Medical Research Archives

ESMED
European Society of Medicine

OPEN ACCESS

Published: March 31, 2023

Citation: Becker T., Mailman M., et al., 2023. Comparison of overall survival prognostic power of contemporary prognostic scores in prevailing tumor indications, Medical Research Archives, [online] 11(3). https://doi.org/10.18103/mra.v11i3.3638

DOI: https://doi.org/10.18103/mra.v11i3.3638

ISSN: 2375-1924

RESEARCH ARTICLE

Comparison of overall survival prognostic power of contemporary prognostic scores in prevailing tumor indications

T. Becker[1], Marc Mailman[2], Sandy Tan[2], Ernest Lo[2], A. Bauer-Mehren[1]*

[1]Data & Analytics, Pharma Research and Development, Roche Innovation Center Munich

[2]Roche Information Solutions, Roche Diagnostics, Santa Clara, United States

*anna.bauer-mehren@roche.com

Abstract

Background: Prognosis of overall survival is instrumental for patient management and can improve conduct of clinical trials and real-world data analysis. With the shift towards cancer immunotherapy, modeling of overall host fitness becomes increasingly important. Here, we compare the performance of contemporary prognostic scores constructed from routinely measured biomarkers.

Patients and methods: We used patient data from the Flatiron Health electronic-health record de-identified oncology database and from 16 clinical studies sponsored by Roche. A total of 64,233 patients were analyzed, covering the most prevailing solid tumor and hematology cancer types.

We compared the Royal Marsden Hospital score (solid tumors), international prognostic index (IPI) (blood tumors), the Eastern Cooperative Oncology Group (ECOG) performance status, and the 'Real wOrld PROgnostic score (ROPRO)'. OS was modeled from the start of treatment using Kaplan-Meier analysis and Cox regression.

Results: All investigated scores proved to be prognostic, both in RWD and clinical trial data, and in all indications from the respectively intended range of application. The ROPRO uniformly outperformed other prognostic scores. Concordance indices / hazard ratios in the range of [0.64;0.73]/[2.80;4.50] were found for ROPRO, and in the range of [0.53;0.65]/[1.55; 3.10] for the remaining scores. In hematology trials, the IPI came close to the performance of ROPRO.

Conclusions: Strong and easy-to-apply prognostic scores for overall survival exist. The usage of all investigated scores can be recommended. With moderate extra effort, the implementation of ROPRO can create considerable improvement.

Introduction

Overall survival (OS) is a typical primary outcome measure in oncology trials. With OS-prognostic biomarkers it is possible to improve design, conduct, and analysis of clinical trials and real-world data. In general, prediction of survival is instrumental for optimal patient management[1]. Prognostic or predictive biomarkers can support the choice of treatment or treatment change physicians make[2].

In addition to indication-specific biomarkers, such as HER2-status in breast cancer[3], general prognostic scores for OS aim to measure overall health status and patient prospects. The Royal Marsden Hospital Score[4] (RMHS), is an established prognostic score for solid tumors, which comprises information on the blood biomarkers albumin and LDH, together with information on metastatic sites, into an overall risk assessment. In hematology, the International Prognostic Index[5] is used, a score comprising information about age, tumor stage, LDH, extranodal sites, and the Eastern Cooperative Oncology Group (ECOG) performance status[6], cf. below. Modifications of the IPI for follicular lymphoma, FL- IPI[7], and chronic lymphocytic leukemia (CLL), the CLL-IPI[8] exist.

A key shift in the oncology treatment paradigm is the emergence of cancer immunotherapy, which treats the patient's immune system, not primarily the tumor. Therefore, modeling of overall host fitness becomes increasingly important. Health authorities such as the Food and Drug Administration recently provided updated guidance on patient enrichment strategies in investigational studies aiming to (i) decrease interpatient variability, (ii) identify high-risk patients to enable prognostic enrichment strategies, and (iii) to identify more responsive patients for predictive enrichment[9]. The ECOG performance status, often referred to as just "ECOG", is a prognostic score which is not based on biomarker information, but assessment of physicians, and "... describes a patient's level of functioning in terms of their ability to care for themself, daily activity, and physical ability"[6]. In extension of the idea of an overall assessment of patient fitness status, we recently developed the 'Real wOrld PROgnostic score (ROPRO)'[10] as a pan-cancer prognostic score based on 27 tumor/lab and vital parameters that are part of standard clinical practice.

In view of the increasing relevance of prognostic tools, we compare in this manuscript the prognostic power of the RMHS, the IPI, ECOG, and ROPRO by applying each of these models to previously unseen data. We evaluate their performance in the most common solid and blood tumor indications, using both real world and clinical trial data.

Methods

Our analysis focused on 9 prevailing cancer indications. We investigated advanced non-small cell lung cancer (aNSCLC) and small cell lung cancer (SCLC), metastatic breast cancer (MetBC), metastatic prostate cancer (MetPC), and metastatic colorectal cancer (MetCRC) as solid tumor indications. Chronic lymphocytic leukemia (CLL), diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and

multiple myeloma (MM) were chosen as hematology cancers.

Real word data (RWD) analysis utilized Flatiron Health's (FH) US nationwide longitudinal, demographically and geographically diverse de-identified database derived from electronic health records data from over 280 oncology clinics (800 sites of care), including more than 2.4 million US cancer patients available for analysis[11]. The patient-level data include structured data (e.g. laboratory values, prescribed drugs) in addition to unstructured data collected via technology enabled chart abstraction from physicians' notes and other unstructured documents (e.g. biomarker reports). For our analysis, data from the June 2022 release were extracted, cf. Table 1. Institutional review board approval of the study protocol was obtained before study conduct and included a waiver of informed consent. We analyzed only patients that were not included in the analysis of the original ROPRO discovery manuscript[10].

Clinical study data from 12 solid tumor trials and 4 blood cancer trials (Table 1) sponsored by Roche were extracted from uniformly curated data marts maintained by the company. Only patients with a waiver of informed consent were included.

Table 1. Cohort and trial overview.

