Supplementary materials for Meng *et al*., "**A machine learning approach to real-world time to treatment discontinuation prediction"**.

Figure S1. Performance of rwTTD prediction in homogeneous population during cross-validation by absolute errors. a. Absolute error at different termination rates. d. Absolute error with different numbers of training examples. e. Absolute error with different numbers of predictive features. f. Absolute error with different feature noise levels.

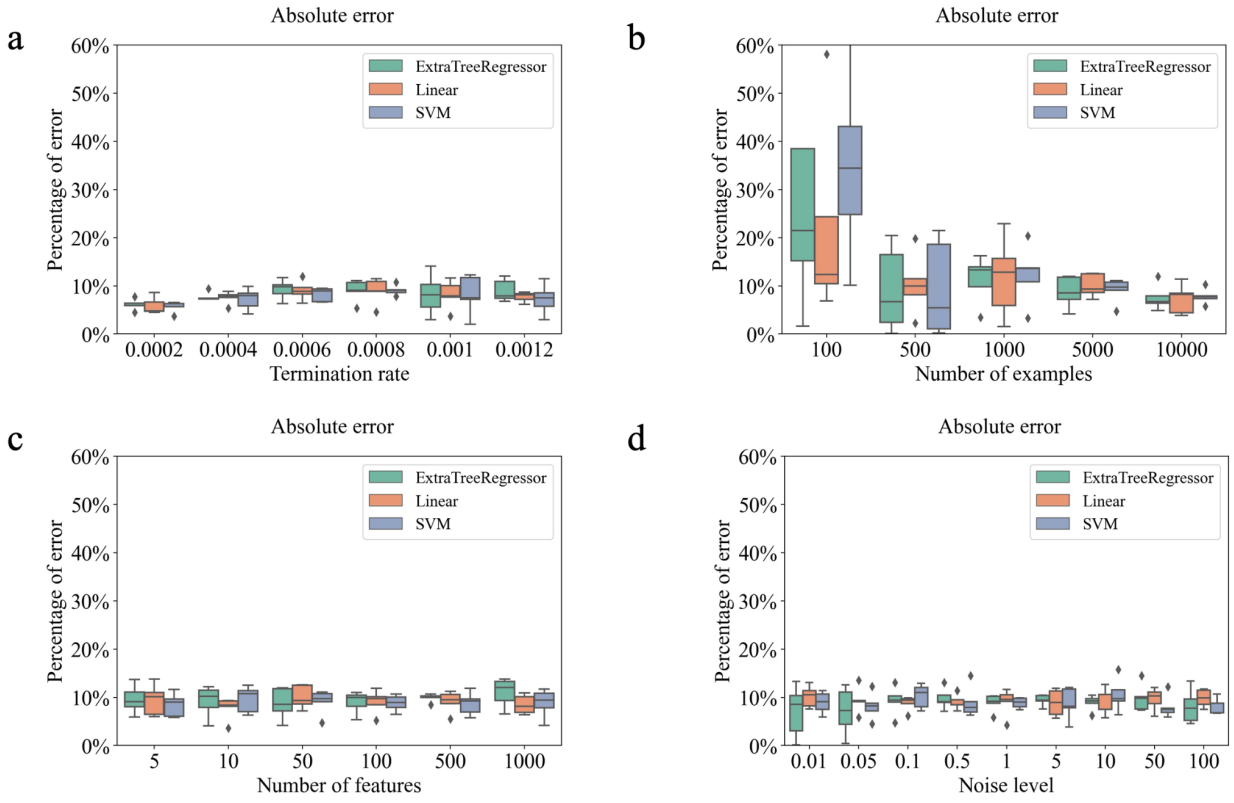

Figure S2. Predicted termination ratio vs. gold standard when changing termination rate. We used fold 1 of each cross-validation experiment to demonstrate the prediction accuracy.

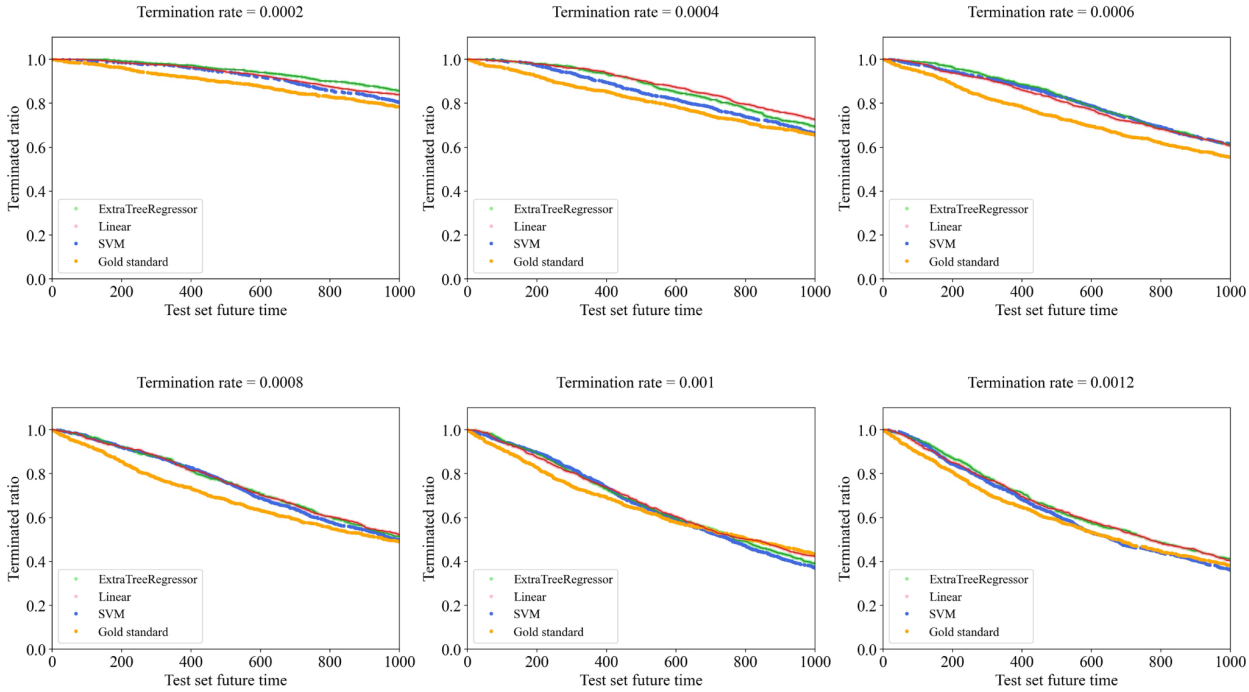

Figure S3. Predicted termination ratio vs. gold standard when changing the number of examples. We used fold 1 of each cross-validation experiment to demonstrate the prediction accuracy.

