An Ensemble Approach to Predicting Overall Survival Benefit Jointly From Multiple Surrogate Endpoints in Oncology: A Case Study in Previously Untreated Metastatic Melanoma

CO29

Murat Kurt, Paul Serafini, Mir-Masoud Pourrahmat, Victoria Wan

Evidinno Research Outcomes Inc., Vancouver, British Columbia, Canada

Background

- Surrogate endpoints (SEs) play an increasingly important role in the drug development process as new health technologies are frequently being licensed by the regulatory agencies, such as the European Medicines Agency (EMA) or the Food and Drug Administration (FDA), based on early indicators of long-term efficacy rather than long-term outcomes.[1]
- SEs provide valuable early insights into overall survival (OS) benefits, particularly in settings where long follow-up times are needed for the collection of statistically mature OS data.
- Clinical trials often report multiple SEs which are evaluated individually for their prognostic role on OS and ability to predict OS benefit. Common bivariate meta-analytical methods evaluating the effect of a single SE on OS benefit include:
  - Daniels and Hughes models
  - Bivariate random-effects meta-analysis
  - Weighted linear regression
- Joint analysis of multiple SEs can capture their combined clinical potential and enhance the precision of early predictions for the treatment effects on OS
- Recent guidelines and studies proposing multivariate analysis in the context of SE modeling include:
  - NICE DSU TSD 20, which proposes multivariate meta-analysis (MVMA) to integrate data on multiple SEs and OS.[2]
  - Bujkiewicz et al. (2022), which introduced a Bayesian framework to extend MVMA.[3]
  - Wang et al. (2022), which proposed a calibrated model fusion approach for combining surrogate markers into a composite predictor. [4]
- Despite the guidelines and proposed innovative approaches, methods jointly evaluating multiple SEs in a single framework to optimize early prediction of OS benefits in randomized controlled trials (RCTs) have not been widely explored.

Objective

- To devise an approach blending in the available predictions obtained from individual SE models to improve the accuracy and reduce the uncertainty in the prediction of OS benefit.
- To explore the sensitivity of the predictive performance of the model with respect to mechanism of action of the experimental therapy in the RCTs

Methods

- The evidence base consisted of N=24 RCTs in previously untreated metastatic melanoma identified by a systematic review.[5] Ten trials in the evidence base investigated immune-checkpoint inhibitors (ICIs) as monotherapy or in combination with other agents as experimental treatment. Majority of the remaining trials investigated targeted therapies.
- Candidate SEs were progression-free survival (PFS), objective response rate (ORR), and complete response rate (CRR).
- Given predictions on OS hazard ratios (HRs) obtained from bivariate random effects meta-analyses individually for each SE in each RCT, an optimization model was developed and solved separately under two different objectives to elicit the weights to be assigned to the predictions of each individual SE.
- In the model, weights assigned to the predictions of each SEs were identical across all trials and indicated their exclusive probability of correctness relative to the predictions from other SEs.
- The model was solved under two different objectives: (1) The Euclidean norm objective aimed to minimize the sum of squared errors between the model-predicted and reported OS HRs across all RCTs, and (2) The supremum norm objective aimed to minimize the maximum gap between the model-predicted and reported OS HRs across all RCTs.
- Average absolute gaps between the model-predicted and reported OS HRs across all trials by the ensemble approach and the best/worst performing individual SE models were compared.
- Due to established positive correlation between the treatment effects on each SE and OS, it was possible to derive an upper bound on the standard error of the OS HR predicted from the ensemble approach. Therefore, improvements in maximum standard error around the predicted log-transformed OS HRs by the ensemble approach versus individual SEs were also measured.
- Predictive performance of the ensemble approach was tested externally on three Phase III trials in previously untreated advanced melanoma (IMSpire-170, PIVOT IO-001, and RELATIVITY-047) that were published after the conduct of systematic review.
- **Mathematical Formulation of the Model**
  - **Decision Variables:**

p_1: Probability that OS HR predicted from the PFS − based SE model is correct

p_2: Probability that OS HR predicted from the ORR − based SE model is correct

p_3: Probability that OS HR predicted from the CRR − based SE model is correct

x_i: Predicted OS HR for trial i by the ensemble approach

  - **Parameters:**

y_i: Observed OS HR for trial i

$z_{1,i}$: Predicted OS HR for trial i using PFS as an SE

$z_{2,i}$: Predicted OS HR for trial i using ORR as an SE

$z_{3,i}$: Predicted OS HR for trial i using CRR as an SE

  - **Constraints:**

$$p_1 + p_2 + p_3 = 1; p_1, p_2, p_3 \geq 0,$$

$$x_i = p_1 z_{1,i} + p_2 z_{2,i} + p_3 z_{3,i} \text{ (blended prediction for trial } i)$$