Source	Cohort/trial	Cancer type	Indication	#patients	Study name
Flatiron Health	aNSCLC	solid tumor	aNSCLC	16591	-
Flatiron Health	MetBC	solid tumor	MetBC	6706	-
Flatiron Health	MetCRC	solid tumor	MetCRC	8224	-
Flatiron Health	SCLC	solid tumor	SCLC	2133	-
Flatiron Health	MetPC	solid tumor	MetPC	7127	-
Flatiron Health	CLL	blood cancer	CLL	3946	-
Flatiron Health	DLBCL	blood cancer	DLBCL	1982	-
Flatiron Health	FL	blood cancer	FL	1995	-
Flatiron Health	MM	blood cancer	MM	3105	-
Roche trial	BO21005	blood cancer	DLBCL	1403	GOYA
Roche trial	BO21223	blood cancer	NHL	1192	GALLIUM
Roche trial	GO27878	blood cancer	NHL/DLBCL	208	-

Source	Cohort/trial	Cancer type	Indication	#patients	Study name
Roche trial	BO25323	blood cancer	CLL	439	-
Roche trial	BO29554	solid tumor	aNSCLC	455	BFAST
Roche trial	GO28915	solid tumor	aNSCLC	1187	Oak
Roche trial	GO29431	solid tumor	aNSCLC	549	IMpower 110
Roche trial	GO29436	solid tumor	aNSCLC	1187	IMpower 150
Roche trial	GO29437	solid tumor	aNSCLC	1000	IMpower 131
Roche trial	GO29438	solid tumor	aNSCLC	727	IMpower 132
Roche trial	GO29527	solid tumor	aNSCLC	990	IMpower 010
Roche trial	GO29537	solid tumor	aNSCLC	705	IMpower 130
Roche trial	GO30081	solid tumor	SCLC	494	IMpower133
Roche trial	GO30182	solid tumor	MetCRC	349	COTEZO IMblaze370
Roche trial	MO39196	solid tumor	MetBC	649	IMpassion131
Roche trial	WO29522	solid tumor	MetBC	890	IMpassion130

Survival was measured from initiation of study treatment (Roche trials) or from the start of the first documented line of treatment (FH data). In case a cohort is named "advanced/ metastatic" it is the first treatment line under advanced/metastatic tumor. FH data were curated as described previously[10]. From the Roche trials, we always used the ITT population and biomarker data consistent with official study filing reports.

We compared the prognostic power of the following scores: the Royal Marsden Hospital score (RMHS)[4], the IPI[5], ECOG (performance status)[6], and ROPRO[10]. For the Roche CLL trials, we used the indication-specific modifications of the IPI/RORPO, the CLL-IPI[8] and ROPRO-CLL[10]. For score computation, we followed the prescriptions in the respective publications. In detail:

Royal Marsden Hospital score

The RMHS score pre-scale ranges from 0 to 3. The score rawards a point for serum albumin <35 g/L, number of metastatic sites >2, and elevated LDH levels (>ULN). ULN of LDN was here specified as 280 U/L. The pre-scale can be further classified into low risk (pre-scale 0-1) or high risk (pre-scale 2-3). For Kaplan-Meier analysis, we used the high/low

classification for better visibility, while we used the full pre-scale range for computation of C-index and ROC-AUC values.

International prognostic index

The IPI index (pre-scale range 0 to 5) awards a point for each of the following risk factors: ager over 60, tumor stage of III or higher, elevated LDH (>280 U/L), ECOG performance score ≥2, and the presence of at least one extranodal site. The pre-scale can be further classified into high risk (4,5) and remaining patients. For Kaplan-Meier analysis, we used the high/low classification for better visibility, while we used the full pre-scale range for computation of C-index and ROC-AUC values.

Eastern Cooperative Oncology Group performance status

The ECOG performance status is assigned by the physician and ".. describes a patient's level of functioning in terms of their ability to care for themself, daily activity, and physical ability"[6]. It ranges from 0 to 4 in alive patients. For Kaplan-Meier analysis, we used a high/low classification for better visibility, while we used the full ECOG range for computation of C-index and ROC-AUC values. High ECOG was assigned to ECOG values 2,3, and 4. Of note, in clinical trials, mostly patients with ECOG value 0 or 1 are enrolled, ECOG values of 2 or higher are rare.

Real wOrld PROgnostic score (ROPRO)

The Flatiron Health data were curated as described previously and the ROPRO score was computed using the formula and measurement units given by Becker et al.[10]. In essence, the ROPRO is computed as a sum over the 27 included variables. For each variable, the difference to the Flatiron Health mean of the variable is computed and weighted with the estimate from the multivariable Cox regression analysis previously performed on the training data set[10]. The exact ROPRO formula is

(1) 0.00948 (age - 67.15) + 0.14162 (sex - 0.502) + 0.19521 (smoking - 0.576) + 0.02833 (number of metastatic sites - 0.897) - 0.04178 (Hgb - 12.087) + 0.19097 (urea nitrogen - 2.777) - 6x10e04 (platelets - 267.423) + 1.0619 (calcium - 2.231) +0.06098 (glucose - 4.733) - 0.11658 (lymphocytes/leukocytes ratio - 2.953) + 0.19019 (alkaline phosphatase - 4.582) - 0.00896 (protein - 68.888) - 0.05113 (alanine aminotransferase - 3.008) - 0.03988 (albumin - 37.851) + 0.08189 [bilirubin - (-0.773)] - 0.02462 (chloride - 101.434) + 0.10671 [monocytes - (e0.548)] - 0.0543 (eosinophils/leukocytes ratio - 0.307) + 1.22733 (lactate dehydrogenase - 1.694) + 0.42372 (heart rate - 4.402) - 0.39878 [systolic blood pressure - (4.846)] - 0.02083 (oxygen - 96.487) + 0.20066 (ECOG - 0.84) + 0.0905 (neutrophil-to-lymphocyte ratio - 1.156) - 0.17076 (body mass index - 3.301) + 0.13122 (aspartate aminotransferase-to-alanine aminotransferase ratio - 0.092) + 0.081 (tumor stage - 3.098).

Individual patient ROPRO scores are derived by inputting their measurements for each of

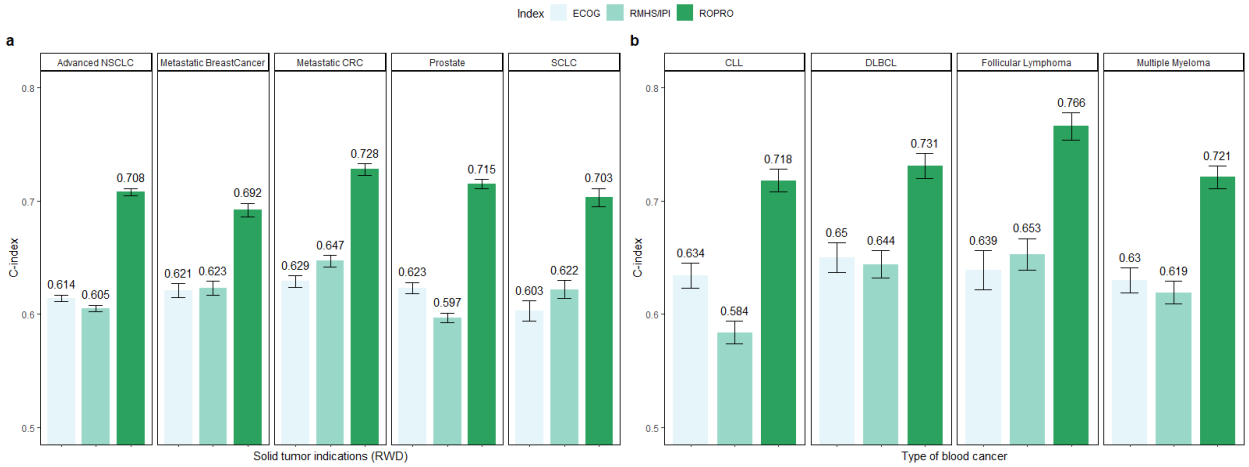
the variables into the formula. Units, variable coding, and log-transformation of a subset of the variables have to be considered prior to using the formula, as described by Becker et al.[10]. In case for a particular patient a particular variable measurement was missing, the respective term of the formula was omitted for the respective patient (equivalent to mean imputation with FH mean). ROPRO values range typically from -2 to 2, are distributed approximately normal, with higher ROPRO values indicating higher risk. For Kaplan-Meier analysis, we dichotomized ROPRO variables for better visibility. The upper 10%-quantile was assigned to be the high-ROPOR quantile. With this assignment the number of high ROPRO patients roughly corresponded to the number of high RMHS/IPI/ECOG patients. We used the full ROPRO range for computation of C-index and ROC-AUC values.