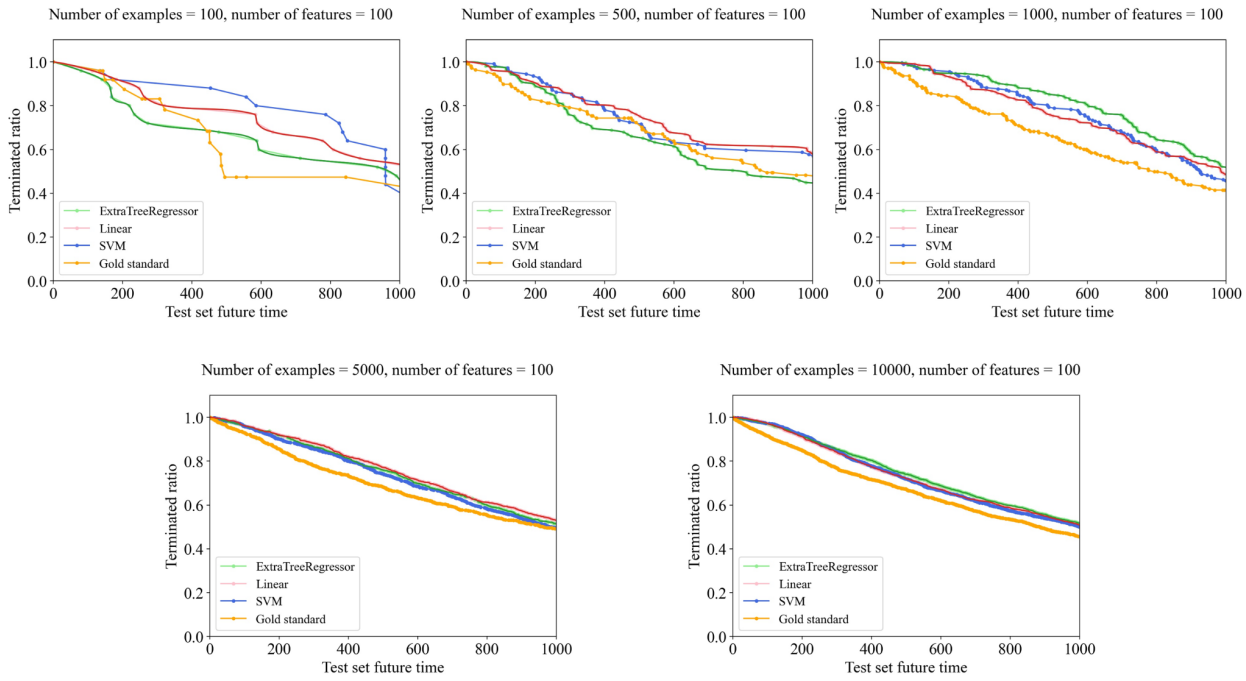

Figure S4. Predicted termination ratio vs. gold standard when changing the number of features. We used fold 1 of each cross-validation experiment to demonstrate the prediction accuracy.

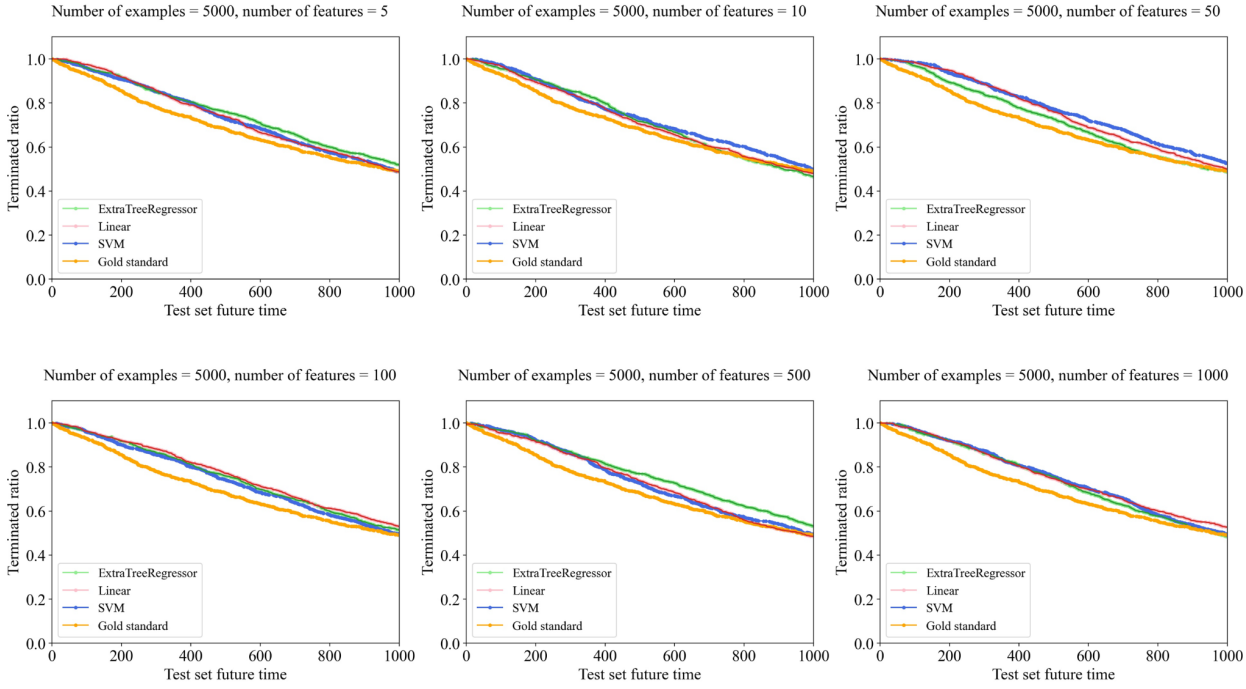

Figure S5. Predicted termination ratio vs. gold standard when changing the noise level. We used fold 1 of each cross-validation experiment to demonstrate the prediction accuracy.

