

# Supplement

eTable 1. Variables in use for all diagnosis years and disease sites included in our analysis.

Category	Variable name	Missing in at least one patient record	Variable of interest
Demographics	AGE		
Demographics	CDCC_TOTAL_BEST		
Demographics	CROWFLY	X	X
Demographics	FACILITY_LOCATION_CD	X	X
Demographics	FACILITY_TYPE_CD	X	X
Demographics	INSURANCE_STATUS	X	X
Demographics	MED_INC_QUAR_00	X	X
Demographics	MED_INC_QUAR_12	X	X
Demographics	MED_INC_QUAR_16	X	X
Demographics	MEDICAID_EXP_N_CODE		
Demographics	NO_HSD_QUAR_00	X	X
Demographics	NO_HSD_QUAR_12	X	X
Demographics	NO_HSD_QUAR_16	X	X
Demographics	PUF_CASE_ID		
Demographics	PUF_FACILITY_ID		
Demographics	PUF_MULT_SOURCE		
Demographics	RACE	X	X
Demographics	REFERENCE_DATE_FLAG		
Demographics	SEX		
Demographics	SPANISH_HISPANIC_ORIGIN	X	X
Demographics	UR_CD_03	X	X
Demographics	UR_CD_13	X	X
Cancer Identification	BEHAVIOR		
Cancer Identification	CLASS_OF_CASE		
Cancer Identification	DIAGNOSTIC_CONFIRMATION	X	X
Cancer Identification	GRADE	X	X
Cancer Identification	HISTOLOGY		
Cancer Identification	LATERALITY	X	X
Cancer Identification	PRIMARY_SITE		
Cancer Identification	REGIONAL_NODES_EXAMINED	X	X
Cancer Identification	REGIONAL_NODES_POSITIVE	X	X
Cancer Identification	SEQUENCE_NUMBER	X	X
Cancer Identification	YEAR_OF_DIAGNOSIS		
Stage	ANALYTIC_STAGE_GROUP	X	X
Stage	CS_EXTENSION	X	X
Stage	CS_METS_AT_DX	X	X
Stage	CS_METS_EVAL	X	X
Stage	CS_TUMOR_SIZEEXT_EVAL	X	X
Stage	CS_VERSION_LATEST	X	X
Stage	DX_STAGING_PROC_DAYS	X	
Stage	RX_SUMM_DXSTG_PROC	X	X
Stage	TNM_CLIN_M	X	X
Stage	TNM_CLIN_N	X	X
Stage	TNM_CLIN_STAGE_GROUP	X	X
Stage	TNM_CLIN_T	X	X
Stage	TNM_EDITION_NUMBER	X	X
Stage	TNM_PATH_M	X	
Stage	TNM_PATH_N	X	
Stage	TNM_PATH_STAGE_GROUP	X	
Stage	TNM_PATH_T	X	
Stage	TUMOR_SIZE	X	X*
Treatment	DX_CHEMO_STARTED_DAYS	X	
Treatment	DX_DEFSURG_STARTED_DAYS	X	
Treatment	DX_HORMONE_STARTED_DAYS	X	
Treatment	DX_IMMUNO_STARTED_DAYS	X	
Treatment	DX_OTHER_STARTED_DAYS	X	
Treatment	DX_RAD_STARTED_DAYS	X	
Treatment	DX_RX_STARTED_DAYS	X	
Treatment	DX_SURG_STARTED_DAYS	X	
Treatment	DX_SYSTEMIC_STARTED_DAYS	X	
Treatment	PALLIATIVE_CARE	X	X
Treatment	PALLIATIVE_CARE_HOSP	X	X
Treatment	RAD_BOOST_DOSE_CGY	X	X
Treatment	RAD_BOOST_RX_MODALITY	X	X
Treatment	RAD_ELAPSED_RX_DAYS	X	X
Treatment	RAD_LOCATION_OF_RX	X	X
Treatment	RAD_NUM_TREAT_VOL	X	X
Treatment	RAD_REGIONAL_DOSE_CGY	X	X
Treatment	RAD_REGIONAL_RX_MODALITY	X	X
Treatment	RAD_TREAT_VOL	X	X
Treatment	READM_HOSP_30_DAYS	X	X
Treatment	REASON_FOR_NO_RADIATION	X	X
Treatment	REASON_FOR_NO_SURGERY	X	X
Treatment	RX_HOSP_CHEMO	X	X
Treatment	RX_HOSP_DXSTG_PROC	X	X
Treatment	RX_HOSP_HORMONE	X	X
Treatment	RX_HOSP_IMMUNOTHERAPY	X	X
Treatment	RX_HOSP_OTHER	X	X
Treatment	RX_HOSP_SURG_PRIM_SITE	X	X
Treatment	RX_SUMM_CHEMO	X	X
Treatment	RX_SUMM_HORMONE	X	X
Treatment	RX_SUMM_IMMUNOTHERAPY	X	X
Treatment	RX_SUMM_OTHER	X	X
Treatment	RX_SUMM_RADIATION	X	X
Treatment	RX_SUMM_SCOPE_REG_LN_SUR	X	X
Treatment	RX_SUMM_SURG_OTH_REGDIS	X	X
Treatment	RX_SUMM_SURG_PRIM_SITE	X	X
Treatment	RX_SUMM_SURGICAL_MARGINS	X	
Treatment	RX_SUMM_SURGRAD_SEQ	X	X
Treatment	RX_SUMM_SYSTEMIC_SUR_SEQ	X	X
Treatment	RX_SUMM_TRNSPLNT_ENDO	X	X
Treatment	SURG_DISCHARGE_DAYS	X	
Outcomes	DX_LASTCONTACT_DEATH_MONTHS	X	
Outcomes	PUF_30_DAY_MORT_CD	X	
Outcomes	PUF_90_DAY_MORT_CD	X	
Outcomes	PUF_VITAL_STATUS		

\*Tumor size was not used for prostate cancer given it is not routinely reported in clinical practice

eTable 2. Tabulation of patient records used to comparison of patient, tumor, and treatment characteristics. Patient records not used that have complete data are due to recoding of the variable for comparison.