  - **Objectives:**

$$\text{Euclidean Norm: } \min \sum_{i=1}^{N} (x_i - y_i)^2$$

$$\text{Supremum Norm: } \min \max_{i=1,\ldots,N} |x_i - y_i|$$

Results

- The optimized weights for the predictions from the HRs of PFS, and the odds ratios of ORR and CRR were 0.254, 0.446, and 0.300, respectively under the Euclidean norm, and 0.606, 0, and 0.394, respectively under the supremum norm.
- Predictions from the ensemble approach improved the maximum standard error around the log-transformed OS benefit obtained across all individual SE models by 0.039 (14.5%, relatively) and 0.044 (18.6%, relatively) on average, under the Euclidean and supremum norm objectives, respectively.
- In the external validation across IMSpire-170, PIVOT IO-001, and RELATIVITY-047 trials, average absolute improvement in the maximum standard error by the ensemble approach was 0.022 (12.1%, relatively) and 0.033 (18.3%, relatively), under the Euclidean and supremum norm objectives, respectively.
- Regardless of the norm of the objective, the ensemble approach overpredicted the reported OS benefit in 14 out of 24 trials, whereas the overprediction of reported OS benefit was seen only in 12 trials from the PFS-based SE model, 10 trials from the ORR-based SE model, and 12 trials from the CRR-based SE model.
- On average, across 24 trials, the absolute gap between the reported and predicted OS benefit was 0.068 and 0.170 from the best- and worst-performing individual SE models. The average absolute gap between the reported and predicted OS benefit was 0.112 and 0.106 from the ensemble approach under the Euclidean and supremum norm objectives, respectively.
- In the external validation, the absolute gap between the reported and predicted OS benefit was 0.020 and 0.090 from the best- and worst-performing individual SE models. The average absolute gap between the reported and predicted OS benefit was 0.053 and 0.048 from the ensemble approach under the Euclidean and supremum norm objectives, respectively.
- In 7 of the 10 trials investigating ICIs, regardless of the norm of the objective, the ensemble approach overpredicted the reported OS benefit whereas the overprediction of reported OS benefit was seen only in 6 ICI trials from the PFS-based SE model, 4 ICI trials from the ORR-based SE model and 5 ICI trials from the CRR-based SE model.

Table 1: Comparison of reported OS benefit versus predicted OS benefits from the ensemble approach and individual SE models

Trial/Publication	Observed OS HR	Predicted OS HR from Individual SEs			Predicted OS HR from Ensemble Approach		Absolute Gap in the OS HR Prediction		
		From PFS	From ORR	From CR	Euclidean Norm*	Supremum Norm*	Euclidean Norm*	Supremum Norm*	Max Across Individual SE Models
Algazi (2020)	1.030	1.000	0.870	0.980	0.936	0.992	0.094	0.038	0.160
Ascierto (2017)	0.840	0.860	0.820	0.920	0.860	0.884	0.020	0.044	0.080
Avril (2004)	0.740	0.780	0.760	0.690	0.744	0.745	0.004	0.005	0.050
BREAK-3	0.550	0.630	0.600	0.760	0.656	0.681	0.106	0.131	0.210
BRIM-3	0.700	0.630	0.610	0.630	0.621	0.630	0.079	0.070	0.090
CheckMate 066	0.500	0.640	0.700	0.480	0.619	0.577	0.119	0.077	0.200
CheckMate 067	0.520	0.650	0.670	0.640	0.656	0.646	0.136	0.126	0.150
CheckMate 069	0.740	0.620	0.620	0.430	0.563	0.545	0.177	0.195	0.310
CheckMate 511	1.090	0.910	0.870	0.870	0.880	0.894	0.210	0.196	0.220
coBRIM	0.700	0.730	0.760	0.780	0.758	0.750	0.058	0.050	0.080
COLUMBUS	0.610	0.700	0.750	0.830	0.761	0.751	0.151	0.141	0.220
COMBI-d	0.800	0.790	0.790	0.840	0.805	0.810	0.005	0.010	0.040
COMBI-v	0.690	0.720	0.790	0.780	0.769	0.744	0.079	0.054	0.100
IMspire150	0.850	0.840	0.840	0.910	0.861	0.868	0.011	0.018	0.060
KEYNOTE-006	0.720	0.720	0.720	0.710	0.717	0.716	0.003	0.004	0.010
KEYNOTE-022	0.760	0.770	0.890	0.820	0.839	0.790	0.079	0.030	0.130
KEYNOTE-029	0.890	1.060	0.860	0.860	0.911	0.981	0.021	0.091	0.170
Lebbe (2020)	0.890	0.760	0.790	0.890	0.812	0.811	0.078	0.079	0.130
Middleton (2000)	1.180	1.000	0.860	0.880	0.901	0.953	0.279	0.227	0.320
NEMO	1.000	0.750	0.760	0.640	0.721	0.707	0.279	0.293	0.360
PACMEL	1.220	0.820	0.790	1.270	0.941	0.997	0.279	0.223	0.430
Patel (2011)	1.000	0.870	0.800	0.770	0.809	0.831	0.191	0.169	0.230
Robert (2011)	0.720	0.810	0.800	0.760	0.791	0.790	0.071	0.070	0.090
Weide (2019)	0.560	0.810	0.700	0.700	0.728	0.767	0.168	0.207	0.250