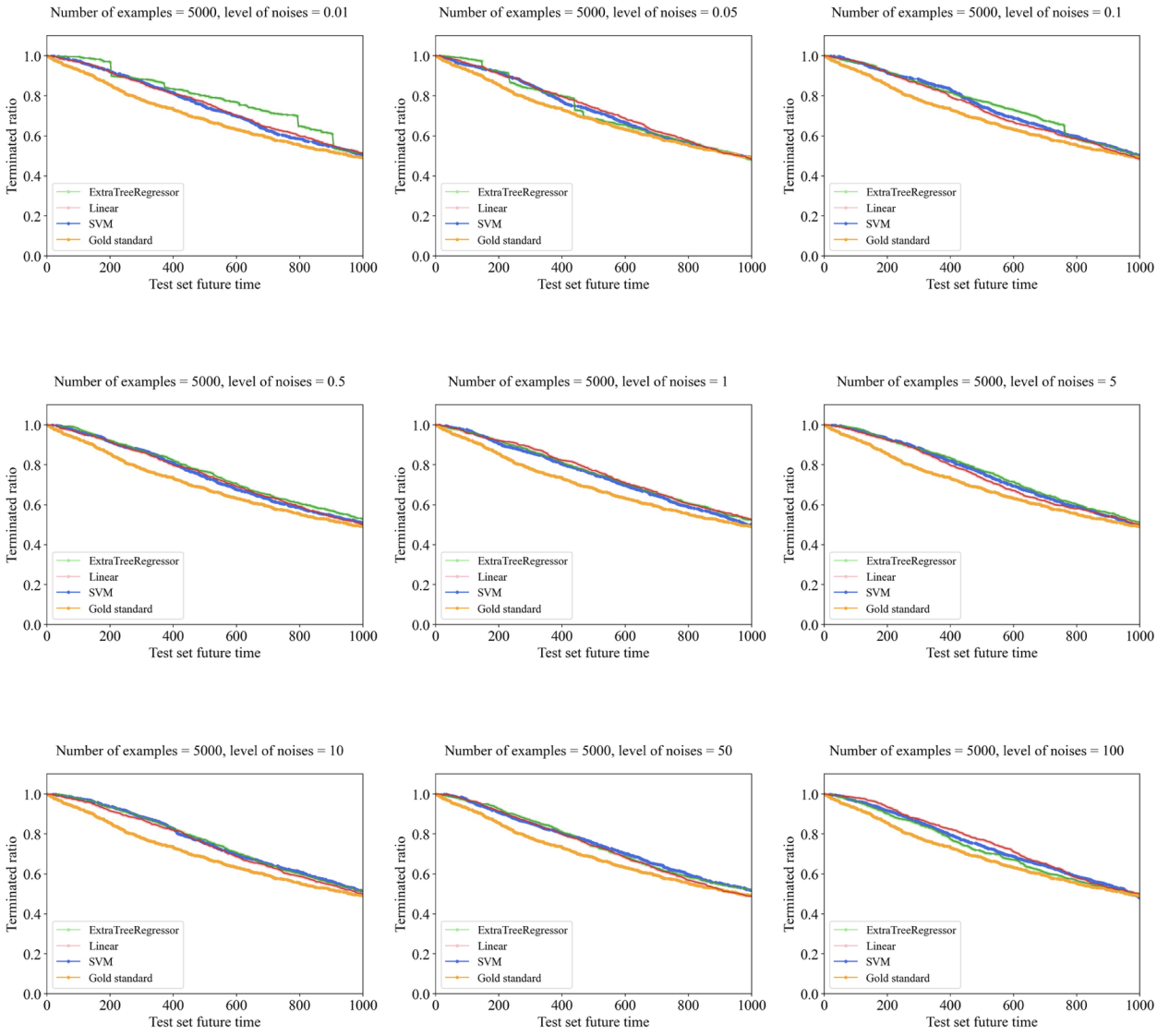

Figure S6. Performance comparison at different test termination rates for each training set termination rate.

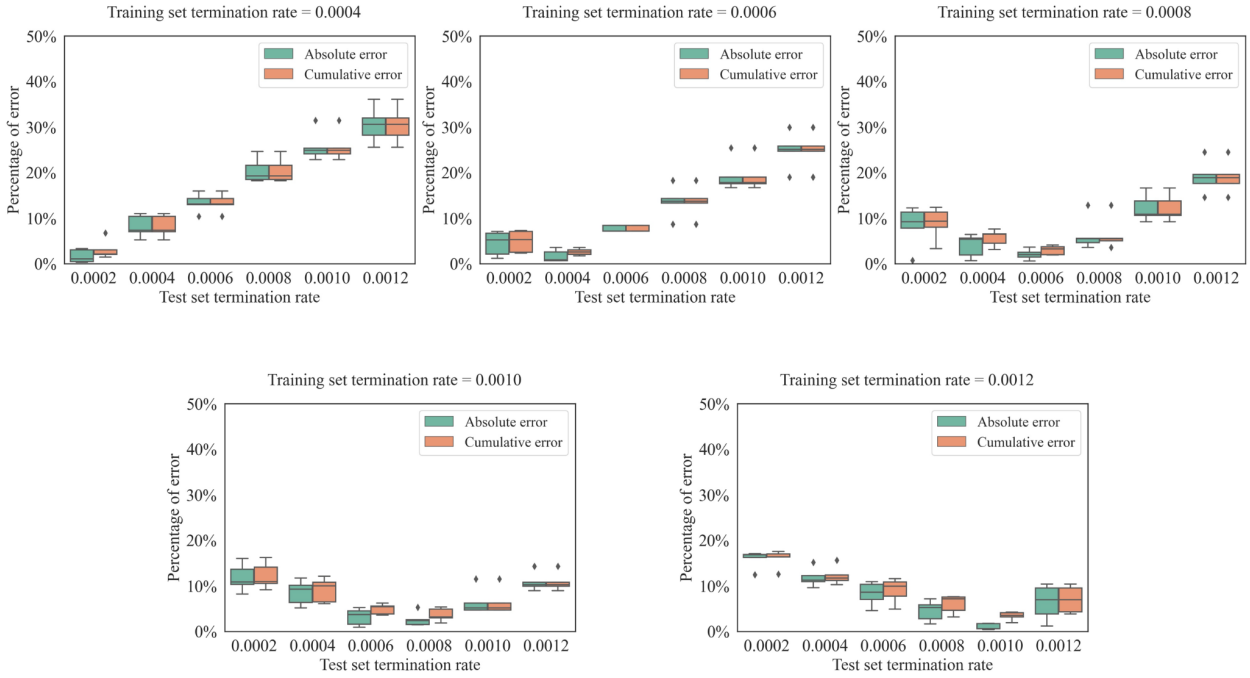

Figure S7. Performance comparison at different training set termination rates for each test set termination rate.