A. Non-small cell lung cancer

	Records used		Records not used	
	Complete data	Missing data	Complete data	Missing data
Age at Diagnosis	347,454	851,295	0	0
Sex	347,454	851,295	0	0
Race	347,454	842,127	0	9,168
Hispanic Ethnicity	347,454	784,015	0	67,280
Charlson-Deyo Score	347,454	851,295	0	0
Insurance	347,454	827,936	0	23,359
Facility Type	347,454	843,657	0	7,638
Year of Diagnosis	347,454	851,295	0	0
Overall Stage	347,395	776,500	59	74,795
Tumor Size	346,962	626,110	492	225,185
Lymph Nodes Involved	336,910	692,475	10,544	158,820
Distant Metastasis	347,454	815,377	0	35,918
Surgery (Primary Site)	347,454	847,710	0	3,585
Radiation	347,454	840,924	0	10,371
Chemotherapy	347,454	811,051	0	40,244

B. Breast cancer

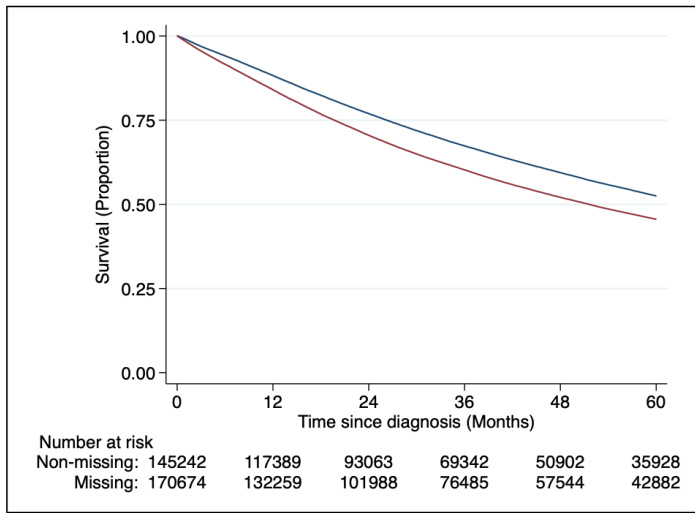
	Records used		Records not used	
	Complete data	Missing data	Complete data	Missing data
Age at Diagnosis	959,679	1,161,096	0	0
Sex	959,679	1,161,096	0	0
Race	959,679	1,138,156	0	22,940
Hispanic Ethnicity	959,679	1,049,841	0	111,255
Charlson-Deyo Score	959,679	1,161,096	0	0
Insurance	959,679	1,122,489	0	38,607
Facility Type	959,679	1,067,375	0	93,721
Year of Diagnosis	959,679	1,161,096	0	0
Overall Stage	959,361	1,089,661	318	71,435
Tumor Size	956,593	954,416	3,086	206,680
Lymph Nodes Involved	945,511	991,196	14,168	169,900
Distant Metastasis	959,361	1,083,580	318	77,516
Surgery (Primary Site)	959,679	1,156,603	0	4,493
Radiation	959,679	1,139,214	0	21,882
Chemotherapy	959,679	1,093,054	0	68,042
Hormonal Therapy	959,679	1,066,418	0	94,678

C. Prostate cancer

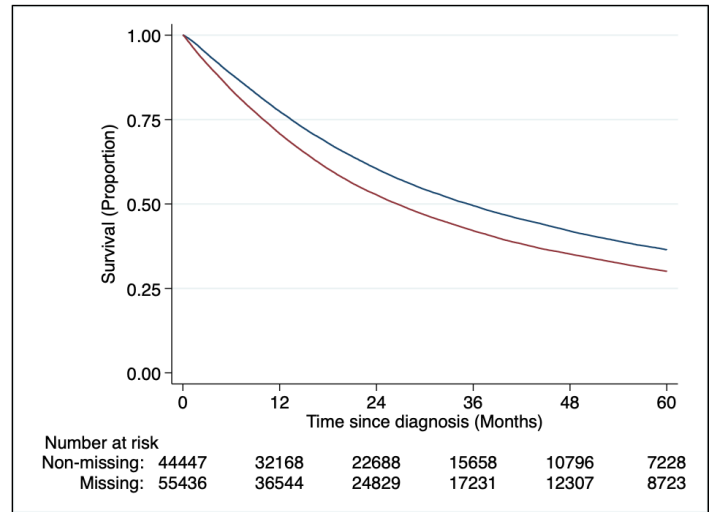
	Records used		Records not used	
	Complete data	Missing data	Complete data	Missing data
Age at Diagnosis	698,468	460,167	0	0
Sex	698,468	460,167	0	0
Race	698,468	441,710	0	18,457
Hispanic Ethnicity	698,468	386,668	0	73,499
Charlson-Deyo Score	698,468	460,167	0	0
Insurance	698,468	436,545	0	23,622
Facility Type	698,468	459,296	0	871
Year of Diagnosis	698,468	456,869	0	3,298
Overall Stage	698,468	460,167	0	0
Tumor Size	698,430	403,550	38	56,617
Lymph Nodes Involved	668,940	356,173	29,528	103,994
Distant Metastasis	698,429	417,866	39	42,301
Surgery (Primary Site)	698,468	456,891	0	3,276
Radiation	698,468	449,371	0	10,796
Chemotherapy	698,468	423,743	0	36,424
Hormonal Therapy	698,468	419,513	0	40,654

eFigure 1. Non-small cell lung cancer patient survival by whether data is missing in key variables and by cancer stage.

