesothelioma incorporating next-generation sequencing data, clinical characteristics, and pathological features. Our application of the novel OncoCast machine-learning platform to mesothelioma has enabled us to create a tool for individual patient prognostication that has a better C index than the eighth edition of the AJCC staging system. These newly described OncoCast-MPM risk groups stratify patient survival as early as 1 year and display durable differences at 3 years, unlike stage-based stratification. Our model found that more than a third of patients are misassigned prognostically based on stage, and that more than half of patients with advanced-stage disease might be low risk. Our model is freely accessible so that it can be further validated and examined prospectively to risk stratify patients into clinically meaningful groups.

Implications of all the available evidence

The results of this study reshape our approach to risk stratification for malignant pleural mesothelioma. The use of our validated model could allow refinement of prognosis for more than a third of patients with malignant pleural mesothelioma who are incorrectly risk stratified based on currently used parameters. Once prospectively validated, our tool could be used to better stratify patients in clinical trials, identify patients with the poorest outcomes to be prioritised for clinical trials, and evaluate real-world datasets.

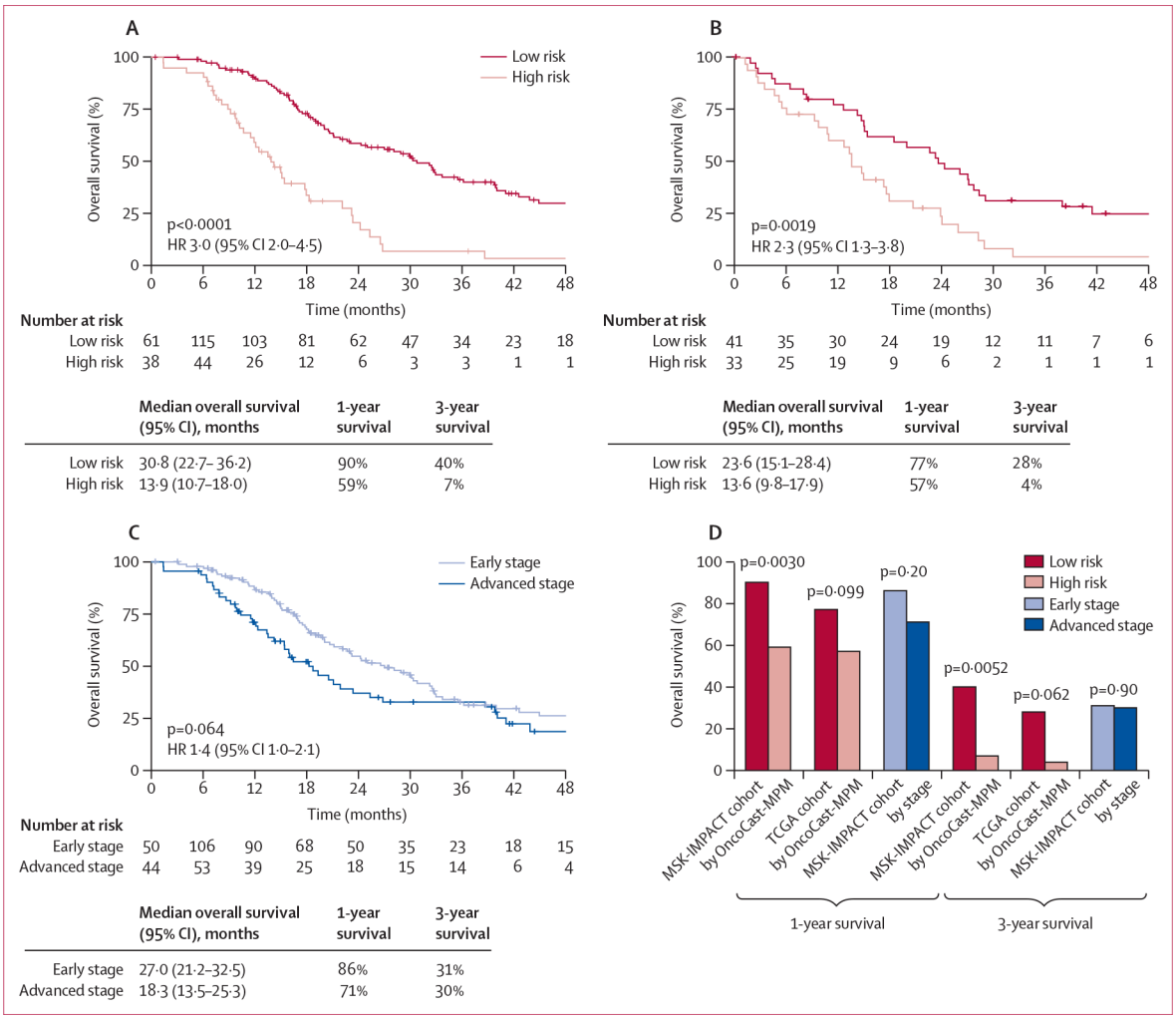

Figure 1: Survival stratification for malignant pleural mesothelioma

(A) Kaplan-Meier plot of overall survival by OncoCast-MPM risk groups in the MSK-IMPACT cohort. (B) Kaplan-Meier plot of overall survival by OncoCast-MPM risk groups in the TCGA cohort. (C) Kaplan-Meier plot of overall survival by stage in the MSK-IMPACT cohort at the time of diagnosis. Patients were grouped by early stage (IA–IIIA) and advanced stage (IIIB–IV). Numbers at risk at time 0 were adjusted; therefore, the number of patients at risk at time 0 differs from the absolute number of patients included in each group. (D) Comparison of landmark survival analyses at 1 year and 3 years for the MSK-IMPACT and TCGA cohorts by OncoCast-MPM as well as the MSK-IMPACT cohort by stage. HR=hazard ratio. MSK-IMPACT=Memorial Sloan Kettering Cancer Center-Integrated Mutation Profiling of Actionable Cancer Targets. TCGA=The Cancer Genome Atlas.