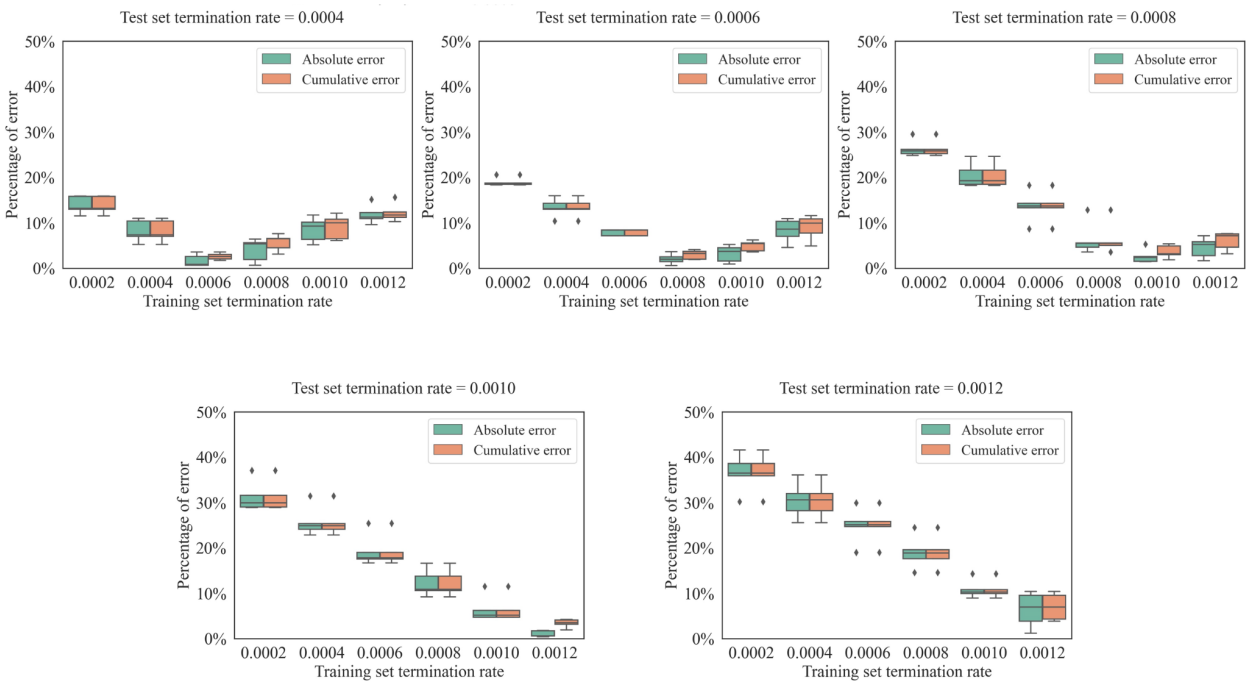

Figure S8. Performance comparison at different numbers of examples in the test set for each training set.

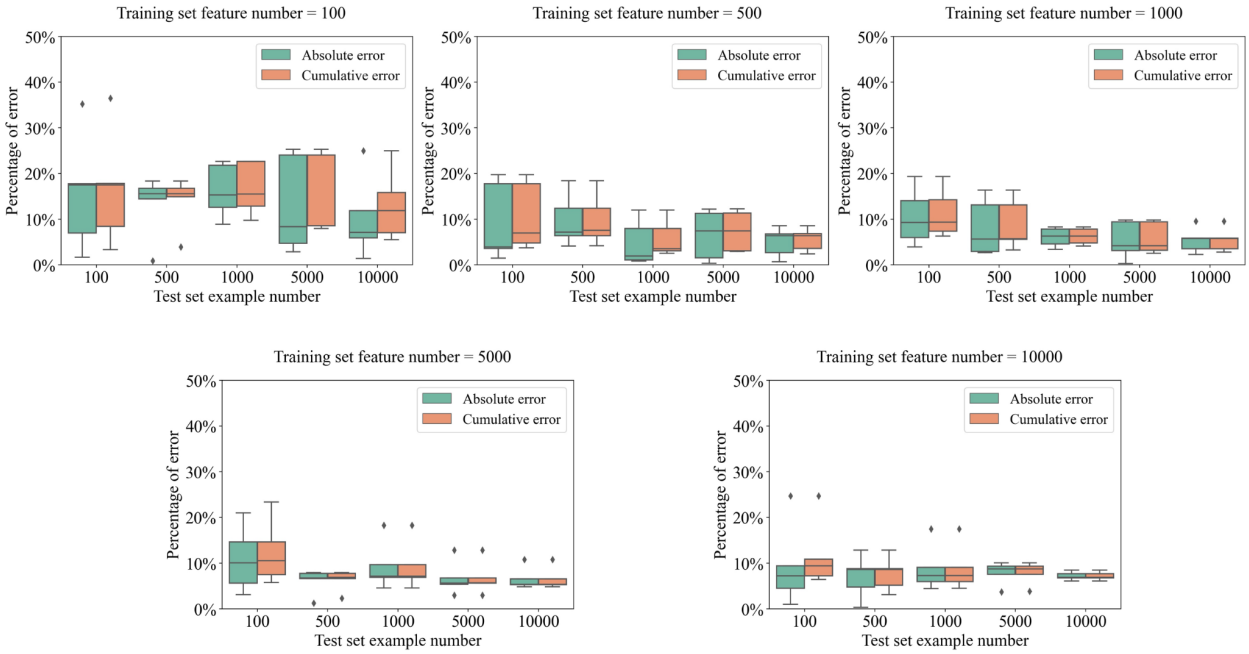

Figure S9. Performance comparison at different numbers of examples in the training set for each test set.

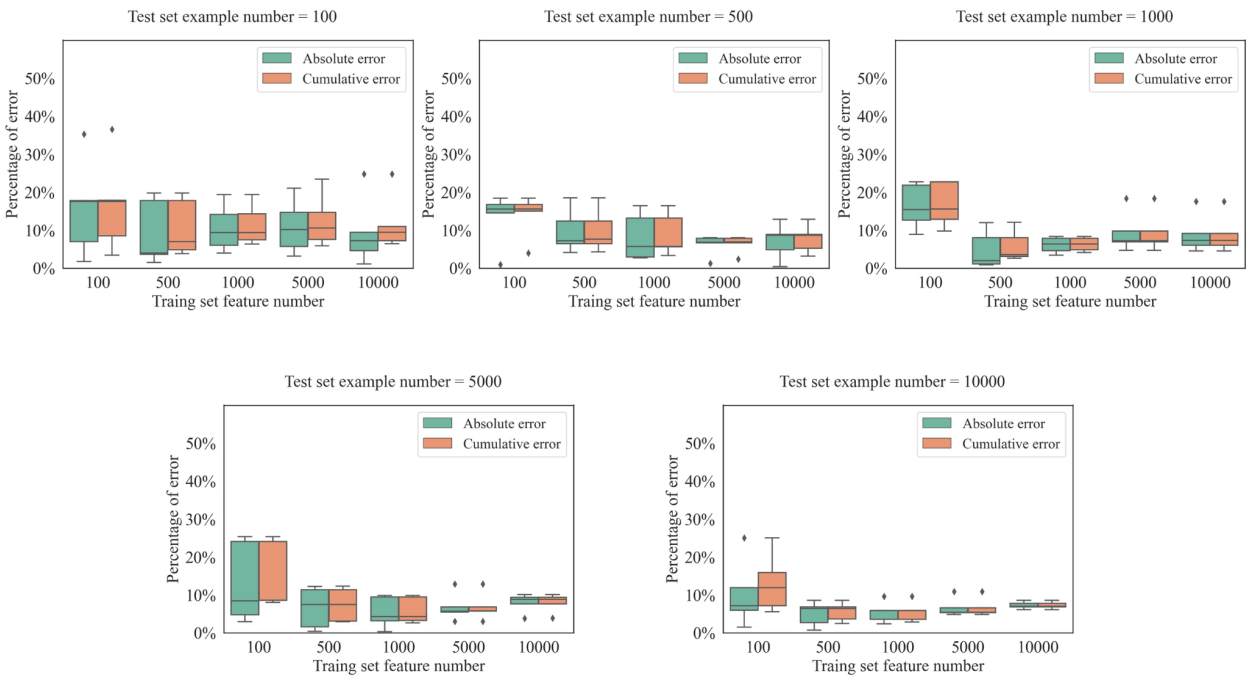

Figure S10. Performance comparison at different test noise levels for each training set noise level.