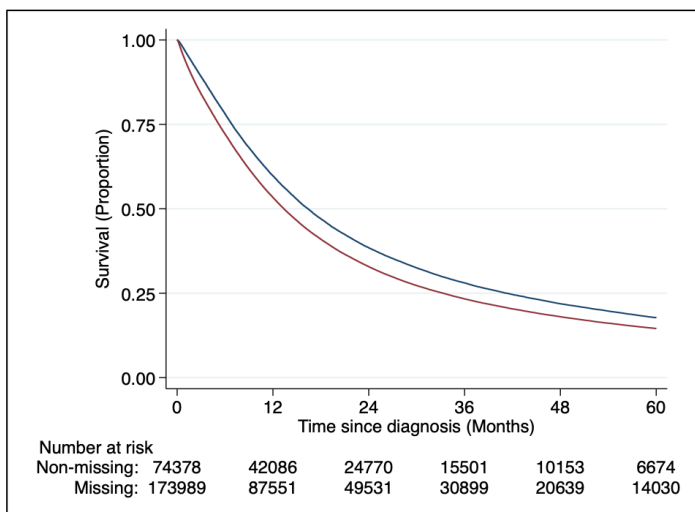
### A. Stage I\*



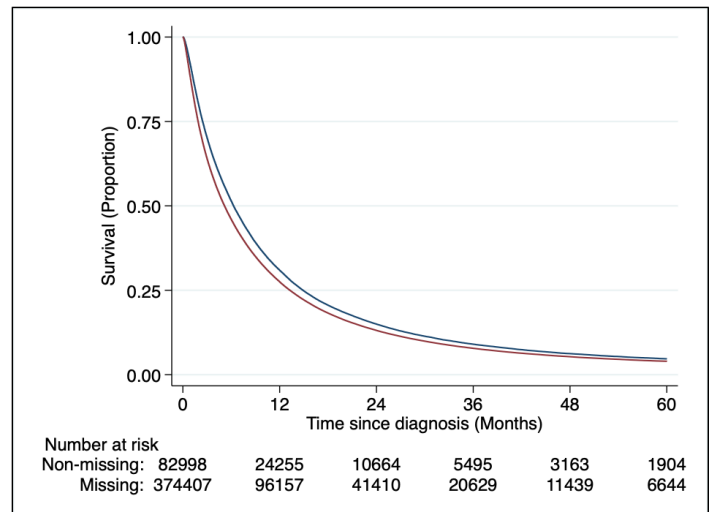
### B. Stage II



### C. Stage III



### D. Stage IV

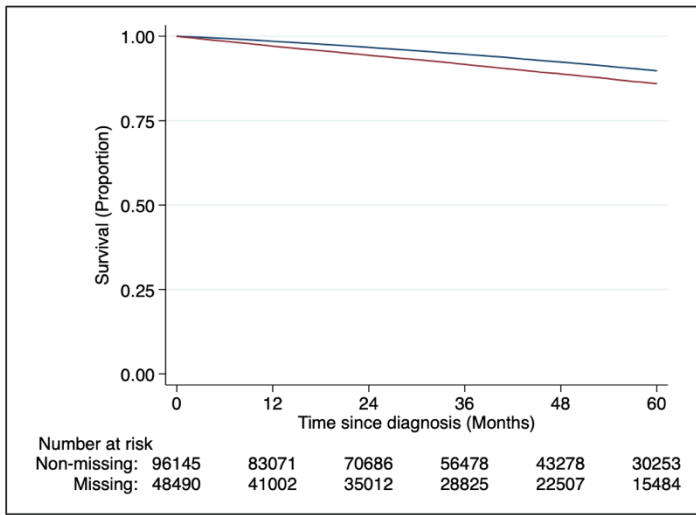


\*Analytic group stages 0 and I were combined for stage I subgroup analysis for NSCLC

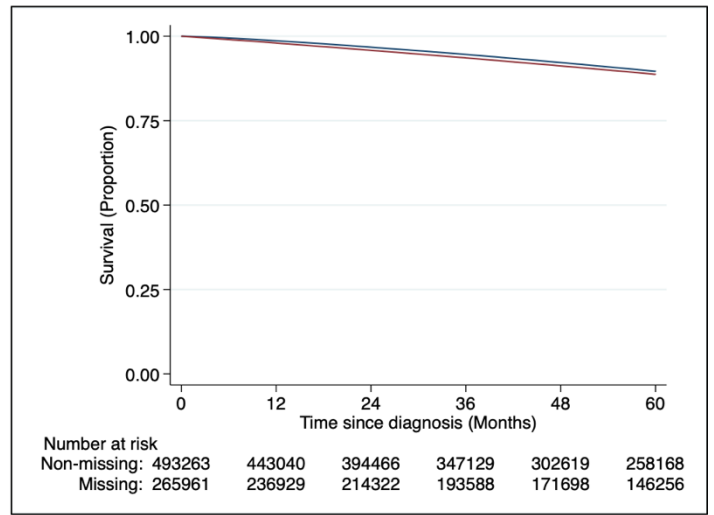
— Non-missing data  
 — Missing data

eFigure 2. Breast cancer patient survival by whether data is missing in key variables and by cancer stage.