All scores were evaluated under time-to-event (death) analysis using Cox regression[12]. We used the R survival package[13], in particular the survfit() function for survival times and the coxph() function for Cox regression[12]. Median survival times were computed using the methodology as described in[14] and visualized as Kaplan-Meier plots[15]. From the models, we report the concordance index (C-index)[16] and ROC-AUC for survival analysis[17] as performance measures. For each score, Kaplan-Meier plots[15] are shown on cohort/study level, contrasting survival curves per high/low-class of the respective score.

Results

Figure 1a shows the C-index from Cox regression analysis of the prognostic scores on time to death for 5 solid tumor indications. In all indications, RMHS, ranging from 0.597 (SE 0.004) to 0.647 (SE 0.005) and ECOG, 0.603 (SE 0.009) to 0.629 (SE 0.005), showed similar performance as measured by C-index. ROPRO showed stronger prognostic power, with C-index ranging from 0.692 (SE 0.006) for MetBC to 0.728 (SE 0.005) for MetCRC. Noteworthy, C-index was also strong for MetPC (0.715, SE 0.004), although the indication was not part of the ROPRO discovery analysis[10]. Analogous observations were made for the blood cancers in the RWD, figure 1b. Here, IPI ranging from 0.584 (SE 0.01), to 0.653 (SE 0.013) and ECOG, 0.630 (SE 0.011) to 0.650 (SE 0.013), showed similar performance, while ROPRO showed stronger prognostic power, with C-index ranging from 0.718 (SE 0.011) for CLL to 0.766 (SE 0.012) for FL.

Figure 1. Model performance (C-index with standard error) of OS prognostic scores in RWD, by indication



In Roche solid tumor trials (Table 2), ECOG had medium prognostic power. In solid tumor trials, RMHS and ROPRO had good prognostic power. ROPRO outperformed RMHS in all studies considered. The difference was strongest in aNSCLC. In the OAK study, for instance, ROPRO had a C-index of 0.66 (SE 0.01) while RMHS reached 0.56 (SE 0.01)

In hematology trials, a difference by indication was observed. ROPRO outperformed IPI and ECOG in DLBCL and NHL. In the largest DLBCL study, GOYA, ROPRO had a C-index of 0.66 (SE 0.02) as compared to 0.61 (SE 0.02) for IPI. In the CLL-trial BO25323, however, IPI (0.64, SE 0.003) and ROPRO (0.63, SE 0.003) had similar prognostic power.

Table 2. Model performance (C-index with standard error) of OS prognostic scores in clinical trials

Trial	Study name	Cancer type	Indication	ECOG	IPI/RMHS	ROPRO
BO21005	GOYA	blood cancer	DLBCL	0.59 (0.02)	0.61 (0.02)	0.66 (0.02)
BO21223	GALLIUM	blood cancer	NHL	0.54 (0.01)	0.63 (0.02)	0.72 (0.02)
GO27878	-	blood cancer	NHL & DLBCL	0.58 (0.04)	0.55 (0.05)	0.69 (0.05)
BO25323	-	blood cancer	CLL	0.56 (0.03)	0.64 (0.03)	0.63 (0.03)

Trial	Study name	Cancer type	Indication	ECOG	IPI/RMHS	ROPRO
BO29554	BFAST	solid tumor	aNSCLC	0.56 (0.01)	0.6 (0.02)	0.71 (0.02)
GO28915	Oak	solid tumor	aNSCLC	0.58 (0.01)	0.56 (0.01)	0.66 (0.01)
GO29431	IMpower 110	solid tumor	aNSCLC	0.56 (0.01)	0.58 (0.01)	0.67 (0.02)
GO29436	IMpower 150	solid tumor	aNSCLC	0.58 (0.01)	0.57 (0.01)	0.65 (0.01)
GO29437	IMpower 131	solid tumor	NSCLC	0.56 (0.01)	0.56 (0.01)	0.62 (0.01)
GO29438	IMpower 132	solid tumor	aNSCLC	0.57 (0.01)	0.6 (0.01)	0.65 (0.01)
GO29527	IMpower 010	solid tumor	aNSCLC	0.55 (0.02)	0.53 (0.02)	0.56 (0.02)
GO29537	IMpower 130	solid tumor	aNSCLC	0.56 (0.01)	0.59 (0.01)	0.67 (0.01)
GO30081	IMpower133	solid tumor	SCLC	0.53 (0.01)	0.58 (0.02)	0.63 (0.02)
GO30182	COTEZO IMblaze370	solid tumor	MetCRC	0.56 (0.02)	0.63 (0.02)	0.69 (0.02)
MO39196	IMpassion131	solid tumor	MetBC	0.58 (0.01)	0.55 (0.01)	0.61 (0.02)
WO29522	IMpassion130	solid tumor	MetBC	0.57 (0.01)	0.59 (0.01)	0.63 (0.01)

Prognostic power became visible also when ROC-AUC for 1/3/6/12/24-months survival were considered. In figure 2 (RWD) and figure 3 (clinical trials), we depict ROC-AUC values over time. In RWD solid tumors (figure 2a), ROPRO could forecast 3-month up to 1 year survival very well, ROC-AUC values were above 0.80 up to 6-months survival and around 0.75 for 1-year survival. Two-year survival ROC-AUC values were lowest, but still at above 0.70 for all indications. ROPRO uniformly performed better than the RMHS, followed by ECOG, irrespective of the indication, trial, or time window considered. For RWD blood tumors (figure 2b), observations were very similar. ROPRO outperformed IPI, followed by ECOG. The performance of IPI in blood cancers was comparable to that of RMHS in solid tumors.