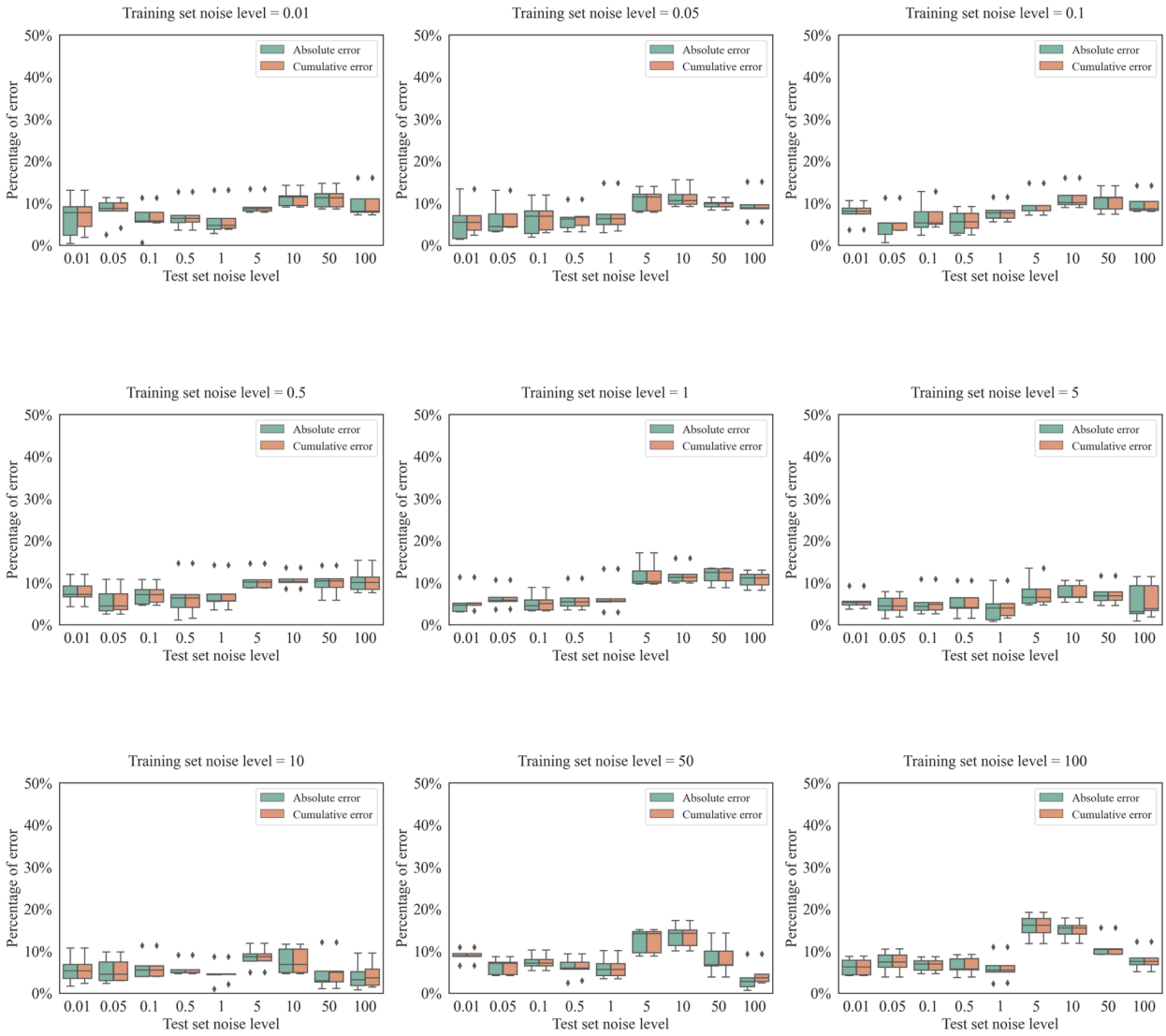

Figure S11. Performance comparison at different training set noise levels for each test set noise level.

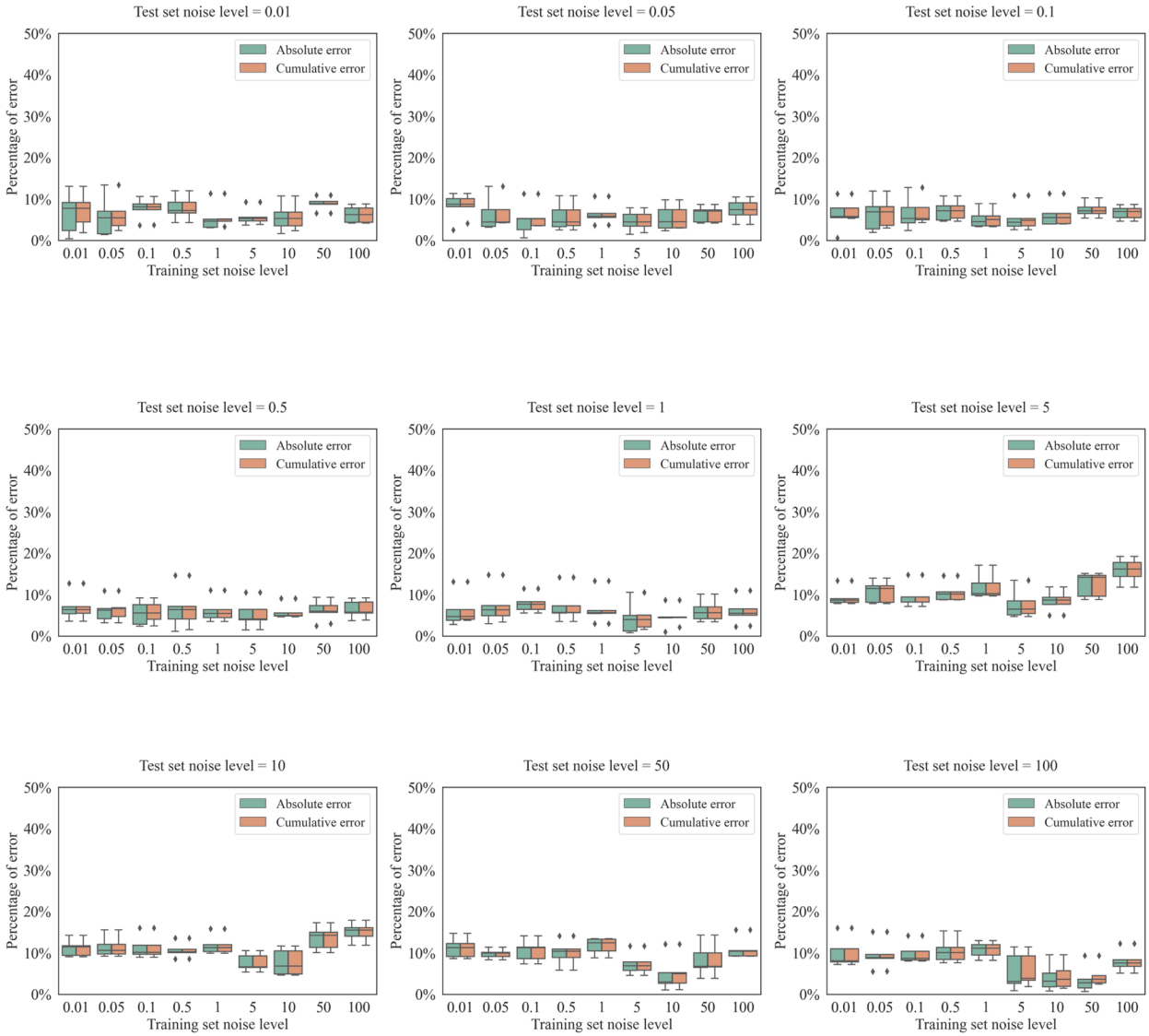

Figure S12. Performance comparison at different test set feature scale for each training set.