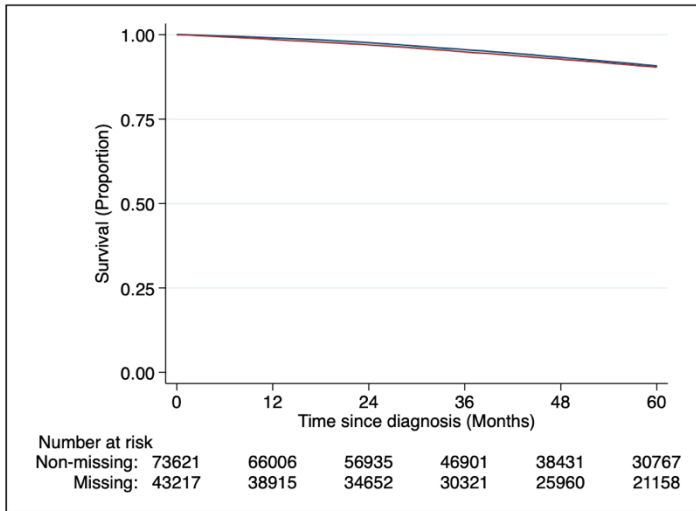
### A. Stage I



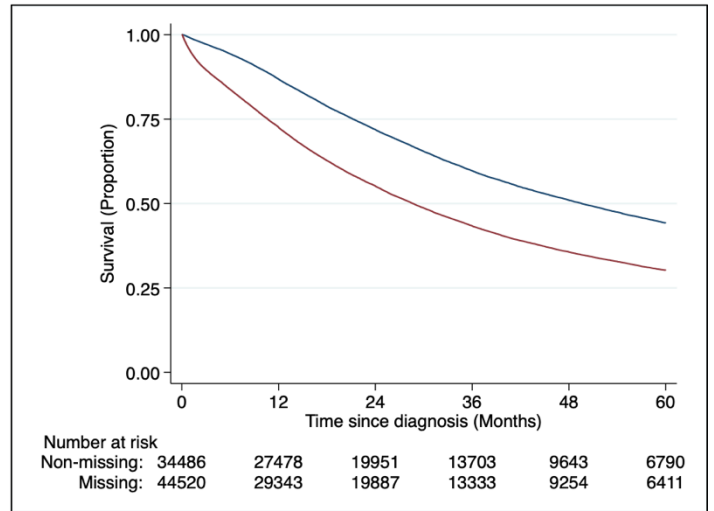
### B. Stage II



### C. Stage III



### D. Stage IV

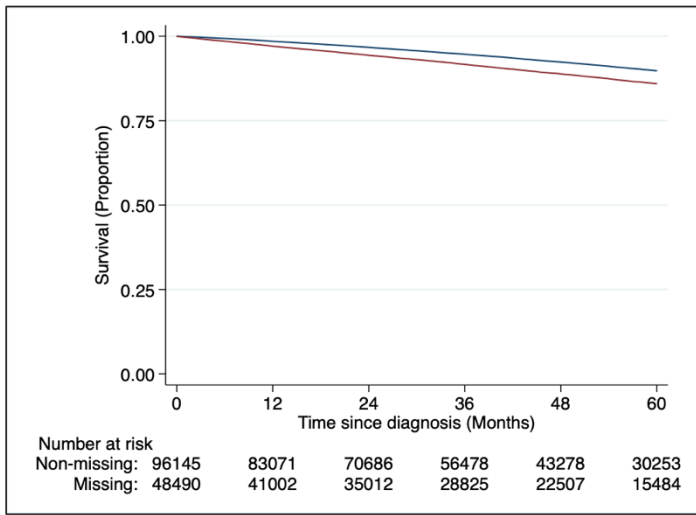


— Non-missing data

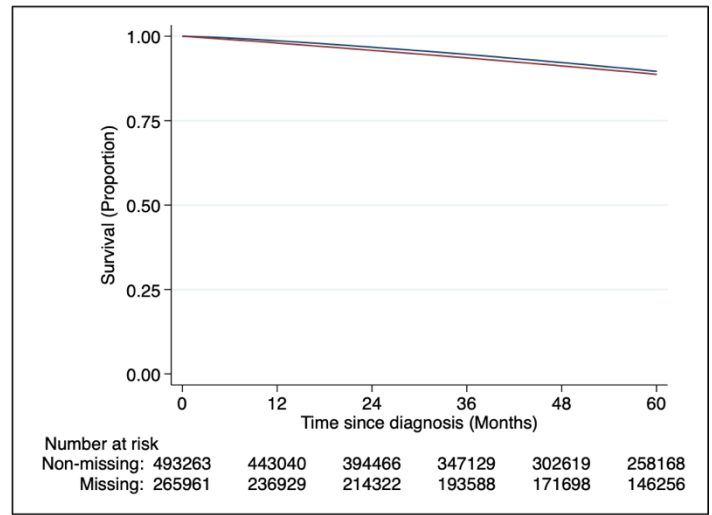
— Missing data

eFigure 3. Prostate cancer patient survival by whether data is missing in key variables and by cancer stage.