Figure 2a. ROC-AUC curves, varied over survival time cut-offs t (days), in RWD solid tumor

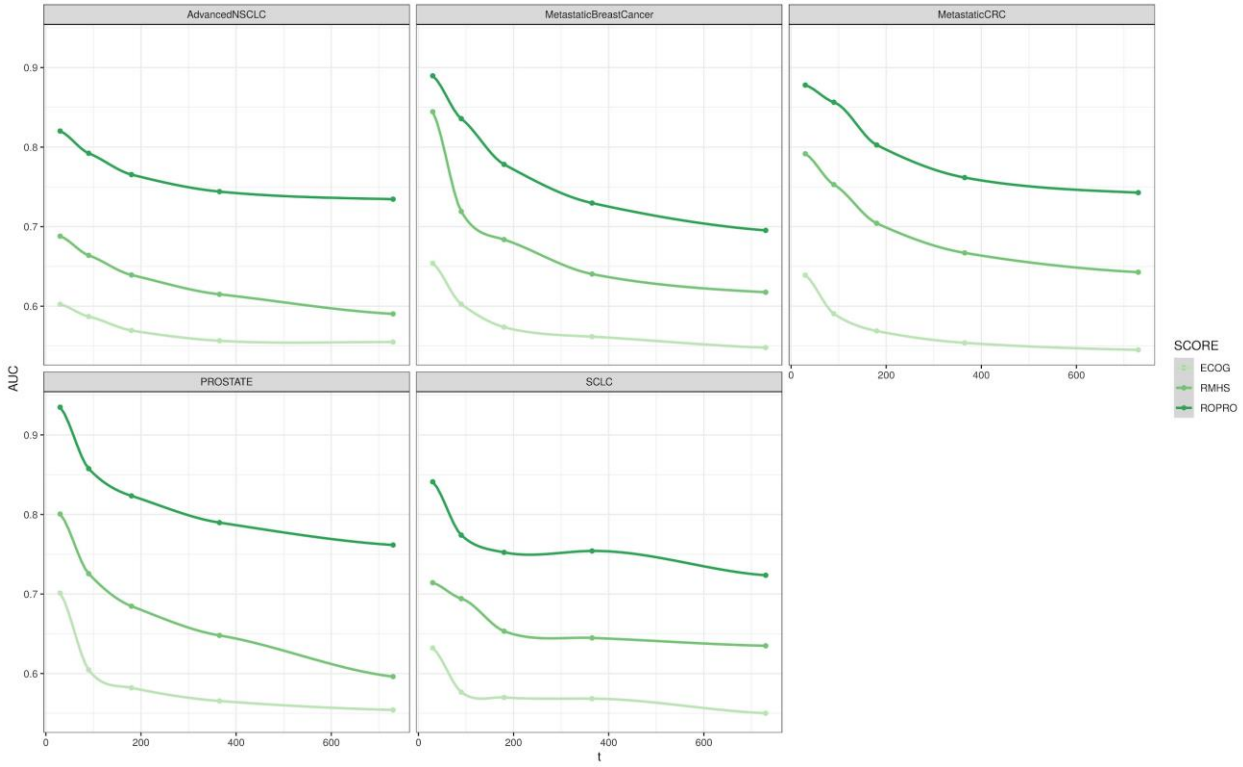
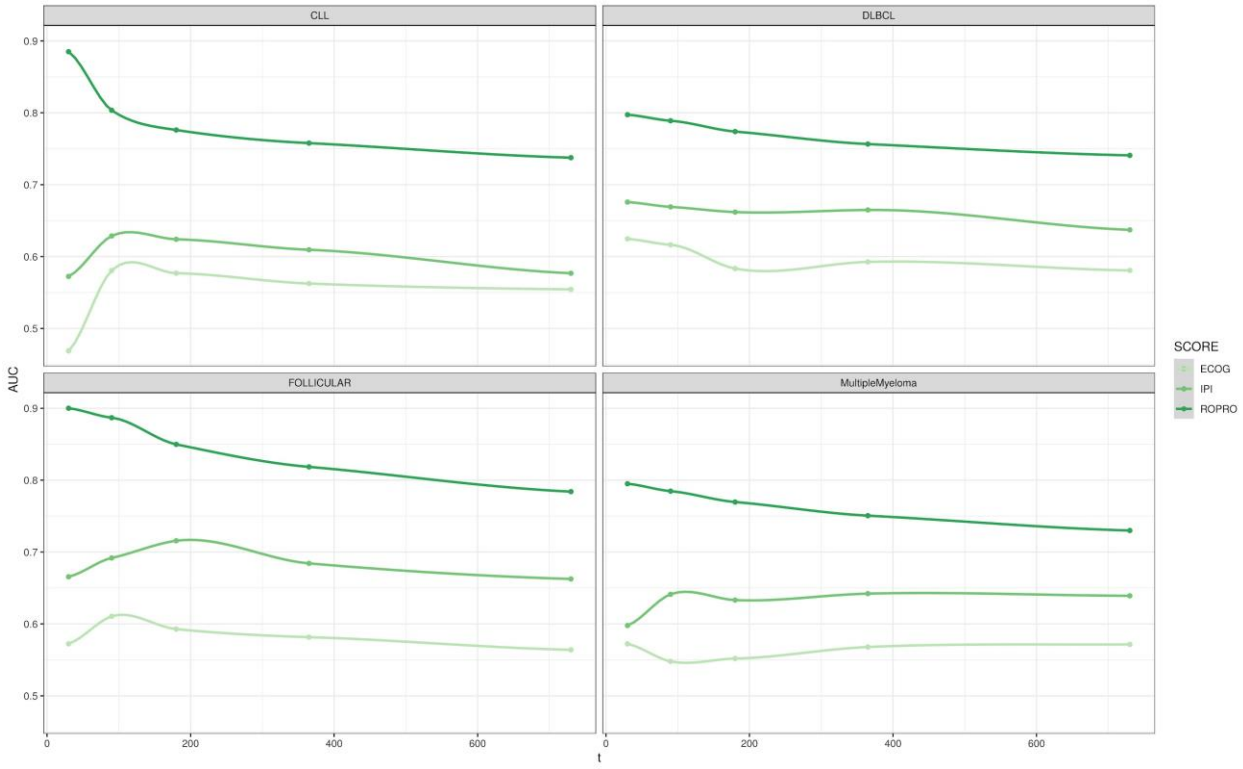


Figure 2b. ROC-AUC curves, varied over survival time cut-offs t (days), in RWD blood cancer



In solid tumor trials (figure 3a), ROPRO survival ROC-AUC were higher than those of RMHS and ECOG. Performance in MetBC and SCLC was somewhat lower than in other indications. Interestingly, ECOG had similar performance as the RMHS in aNSCLC, in contrast to RWD results. A possible reason for this is a higher missing rate of ECOG in RWD, as compared to clinical trials where ECOG values were nearly complete.

In blood cancer trials (figure 3b) of DLBCL/NHL, ROPRO survival ROC-AUC were higher than those of IPI and, next, ECOG, consistent with the results for RWD. In the CLL trial, ROPRO and IPI outperformed ECOG. ROPRO and IPI had similar performance for 3-months survival ROC-AUC. For 6/12-months survival, IPI outperformed ROPRO.