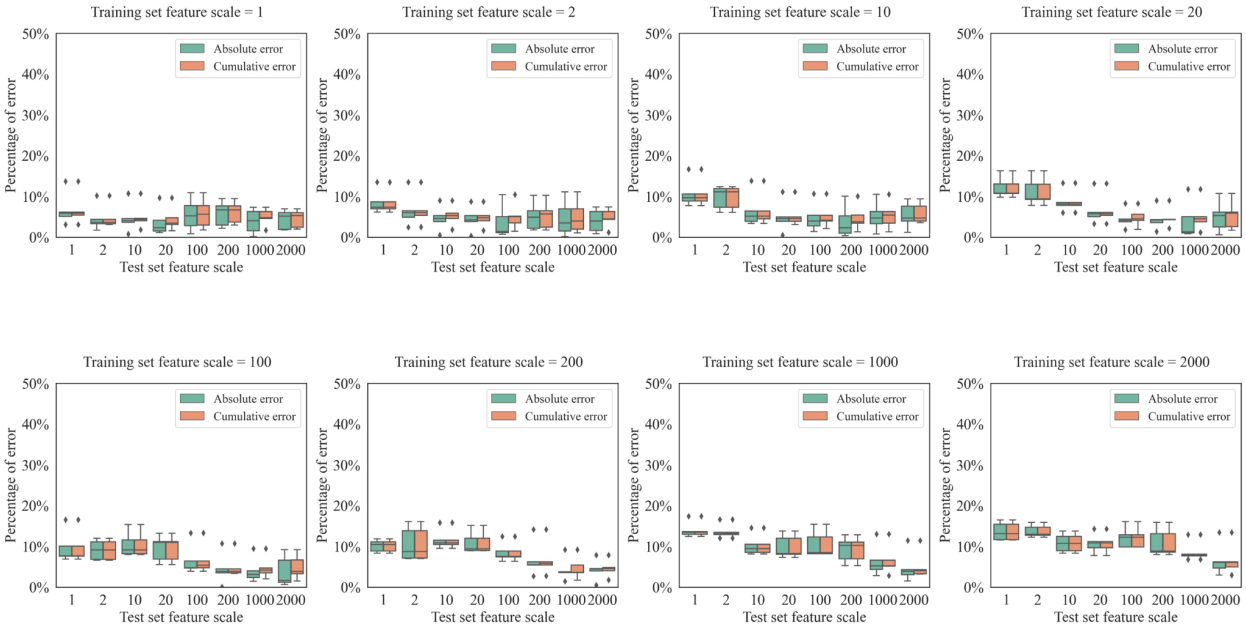

Figure S13. Performance comparison at different training set feature scale for each test set.

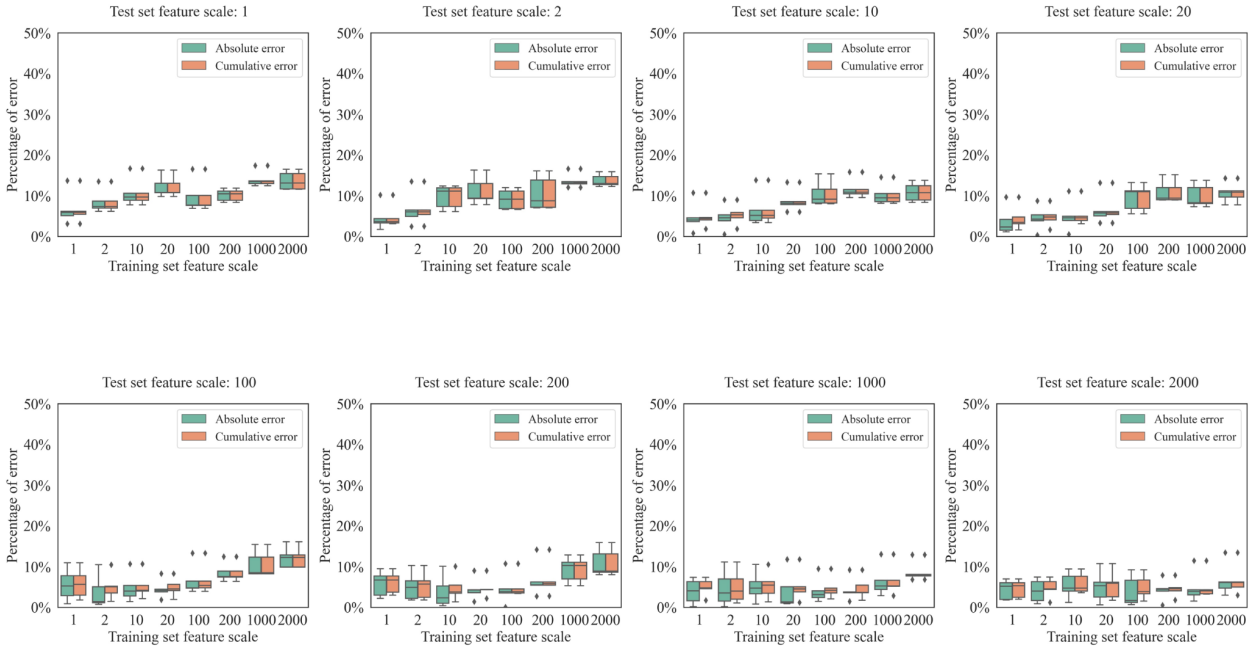

Figure S14. Flow chart of rwTTD prediction for cancers treated with pembrolizumab. a. Models were tested by cross-validation with random cutoff between the first dosing and the last contact date. b. Models were tested by cutting off 30 days after the first dosing. c. Accumulated errors are evaluated at different month cutoffs. d. rwTTD is defined through four different situations. e. Feature extraction and normalization for each data table.