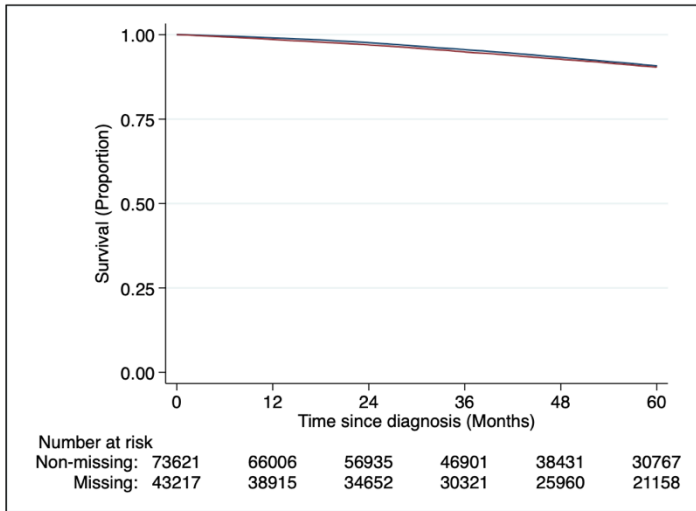
### A. Stage I



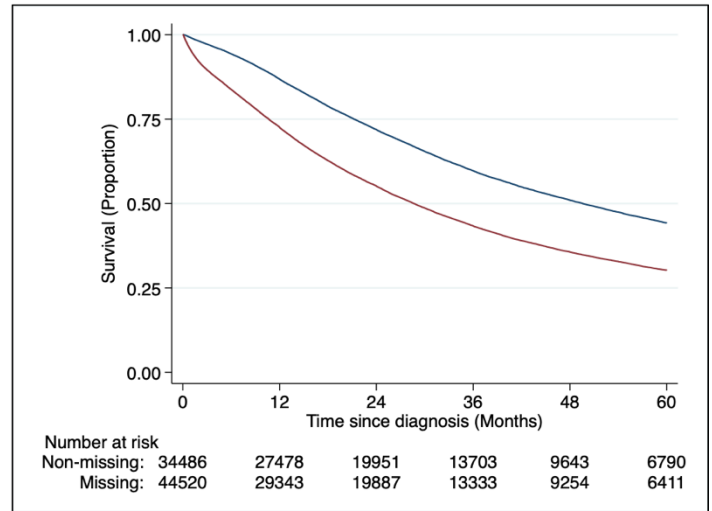
### B. Stage II



### C. Stage III



### D. Stage IV

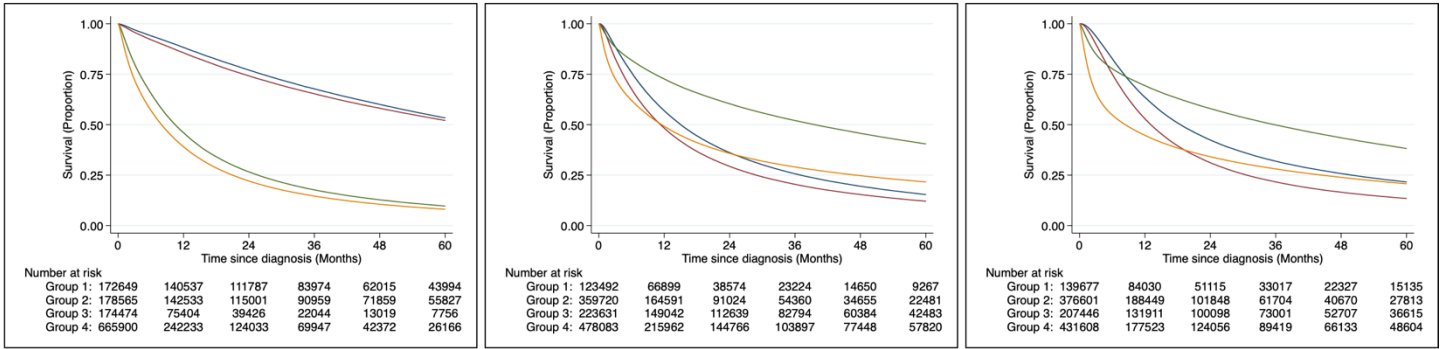


— Non-missing data

— Missing data

eFigure 4. Cancer patient survival by whether data is missing in key variables and by treatment received.