Figure 3a. ROC-AUC curves, varied over survival time cut-offs t (days), in clinical solid tumor trials, aggregated over indication

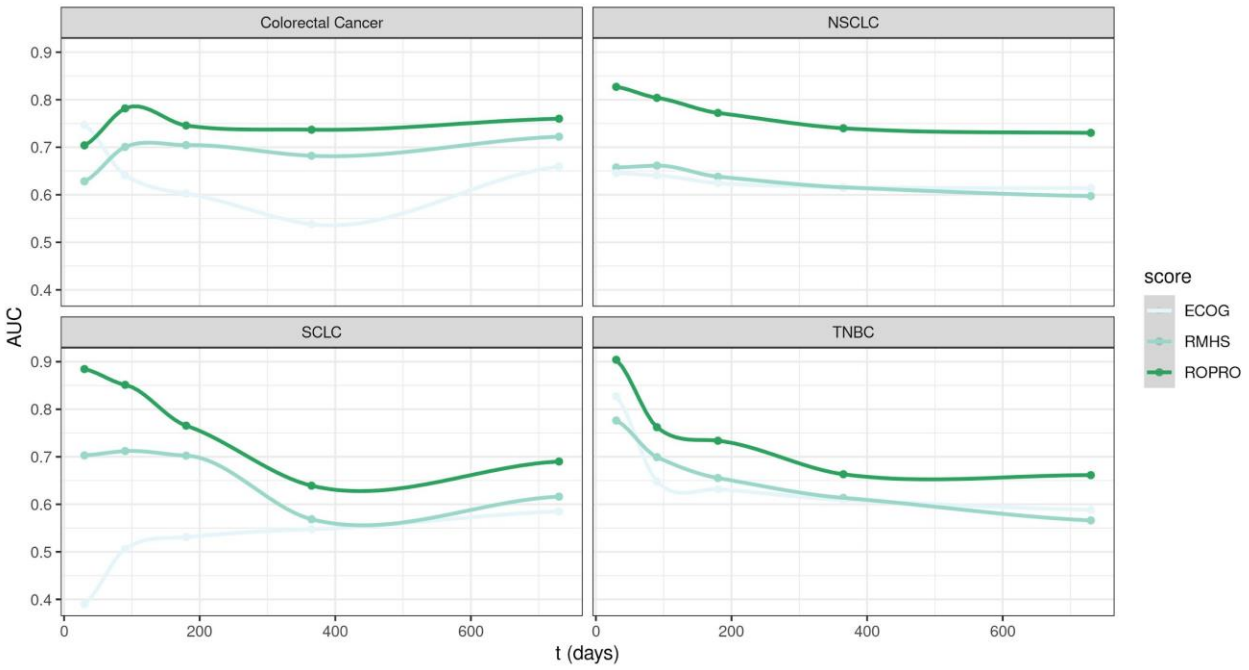
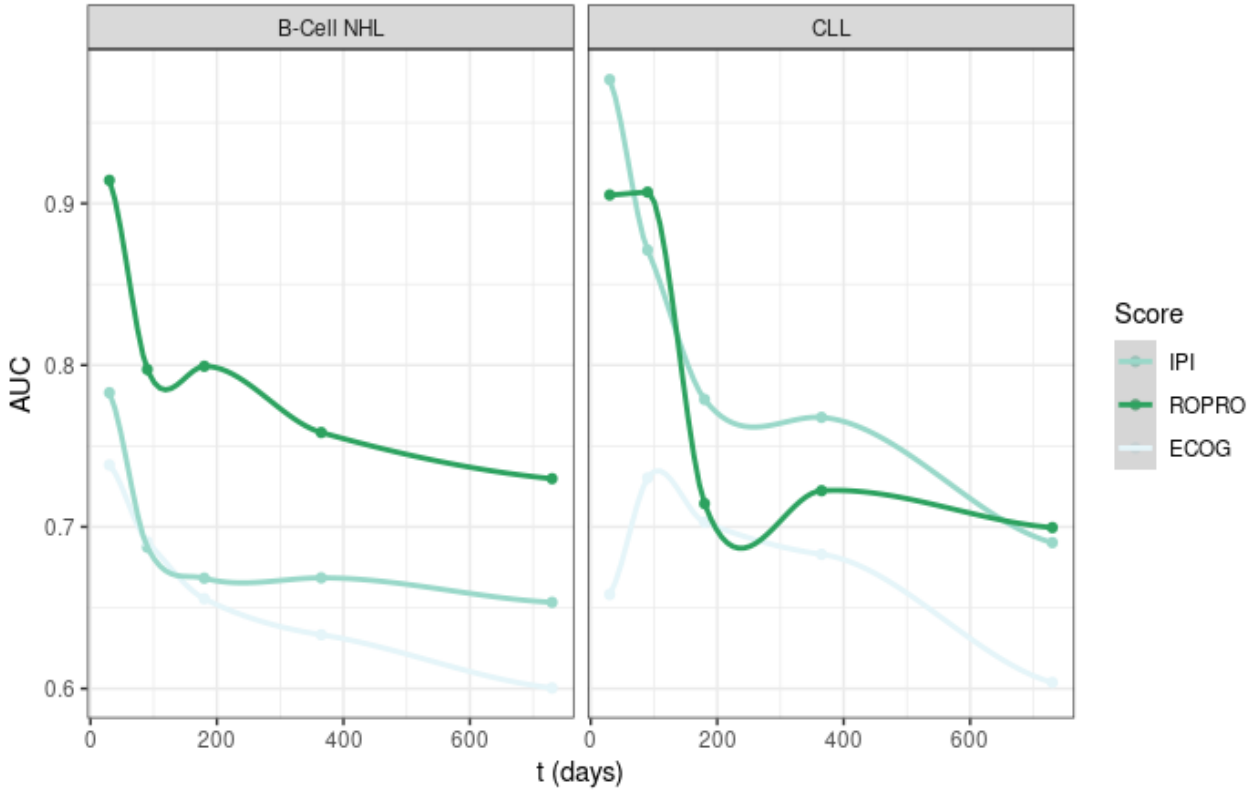


Figure 3b. ROC-AUC curves, varied over survival time cut-offs t (days), in clinical blood cancer trials, aggregated over indication



Finally, we also made a first investigation whether prognostic power of the scores was dependent on the type of treatment administered, using the Roche trial data. In figures 4 (aNSCLC, OAK study) and 5 (DLBCL, GOYA study) we contrast high/low ROPRO and high/low RMHS/IPI, each facet showing one treatment arm of a clinical trial.

In the aNSCLC study, the high-ROPRO class (red curve) had lowest median survival, in both treatment arms. Median survival was reduced also in high-RMHS patients. Median survival of low-ROPRO and low-RMHS patients was not distinguishable.

Likewise in GOYA (DLBCL), the high-ROPRO class (red curve) had lowest median survival, in both treatment arms. Median survival was reduced also in high-IPI patients, but the difference to high-ROPRO patients was substantial. Median survival of low-ROPRO and low-RMHS patients was not distinguishable.

Figure 4. Kaplan-Meier plots for high/low ROPRO and high/low RMHS in OAK study, by treatment arm. Docetaxel (a) and Atezolizumab (b).