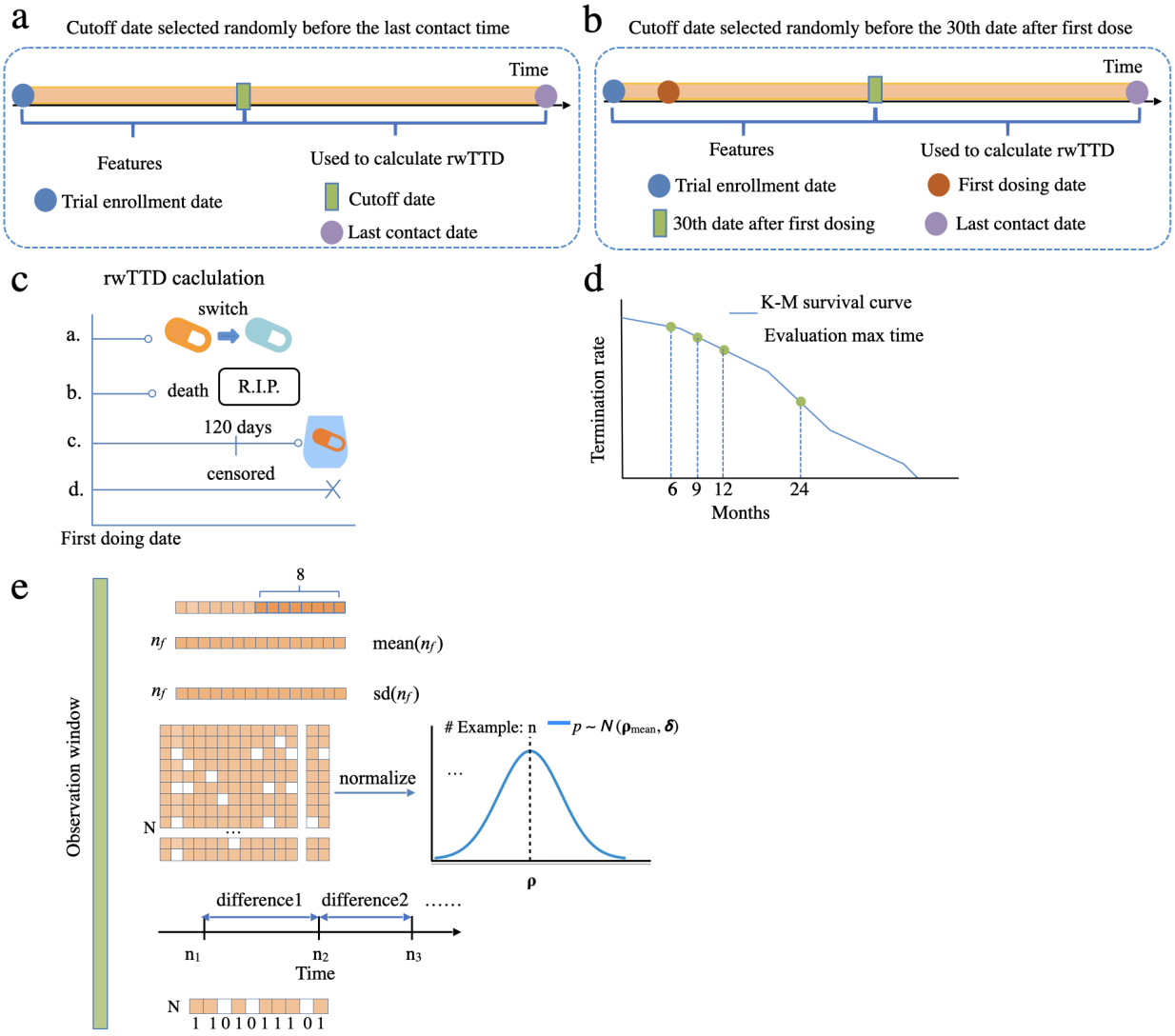

Figure S15. Performance of five fold cross validation for random cutoff, lung cancer patients treated with pembrolizumab.

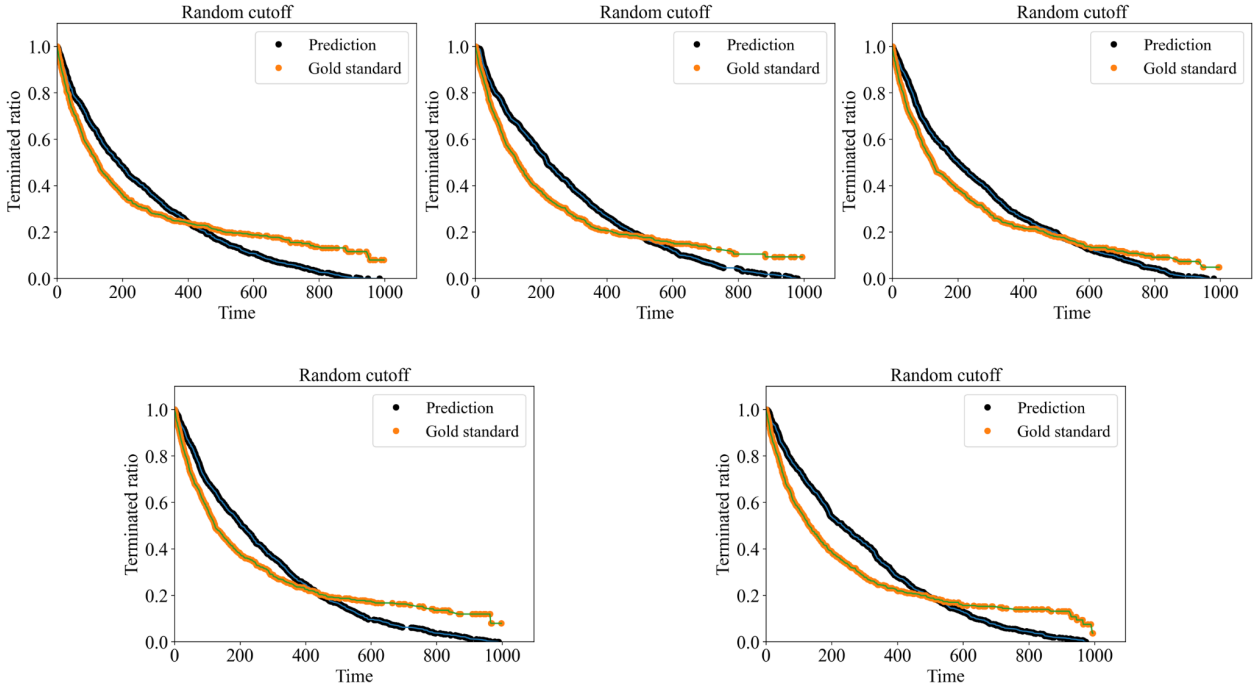

Figure S16. Performance of five fold cross validation for 30 day cutoff, lung cancer patients treated with pembrolizumab.