**A. Non-small cell lung cancer**

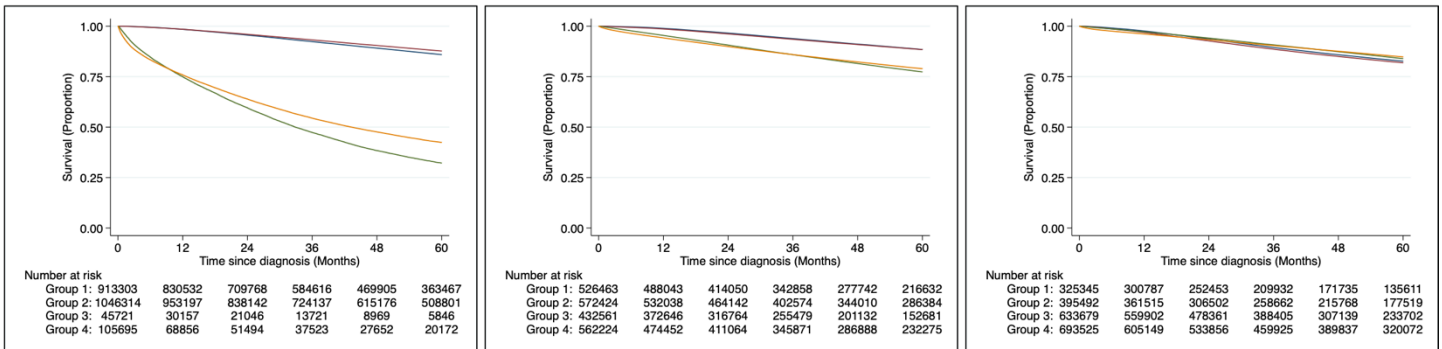


Surgery

Radiation

Chemotherapy

**B. Breast cancer**

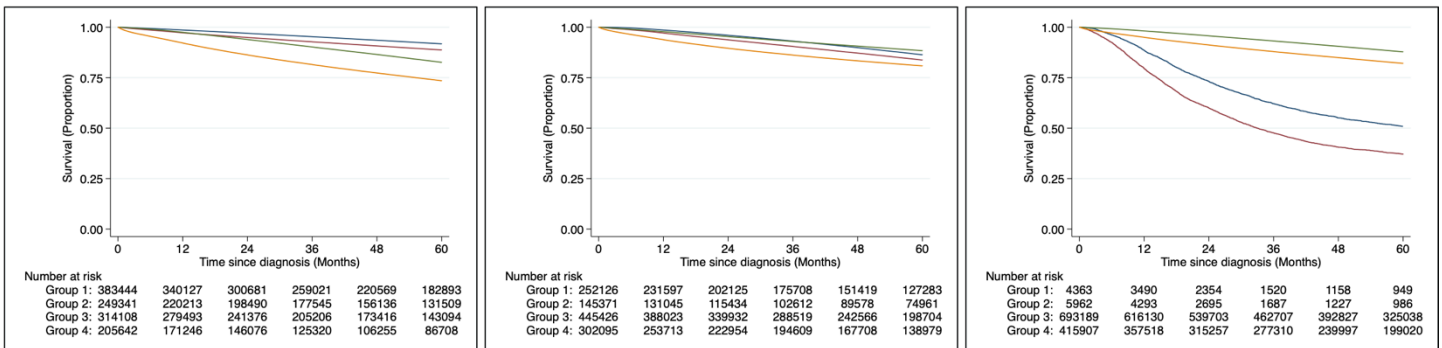


Surgery

Radiation

Chemotherapy

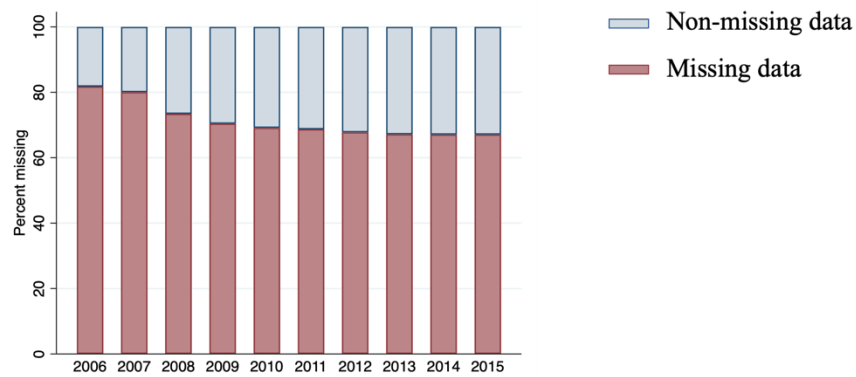
**C. Prostate cancer**



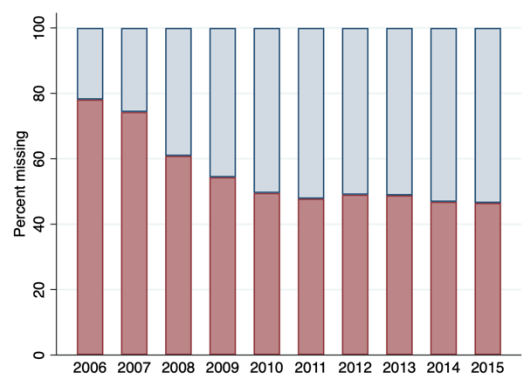
- Group 1: Received treatment, non-missing data
- Group 2: Received treatment, missing data
- Group 3: No treatment, non-missing data
- Group 4: No treatment, missing data

eFigure 5. Distribution of missingness by year of diagnosis.

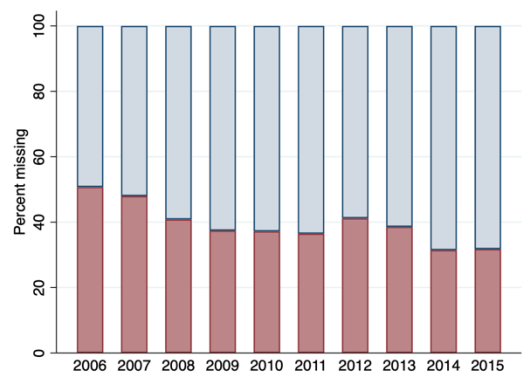
A. Non-small cell lung cancer



B. Breast cancer

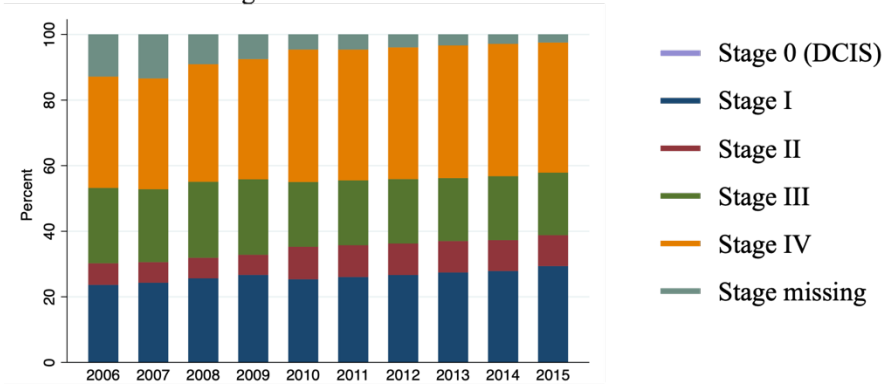


C. Prostate cancer

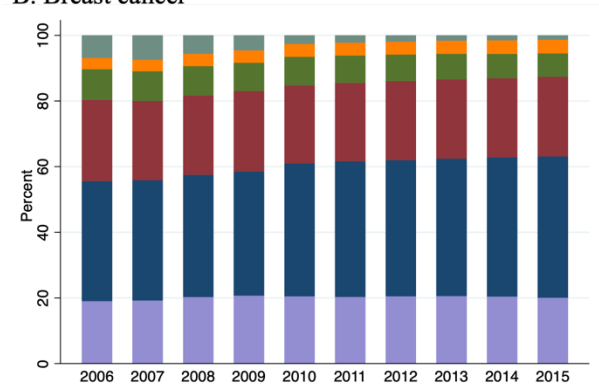


eFigure 6. Distribution of cancer stage by year of diagnosis.