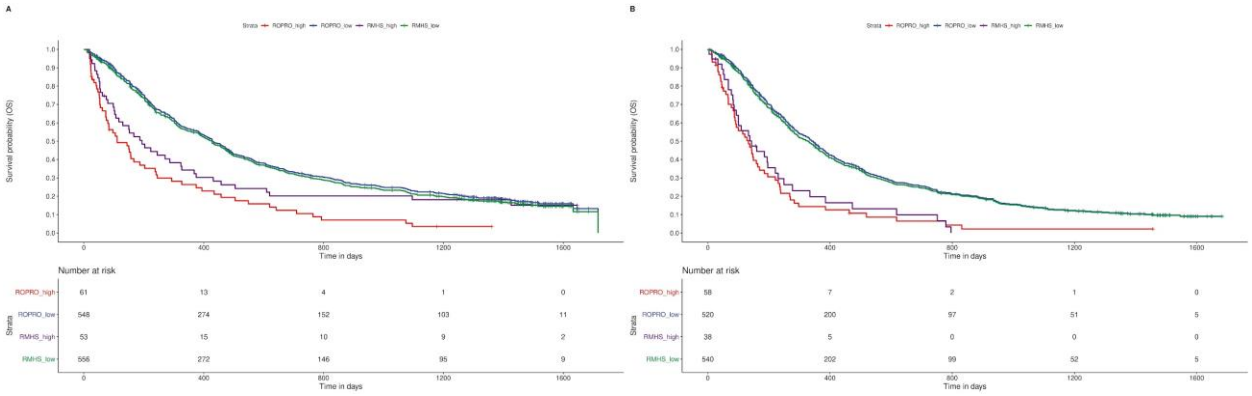
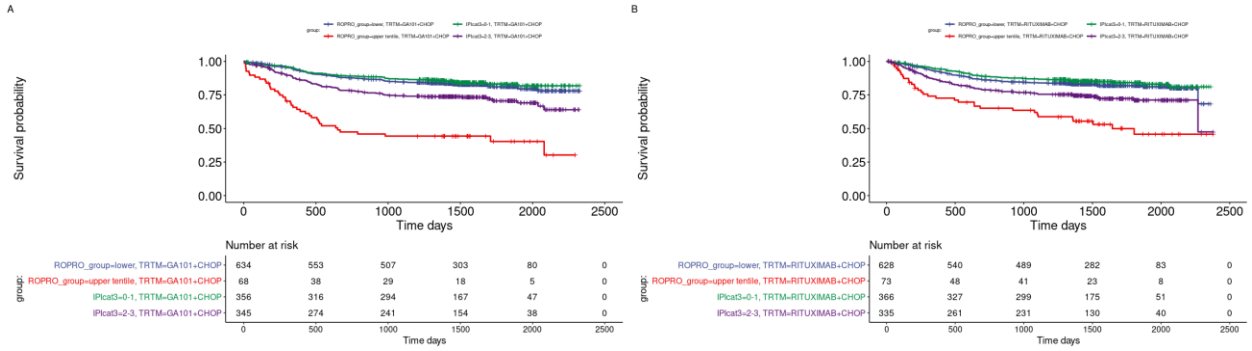


Figure 5. Kaplan-Meier plots for high/low ROPRO and high/low IPI in GOYA study, by treatment arm. Rituximab+CHOP (a) and Obinutuzumab+CHOP (b).



Discussion

Our analysis demonstrated the existence of easy-to-apply prognostic scores for OS and confirmed their power, in previously unseen patient data, and, in the case of metastatic prostate cancer, in a previously not analyzed solid tumor indication. Overall, the usage of all investigated scores can be recommended. With moderate extra effort, the implementation of the ROPRO model can create substantial improvement in predicting OS compared to other existing prognostic models. It describes very well the baseline prognosis of patients and can predict 3-month to 1-year survival particular well. Therefore, ROPRO can be used as an inclusion criteria in phase 1 studies[18] and should be evaluated for its potential as a decision making tool, e.g. fit-for chemotherapy decisions, in the future.

All scores proved to be prognostic in both RWD and clinical trial data. In general, the performance was stronger in RWD. A likely explanation is the more homogeneous patient samples in clinical trials. Several of the variables of the score are used as trial inclusion/exclusion criteria and patients with particular bad health status (ECOG values of 2 or higher) are not part of clinical trials. Since differentiating prognosis within more homogeneous patient data sets is generally

harder, the performance measures tended to be weaker in the clinical trials. It was also northworthy that ROPRO and IPI performed similarly in trial data while there was a substantial difference in RWD. Besides the homogeneity effect in the clinical trial data, which will reduce the benefit of a more complex score like the ROPRO over a less complex score, it is also not unlikely that the performance potential of IPI in RWD is slightly underestimated. In particular the number of extranodal sites is measured at some uncertainty and might impact the result. While this variable is also part of the ROPRO, the ROPRO score can compensate for the shortcoming since it considers a large panel of variables simultaneously. The wide applicability of the ROPRO score was also demonstrated in the clinical trials where it proved to be prognostic in all types of treatment. Therefore, it can also be used to stratify high risk patients for dedicated analysis.

Practical considerations

The ROPRO model combines clinical parameters that are typically available in routine clinical practice and, therefore, can be readily applied to new datasets. It requires 27 variables to be entered into a formula, an additional effort. However, given the substantial improvement in prognostic power, we consider this additional effort to be well justified. In any case, the ROPRO is a strong candidate for prognostic enrichment strategies in clinical trial design[19].

Limitations

Uncertainties and inaccuracies are expected to be encountered with retrospective RWD analysis. Our findings suggest, however, that RWD does provide valuable and stable insights: first, there is high consistency within the RWD, the ROPRO score remained strongly prognostic in the patient collection analyzed here, which was independent of the original discovery cohort of ROPRO[10]. Moreover, the score formula could be successfully transferred to a previously unseen indication (MetPC). Second, the prognostic score could be also validated in a series of clinical trials, again with high consistency across indications.

ECOG performance status demonstrated prognostic power, but performed worse than IPI/RMHS, except for the case of aNSCLC clinical trials. The prognostic power of ECOG might be somewhat underestimated. In RWD, ECOG performance status had a rather high missing rate. In clinical trials, the range of ECOG values was typically limited to 0 and 1, due to study protocol criteria, thereby limiting the prognostic range.

The list of prognostic scores which we investigated is not complete. For renal cell carcinoma (RCC), for instance, the International Metastatic Renal-Cell Carcinoma Database Consortium Score (IMDC)[20] is an established score which could not be investigated for lack of access to sufficiently large cohorts. Likewise, it was not possible to explore the Glasgow Prognostic Score (GPS) for cancer outcomes[21], for complete lack of availability of the CRP biomarker in the RWD and incompleteness in trial data. Furthermore, established indication-specific prognostic and predictive genetic markers[22-24] and tumor markers were not included for a lack of a

sufficient amount of data. We did not explore circulating tumor DNA (ctDNA)[25] since respective data is much more difficult to generate and since our focus was on comparatively easy to apply methods. Patient reported outcomes[26] were not included since they were not available in RWD and not in all clinical trials. Overall, the prognostic scores considered here cover host fitness in general, rather than indication or tumor biology specific features. Therefore, more specific biomarkers[22-24] will very likely further improve the already good prognostic power of the generic scores investigated here.