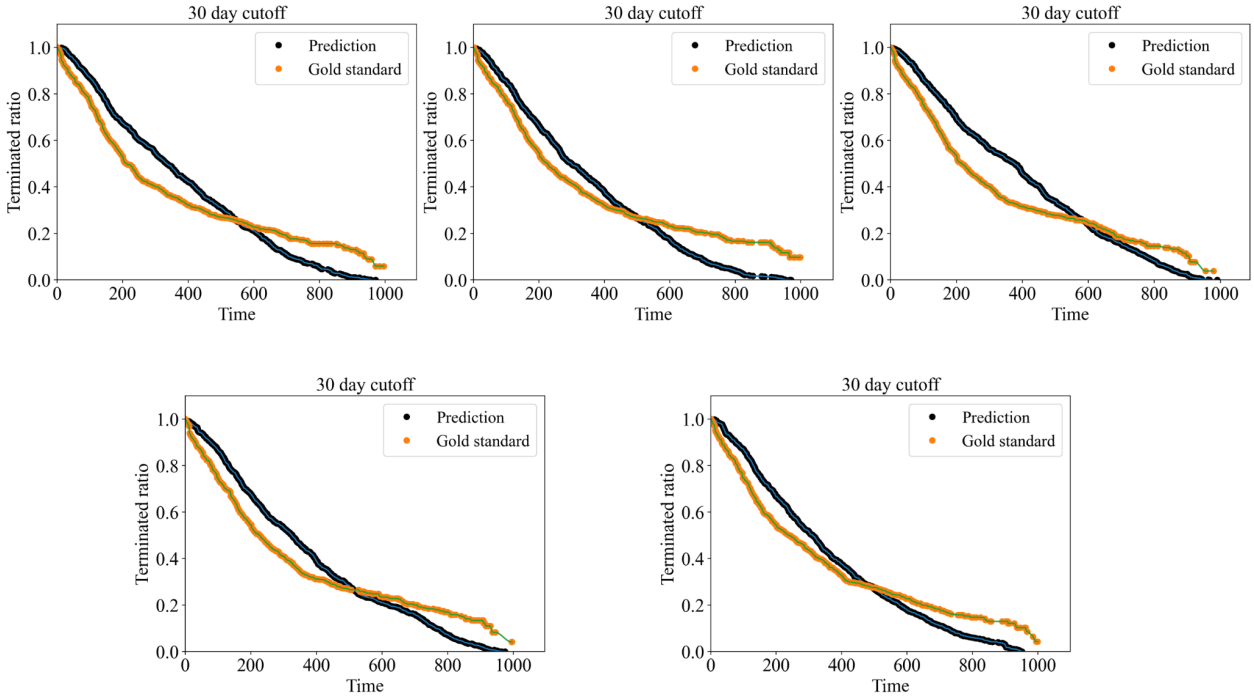

Figure S17. Performance of cross-disease prediction, trained with lung cancer patients treated with pembrolizumab and tested on head and neck patients.

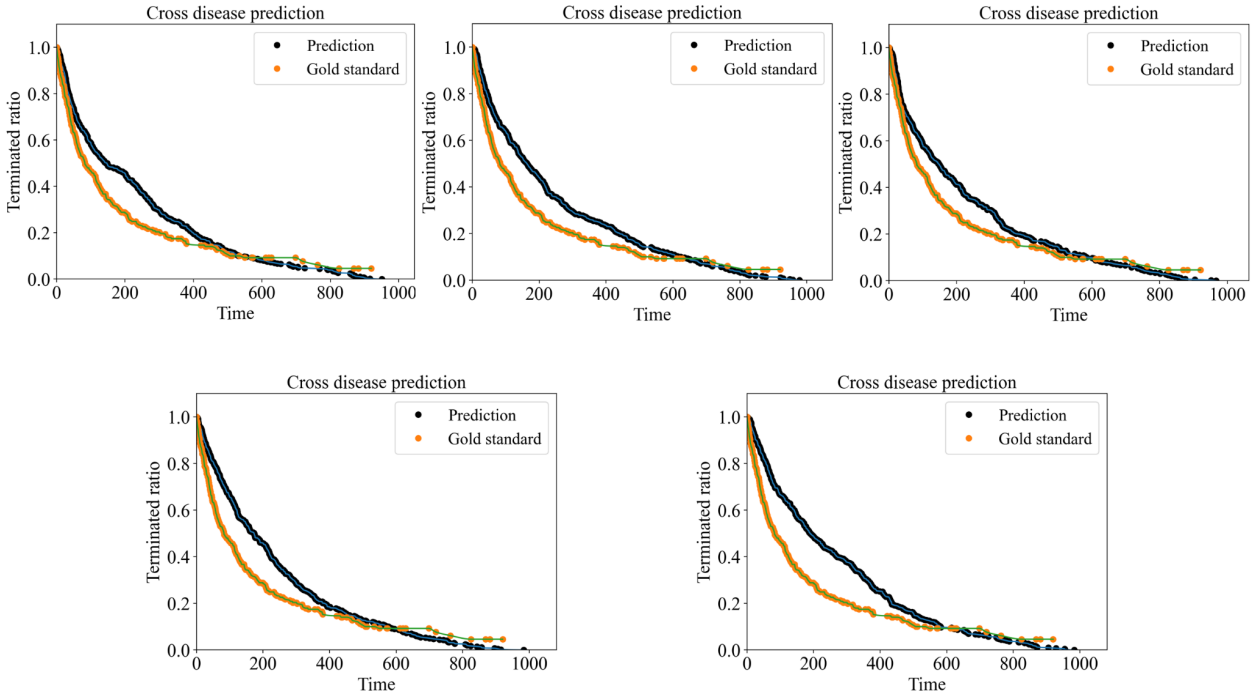

Table S1. Demographic profiles of training and test cohorts.

		NSCLC	Head and neck cancer
Sex	F	2,271	99
	M	2,512	323
Race	Asian	75	3
	Black or African American	384	23
	Hispanic or Latino	3	–
	White	3,388	273
	Other Race	443	92
Ethnicity	Hispanic or Latino	134	16
Birth year	min	1936	1936
	mean	1948	1951
	median	1947	1951
	max	1989	2000

Table S2. Performance summary of different models across different termination rates.

Base learner	Termination Rate	Median Error Rate
Linear	0.0012	0.1248229157
Linear	0.0002	0.1206743285
Linear	0.0004	0.1356050511
Linear	0.0008	0.1338715308
Linear	0.0006	0.136311454
Linear	0.001	0.1358148427
DL	0.0012	0.1139465147
DL	0.0002	0.1220473496
DL	0.0004	0.1353276894
DL	0.0008	0.1268277211
DL	0.0006	0.1816106844
DL	0.001	0.1302750479
SVM	0.0012	0.1201241845
SVM	0.0002	0.1209352065
SVM	0.0004	0.126091972
SVM	0.0008	0.1258197652
SVM	0.0006	0.1302979533
SVM	0.001	0.132713221
ExtraTreeRegressor	0.0012	0.1262252106
ExtraTreeRegressor	0.0002	0.1188941781
ExtraTreeRegressor	0.0004	0.1658188705
ExtraTreeRegressor	0.0008	0.189690489
ExtraTreeRegressor	0.0006	0.142007922
ExtraTreeRegressor	0.001	0.103503574

Table S3. Performance summary of different models across different numbers of examples.

Base learner	Number of examples	Median Error Rate
Linear	100	0.2292179095
Linear	5000	0.09764749906
Linear	500	0.09567383963
Linear	1000	0.1396747775
Linear	10000	0.07946776141
DL	100	0.2277279437
DL	5000	0.08397675668
DL	500	0.04355277992
DL	1000	0.1015391315
DL	10000	0.08257765273
SVM	100	0.2022373928
SVM	5000	0.08948145398
SVM	500	0.07706892652
SVM	1000	0.1349773432
SVM	10000	0.06279371145
ExtraTreeRegressor	100	0.1984001184
ExtraTreeRegressor	5000	0.1042650507
ExtraTreeRegressor	500	0.09840050794
ExtraTreeRegressor	1000	0.127066872
ExtraTreeRegressor	10000	0.06810466375