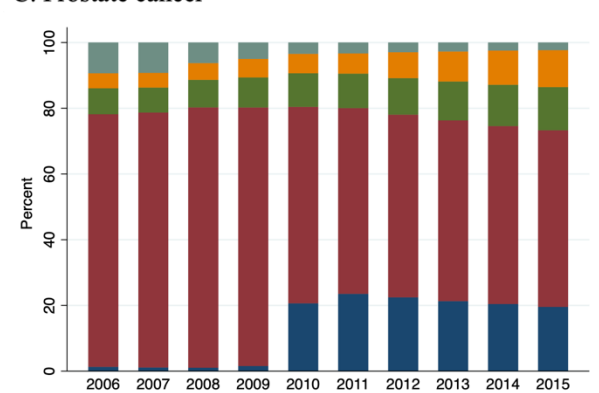
A. Non-small cell lung cancer



B. Breast cancer



C. Prostate cancer



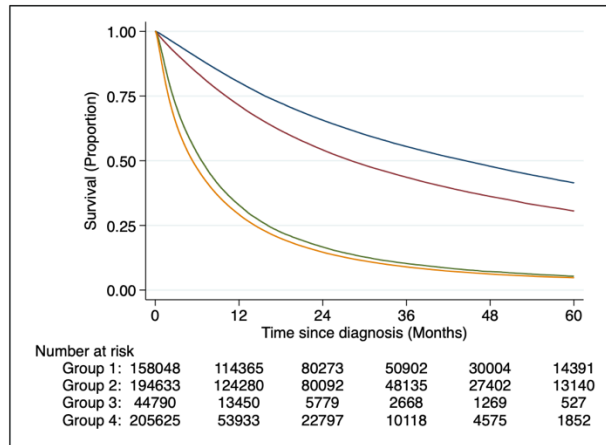
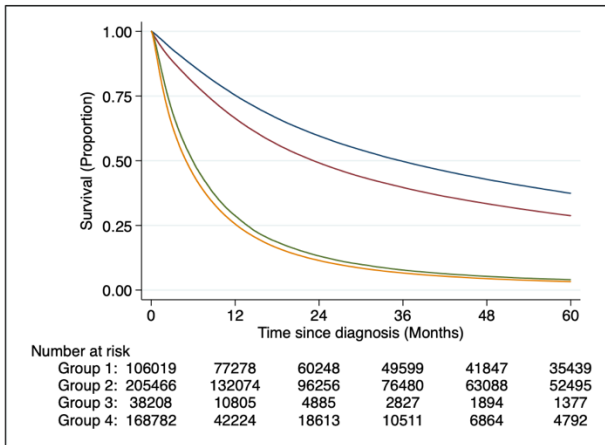


eFigure 7. Cancer patient survival whether data is missing in key variables and year of diagnosis.

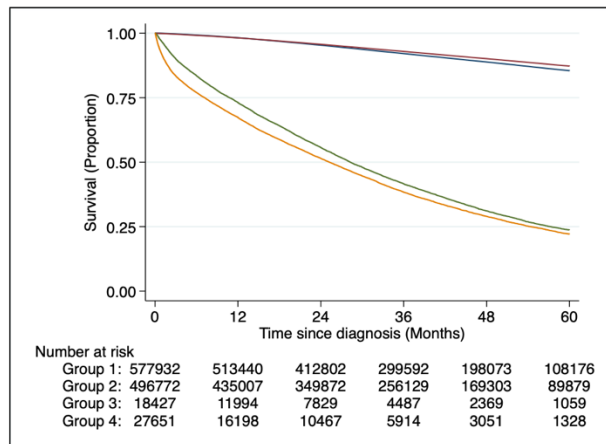
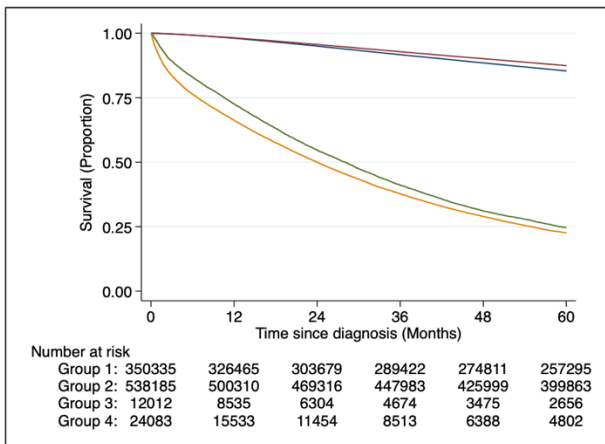
Diagnosed between 2006 to 2010

Diagnosed between 2011 to 2015

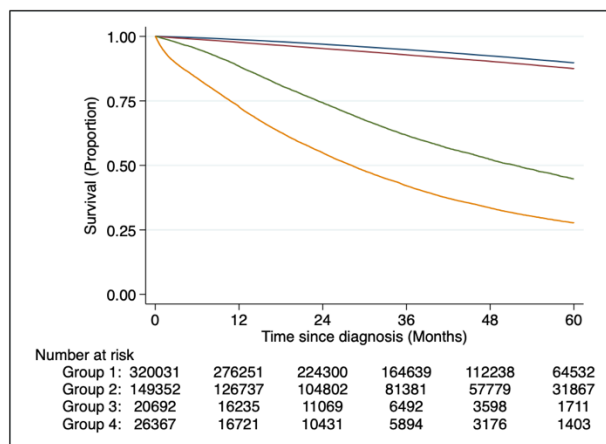
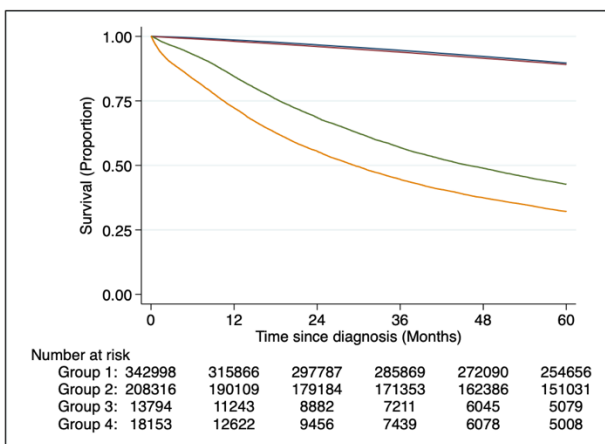
A. Non-small cell lung cancer