Future potential

Beyond the evaluation of the prognostic power of risk scores, we discuss future potential of strong biomarkers, both in research and development as well as in clinical practice. Here, we wish to *sketch* respective concepts. The actual realization of the ideas we are going to present will require thorough investigation and substantial validation efforts in the future.

Patient enrichment in phase I studies: oncology phase I studies are typically conducted with patients for whom no standard therapy option is left. As a consequence, phase I populations are often very vulnerable and have a short median survival perspective[27]. A considerable number of patients die within a short time, often in terms in which a novel drug does not even have the chance to become effective[28]. Such patients might be uninformative and lead to an overall wrong conclusion about drug efficacy, since a potential benefit in a general patient might become invisible. To overcome this issue, a life expectancy of at least 12 weeks, as judged by the study physician, is a patient exclusion criterion in phase I trials[29]. With a strong OS biomarker, an alternative approach to assess the 12-week survival prognosis is possible. Recently, we implemented the ROPRO as an inclusion criteria in a phase 1 study[18] which is currently conducted by Roche. Here, a ROPRO score above a calibrated cut-off (0.7) is used to identify patients that are not eligible for the study. Since the study was the first respective application, physicians have the possibility to overrule the recommendation implied by the biomarker, if they judge that a patient has a better perspective.

External control arms: randomized control trials constitute a standard for the comparison of treatment arms. Nevertheless, a randomized study may not always be feasible[30] and using external control data can be "an effective way to expand the interpretability of the results of an experimental arm by introducing the ability to carry out a formal or an informal comparative analysis.", as the U.S. Food and drug administration (FDA) states[31]. In addition, it was suggested that external control arms can be used in settings where control arms are not part of the standard design paradigm, as for instance in phase I trials[32]. When working with external controls, it is essential to compare treatment data against patients that have a comparable prognosis and to avoid systematic differences. In this context, a powerful OS prognostic marker is a valuable tool, as it allows to match patients according

to their OS risk assessment at baseline. Matching and stratification efforts based on obvious and strong clinical confounders have been discussed in[32]. With matching by an OS biomarker, matching balance can be improved and bias in treatment against external control comparisons can potentially be reduced. In particular, real world data can serve as a resource to construct external control arms[33-34].

Enrichment in phase III studies: patient enrichment can also be an option in phase III studies. The phase I enrichment strategy discussed above, postulates that there are patients which are effectively incurable by any realistic treatment approach. Such patients will also be part of phase III studies, in particular in studies in an advanced tumor setting. In a randomized phase III trial, "untreatable" patients will be distributed randomly across arms and, therefore, will not create a systematic bias. However, in case there is a real benefit of the treatment over the control in patients who are still accessible by treatment, the true signal will be diluted. Technically speaking, the power of a phase III study is reduced and false negative results are possible. Exclusion of patients with particularly bad prognosis, according to the OS biomarker, might remove the loss in power.

Adverse enrichment in phase III studies: Study duration is long in cancer indications with an overall good survival prognosis. The Cleopatra study, "A Study to Evaluate Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer"[35] detected a treatment benefit as measured by progression-free survival at an early stage[36]. Due to the overall comparatively good OS prognosis in breast cancer, consistent results on median overall survival could be published only three years later[37]. In general, such delay and long study duration time is to be expected in any indication setting with a long survival time. With the availability of a strong OS prognostic marker, it is conceivable to reduce a trial to patients which have an elevated death risk (but are still expected to be potentially receptive to treatment). With this "adverse" enrichment, median overall survival time is reached earlier and evidence for OS efficacy can be gained earlier. Adverse enrichment might either be part of the study design or can, alternatively, be applied as a secondary endpoint subgroup analysis.

Study evaluation: drug efficacy assessment by RECIST[38] criteria is a standard. While response events are a guideline for internal decision making in phase I studies[39], progression-free survival is an OS surrogate endpoint in later stages[40]. An advantage of a biomarker like ROPRO, is its availability over time, at dense intervals. Indeed, the underlying variables are repeatedly measured in clinical trials, typically at least once per treatment cycle[41]. Consequently, changes and improvements might be detected earlier than tumor changes as measured by RECIST, just for logistic reasons. For phase I studies, earlier decision making is a hypothetical benefit of the evaluation of biomarker time course. A formal, systematic evaluation of a longitudinal biomarker as an OS surrogate marker[42-44]

would also leverage potential for later stages. At this time, surrogacy validation is technically and logistically challenging. Modeling within a joint model framework is required[45], could be an option to address OS biomarker time course alongside actual survival events. Establishment as a surrogate endpoint would then require international collaboration across pharmaceutical companies. Before such organizational effort is a realistic option, more groundwork needs to be conducted.

Patient monitoring and treatment decisions: In the future, the ROPRO score might also be used to monitor and reassess a patient's prognosis longitudinally over the course of treatment, since its variables are routinely and frequently measured. The quantitative nature of the score can allow detecting changes more readily than with the IPI or RMHS: extranodal/metastatic sites and tumor stage are measured in larger time intervals. Moreover, age is an important baseline prognostic variable of the IPI score, but has no practical value in longitudinal assessment. More thorough analysis and validation studies will be required to investigate the potential of longitudinal assessment of the prognostic scores. Ideally, biomarkers cut-offs can be identified which can indicate the need for a treatment change. Also primary treatment decisions might be supported by OS biomarker status. Future research is necessary to investigate if fitness for chemotherapy or potential responsiveness to immunotherapy can be measured.

Life-style biomarkers: so far, the OS biomarkers presented here were evaluated in patients with typically substantially reduced life expectancy. On the other hand, in particular ROPRO, measures overall fitness and it would be highly interesting to investigate the potential in long-term prognosis. Population bases resources as the UK biobank[46] could be leveraged to explore the actual value in a general population.

Conclusion

Strong and easy-to-apply OS prognostic scores exist. The usage of all investigated scores can be recommended. With moderate extra effort, the implementation of ROPRO can create considerable improvement over existing prognostic scoring systems. Further improvement is likely to be achieved with specific biomarkers which could not be investigated here.

Corresponding Author:
A. Bauer-Mehren
Data & Analytics, Pharma Research and Development, Roche Innovation Center Munich
Email: anna.bauer-mehren@roche.com

Acknowledgements:
The authors thank Jessica Davies, PHC Data Science, Roche Products Ltd, Welwyn, UK for valuable input and discussions regarding methodology and results.

Funding:
This study was funded by F. Hoffmann-La Roche LTD. No grant number is applicable.

Disclosure:
All authors are employed or contracted by Roche.

Data Availability:
The data presented is owned by either Flatiron Health (real world data oncology database) or F. Hoffmann-la Roche LTD (clinical studies). Access to Flatiron Health can be licensed.

References:

1. Halabi S, Owzar K. The importance of identifying and validating prognostic factors in oncology. Semin Oncol. 2010;37(2):e9ee18.

2. Meera N, Sandhu SS, Sharma AK. Prognostic and Predictive Biomarkers in Cancer. Current Cancer Drug Targets. 2014; 14(5): 477-504.