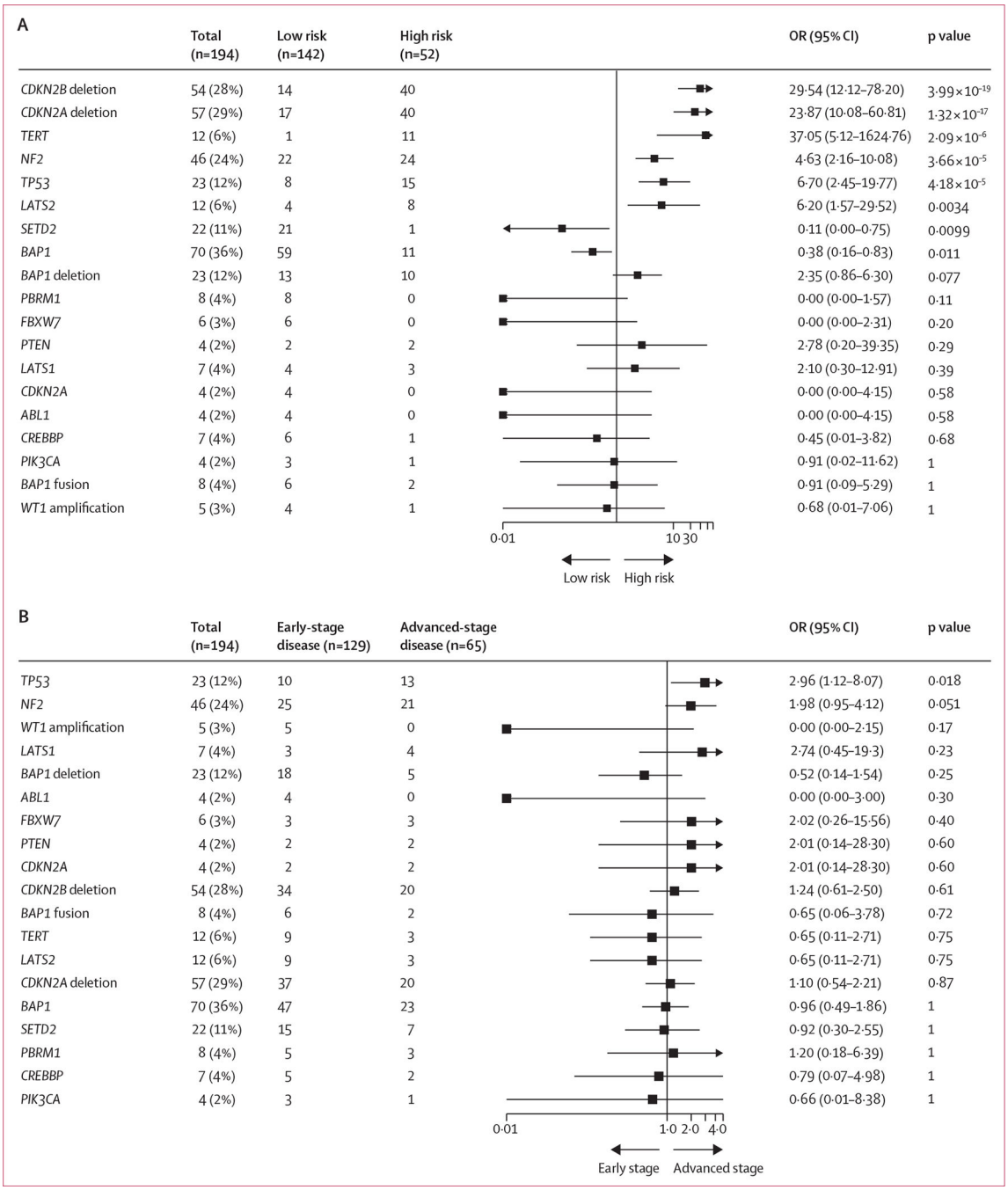

Figure 2: Comparison of gene ORs between stage stratification and OncoCast-MPM risk stratification

(A) Error bars are 95% CIs. ORs by genes between OncoCast-MPM low-risk and high-risk groups. Six alterations were enriched in the high-risk group and two alterations were enriched in the low-risk group. (B) Error bars are 95% CIs. ORs by genes between early-stage malignant pleural mesothelioma and advanced-stage malignant pleural mesothelioma. Across both stages, only one gene was enriched in the advanced-stage group. OR=odds ratio.

Figure 3: OncoPrint plots of risk scores with mutational burden, sex, smoking status, asbestos exposure, histology, and mutation status for the top mutated genes

Percentages are for positive mutated or deleted genes, unless otherwise specified. (A) 142 (73%) of 194 patients with a low-risk score (≤4·18) and 52 (27%) with a high-risk score (>4·18) from the MSK-IMPACT cohort. (B) 41 (55%) of 74 patients with a low-risk score (≤4·18) and 33 (45%) with a high-risk score (>4·18) from the TCGA cohort. MSK-IMPACT=Memorial Sloan Kettering Cancer Center-Integrated Mutation Profiling of Actionable Cancer Targets. TCGA=The Cancer Genome Atlas. *Percentage of males. †Percentage of advanced-stage disease. ‡Percentage of ever smokers. §Percentage of those previously exposed to asbestos. ¶Percentage of epithelioid histology.