Table 2: Comparison of standard errors for the log-transformed OS benefit predicted from the ensemble approach and individual SE models

Trial/Publication	Max Std Error			Absolute Improvement in Std Error by Ensemble Approach		Relative Improvement in Std Error by Ensemble Approach	
	Euclidean Norm*	Supremum Norm*	Across Individual SE Models	Euclidean Norm*	Supremum Norm*	Euclidean Norm*	Supremum Norm*
Algazi (2020)	0.166	0.166	0.199	0.033	0.033	0.166	0.167
Ascierto (2017)	0.149	0.142	0.167	0.018	0.025	0.106	0.147
Avril (2004)	0.148	0.142	0.172	0.024	0.029	0.138	0.170
BREAK-3	0.153	0.136	0.174	0.021	0.038	0.119	0.220
BRIM-3	0.177	0.168	0.217	0.040	0.049	0.183	0.225
CheckMate 066	0.221	0.237	0.394	0.173	0.157	0.439	0.399
CheckMate 067	0.164	0.158	0.205	0.041	0.047	0.198	0.230
CheckMate 069	0.252	0.269	0.469	0.217	0.200	0.462	0.427
CheckMate 511	0.148	0.139	0.156	0.008	0.017	0.053	0.110
coBRIM	0.137	0.126	0.148	0.011	0.022	0.074	0.149
COLUMBUS	0.140	0.130	0.150	0.010	0.020	0.068	0.134
COMBI-d	0.139	0.128	0.149	0.010	0.021	0.067	0.142
COMBI-v	0.138	0.127	0.149	0.011	0.022	0.075	0.148
IMspire150	0.150	0.140	0.165	0.015	0.025	0.090	0.153
KEYNOTE-006	0.145	0.138	0.159	0.014	0.021	0.091	0.134
KEYNOTE-022	0.142	0.125	0.161	0.018	0.036	0.114	0.222
KEYNOTE-029	0.152	0.151	0.156	0.004	0.005	0.026	0.033
Lebbe (2020)	0.144	0.136	0.157	0.013	0.021	0.083	0.132
Middleton (2000)	0.154	0.149	0.158	0.004	0.009	0.026	0.054
NEMO	0.157	0.154	0.205	0.047	0.051	0.232	0.247
PACMEL	0.205	0.216	0.361	0.156	0.144	0.432	0.401
Patel (2011)	0.139	0.130	0.148	0.009	0.018	0.060	0.121
Robert (2011)	0.138	0.128	0.148	0.009	0.019	0.062	0.131
Weide (2019)	0.149	0.140	0.169	0.020	0.029	0.119	0.171

* The objective of the optimization model in the ensemble approach, Max: Maximum, Std: Standard, OS: Overall survival, HR: Hazard ratio, SE: Surrogate endpoint. Rows shaded in light green represent immune-checkpoint inhibitor trials.

Conclusions

- Our ensemble approach consolidates predictions from multiple SEs, maximizing data use and reducing uncertainty compared to single SE models.
- In the ensemble approach, standard error for the log-transformed OS benefit may be non-estimable if SEs are negatively correlated or if interaction terms are present.
- The proposed method is generalizable across tumor types with flexible SE inclusion. However, adding more SEs without interaction terms can introduce bias.
- A preliminary meta-analysis to assess each SE's correlation with OS is required, though sparse data may pose computational challenges.
- A comprehensive SE list is essential for optimal performance, especially in heterogeneous evidence bases across treatment classes.

References

1. Chen et al. *JAMA Internal Medicine.* 2020. 180(6), 912–914
2. Bujkiewicz S et al. *NICE DSU technical support document 20.* [Accessed October 7, 2024].
3. Bujkiewicz S et al. *Stat Med*. 2016; 35(7):1063-89.
4. Wang X et al. *Biometrics*. 2023;79(2):788-798.
5. Leung L et al. *Value in Health*. 2024;27(6):S46.

Acknowledgments

MK reports contract and PS, MMP and VW report employment with Evidinno Outcomes Research Inc.

EVIDINNO
EVIDENCE NAVIGATION & SYNTHESIS

Presented at ISPOR Europe 2024; November 17-20; Barcelona, Spain

Email: mkurt@evdinno.com