3. Tarighati E, Keivan H, Mahani H. A review of prognostic and predictive biomarkers in breast cancer. Clin Exp Med. 2022. Online ahead of print.

4. Nieder C, Dalhaug A. A new prognostic score derived from phase I study participants with advanced solid tumours is also valid in patients with brain metastasis. Anticancer Res. 2010; 30(3):977-979.

5. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. N Engl J Med. 1993; 329(14): 987-994.

6. https://ecog-acrin.org/resources/ecog-performance-status/

7. https://www.mdcalc.com/calc/2320/follicular-lymphoma-international-prognostic-index-flipi

8. Al-Sawaf O, Can Zhang C, Maneesh Tandon M, Arijit Sinha A, Anna-Maria Fink AM, Sandra Robrecht S et al. Venetoclax plus obinutuzumab versus chlorambucil plus obinutuzumab for previously untreated chronic lymphocytic leukaemia (CLL14): follow-up results from a multicentre, open-label, randomised, phase 3 trial. The Lancet. 2020; 21(9):1188-1200.

9. Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products. Available at: http://www.fda.gov/regulatory-information/search-fda-guidance-documents/enrichmentstrategies-clinical-trials-support-approval-human-drugs-and-biologicalproducts. Accessed August 6, 2020.

10. Becker T, Weberpals J, Jegg AM, So WV, Fischer A, Weisser M, Schmich F, Rüttinger D, Bauer-Mehren A. An enhanced prognostic score for overall survival of patients with cancer derived from a large real-world cohort. Ann Oncol. 2020;31(11):1561-1568.

11. https://flatiron.com

12. Enderlein G, Cox DR, Oakes D. Analysis of Survival Data. Biom J. 1987; 29(1):114–114.

13. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model, New York: Springer-Verlag, 2000.

14. Kalbfleisch, JD, Prentice, RL. The statistical analysis of failure time data, Second Edition, Wiley Series in Probability and Statistics, 2002.

15. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958; 53(28):457–81.

16. Steck H, Krishnapuram B, Dehing-Oberije C et al. On Ranking in Survival Analysis: Bounds on the Concordance Index. In Platt JC, Koller D, Singer Y, Roweis ST (eds): Adv. Neural Inf. Process. Syst. 20, Curran Associates, Inc., 2008:1209–1216.

17. https://cran.r-project.org/web/packages/timeROC/timeROC.pdf

18. https://clinicaltrials.gov/ct2/show/NCT04158583

19. Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products. Available at: http://www.fda.gov/regulatory-information/search-fda-guidance-documents/enrichmentstrategies-clinical-trials-support-approval-human-drugs-and-biologicalproducts.Accessed August 6, 2020.

20. Ko JJ, Xie W, Kroeger N, Lee JL, Rini BI et al. The International Metastatic Renal Cell Carcinoma Database Consortium model as a prognostic tool in patients with metastatic renal cell carcinoma previously treated with first-line targeted therapy: a population-based study. Lancet Oncol. 2015; 16(3):293-300.

21. Forrest LM, McMillan DC, McArdle CS, Angerson WJ, Dunlop DJ. Evaluation of cumulative prognostic scores based on the systemic inflammatory response in patients with inoperable non-small-cell lung cancer. British Journal of Cancer. 2003. 89: 1028–1030.

22. Tarighati E, Keivan H, Mahani H. A review of prognostic and predictive biomarkers in breast cancer. Clin Exp Med. 2022.

23. Liu D. Cancer biomarkers for targeted therapy. Biomark Res. 2019. 7(25).

24. Nair M, Sandhu SS, Sharma AK. Prognostic and Predictive Biomarkers in Cancer. Curr Cancer Drug Targets. 2014. 14(5):477-504

25. Reichert ZR,Morgan TM,Li G,Castellanos E, Snow T, Dall'Olio FG et al. Prognostic value of plasma circulating tumor DNA fraction across four common cancer types: a real-world outcomes study. Ann Oncol. 2022; S0923-7534(22):04186-2.

26. Teixeira MM, Borges FC, Ferreira PS, Rocha J, Sepodes B, Torre C. A review of patient-reported outcomes used for regulatory approval of oncology medicinal products in the European Union between 2017 and 2020. Front Med (Lausanne). 2022;9:968272.

27. Chau, N.G., Florescu, A., Chan, K.K. et al. Early mortality and overall survival in oncology phase I trial participants: can we improve patient selection? BMC Cancer. 2011. 11:426.

28. Ivy SP, Siu LL, Garrett-Mayer E, Rubinstein L. Approaches to phase 1 clinical trial design focused on safety, efficiency, and selected patient populations: a report from the clinical trial design task force of the national cancer institute investigational drug steering committee. Clin Cancer Res. 2010. 16(6):1726-36.

29. Jin S, Pazdur R, Sridhara R. Re-Evaluating Eligibility Criteria for Oncology Clinical Trials: Analysis of Investigational New Drug Applications in 2015. Clin Oncol. 2017. 35(33):3745-3752.

30. https://doi.org/10.1016/j.annonc.2021.12.015

31. https://doi.org/10.1016/j.annonc.2021.12.015

32. Carrigan G, Whipple S, Capra WB, et al. Using electronic health records to derive control arms. for early phase single-arm lung cancer trials: proof-of-concept in randomized controlled trials. Clin Pharmacol Ther. 2020. 107(2):369-377.

33. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials - a primer for researchers. Clin Epidemiol. 2020.12:457-467.

34. Ventz S, Lai A, Cloughesy TF, et al. Design and evaluation of an external control arm using prior clinical trials and real-world data. Clin Cancer Res. 2019. 25(16):4993-5001.

35. https://clinicaltrials.gov/ct2/show/NCT00567190

36. Baselga J, Cortes J, Kim SB, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl J Med. 2012. 366:109-119

37. DOI: 10.1056/NEJMoa1413513

38. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst. 2000. 92:205-21

39. Aykan NF, Özatlı T. Objective response rate assessment in oncology: Current situation and future expectations. World J Clin Oncol. 2020. 11(2):53-73.

40. Belin L, Tan A, De Rycke Y, Dechartres A. Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. Br J Cancer. 2020. 122(11):1707-1714.

41. Goulart BH, Clark JW, Pien HH, Roberts TG, Finkelstein SN, Chabner BA. Trends in the use and role of biomarkers in phase I oncology trials. Clin Cancer Res. 2007. 13(22 Pt 1):6719-26.

42. Alonso A, Bigirumurame T, Burzykowski T, Buyse M, Molenberghs G, Muchene L, Perualila NJ, Shkedy Z, Van der Elst W. Applied surrogate endpoint evaluation methods with SAS and R. CRC Press. 2016.

43. Burzykowski T, Molenberghs G, Buyse M. The Evaluation of Surrogate Endpoints. 2005. Springer.

44. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000. 1(1):49-67.

45. Rizopoulos, Dimitris. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. 2012. CRC press.

46. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. Lancet. 2015;386(9993):533e540.