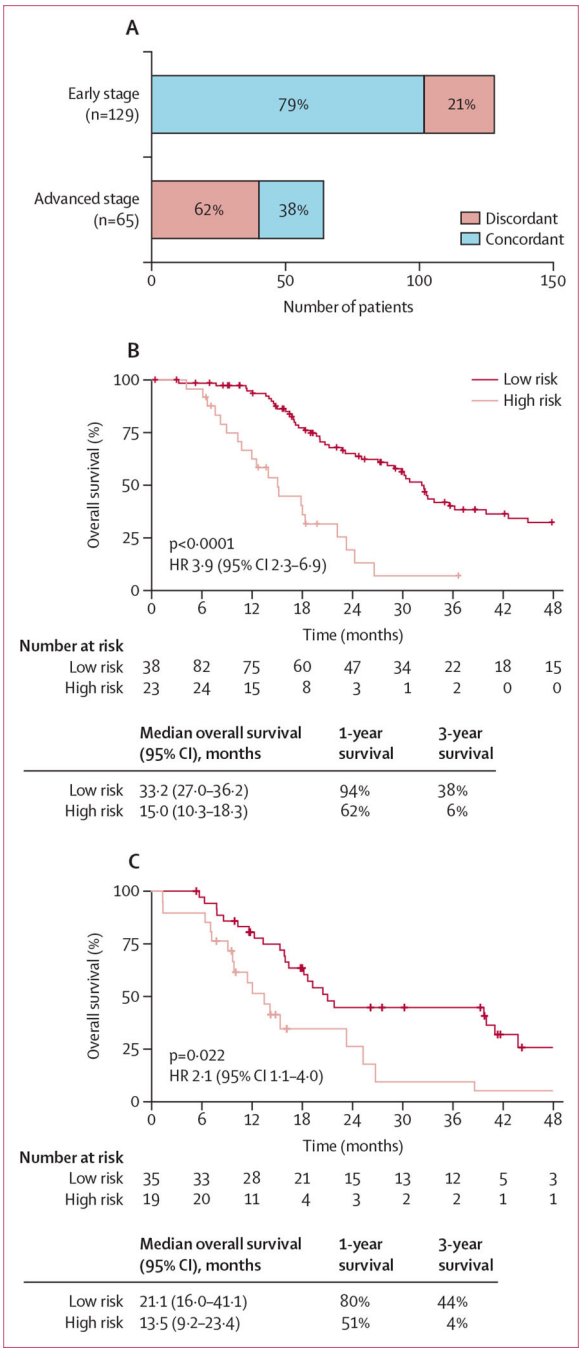

Figure 4: Reassignment of OncoCast-MPM risk categorisation and differentiation of survival among stage classifications

(A) OncoCast-MPM risk stratification by stage. Both concordant and discordant risk categorisation between stage and OncoCast-MPM are shown. (B) Kaplan-Meier plots of overall survival for patients with early-stage disease stratified by OncoCast-MPM risk group. (C) Kaplan-Meier plots of overall survival for patients with advanced-stage disease stratified by OncoCast-MPM risk group. HR=hazard ratio.

Table:

Patient characteristics

	MSK-IMPACT cohort (n=194)	TCGA cohort (n=74)	p value
Age (years)	70 (63–74)	64 (56–69)	<0·0001
Sex	··	··	0·33
Male	148 (76%)	61 (82%)	··
Female	46 (24%)	13 (18%)	··
Ethnicity	··	··	··
Non-Hispanic	174 (90%)	NR	··
Hispanic (not otherwise specified)	6 (3%)	NR	··
South or central America	4 (2%)	NR	··
Other	10 (5%)	NR	··
Smoking status	··	··	··
Current or former	124 (64%)	NR	··
Never	70 (36%)	NR	··
Asbestos exposure *	··	··	0·48
Yes	105 (54%)	46 (62%)	··
No	39 (20%)	13 (18%)	··
Unknown	50 (26%)	15 (20%)	··
Histology	··	··	0·18
Epithelioid	163 (84%)	52 (70%)	··
Biphasic	20 (10%)	13 (18%)	··
Sarcomatoid	9 (5%)	3 (4%)	··
Not otherwise specified	2 (1%)	6 (8%)	··
Stage	··	··	0·20
I–IIIA	129 (66%)	43 (58%)	··
IIIB–IV	65 (34%)	31 (42%)	··
Tumour mutation burden	··	··	··
Mutations (Mb)	2(0–32)	NR	··

Data are median (IQR) or n (%). MSK-IMPACT=Memorial Sloan Kettering Cancer Center-Integrated Mutation Profiling of Actionable Cancer Targets. NR=not reported. TCGA=The Cancer Genome Atlas.

* Self-reported classic occupational exposure.