B. Breast cancer



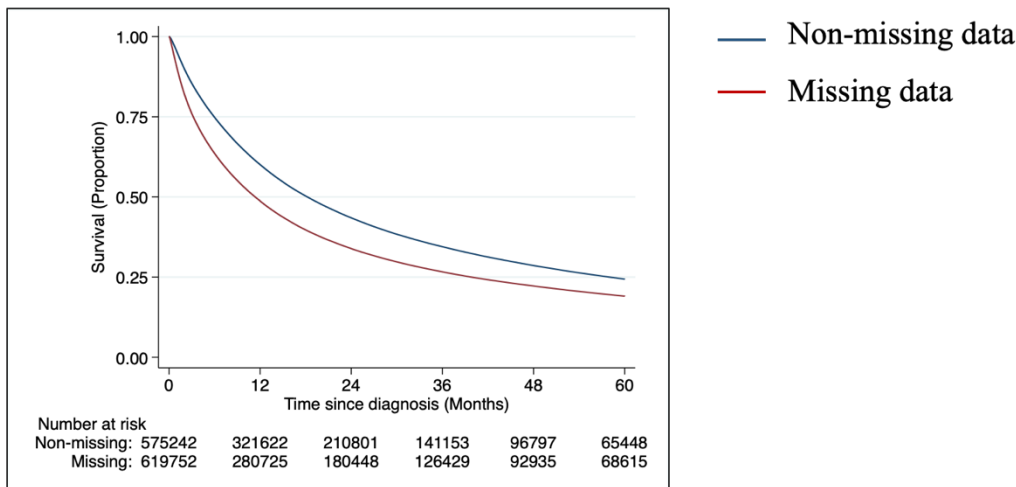
C. Prostate cancer



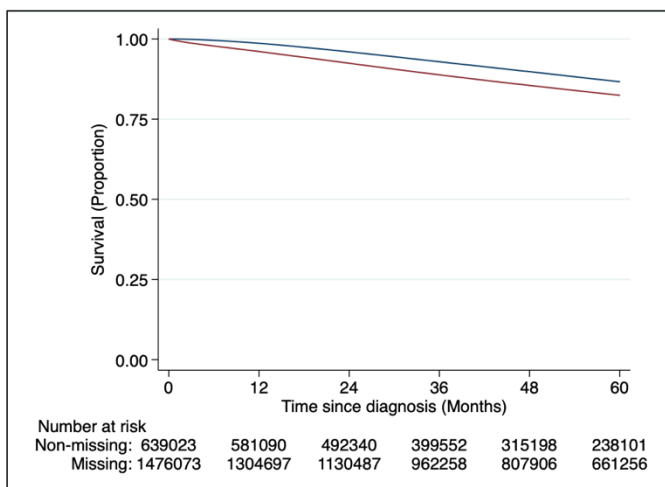
- Group 1: Non-metastatic patients, non-missing data
- Group 2: Non-metastatic patients, missing data
- Group 3: Metastatic patients, non-missing data
- Group 4: Metastatic patients, missing data

eFigure 8. Survival differences between patients with and without complete data in variables with missing data in 1-20% of patients.

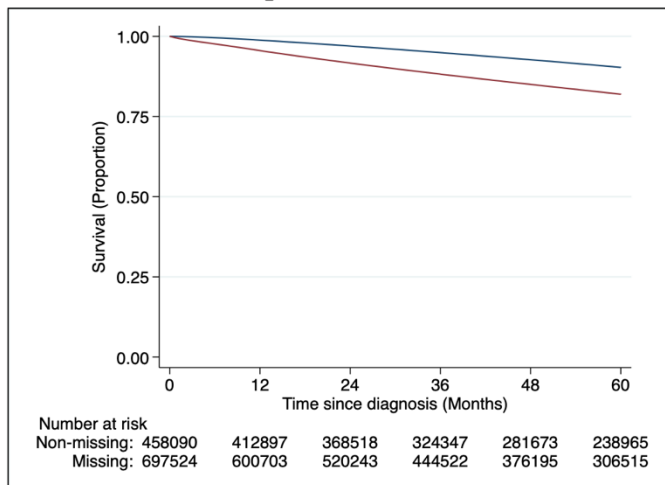
### A. Non-small cell lung cancer patients



### B. Breast cancer patients



### C. Prostate cancer patients

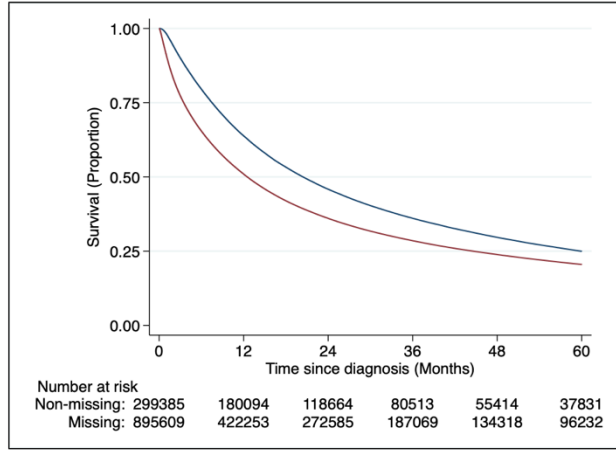
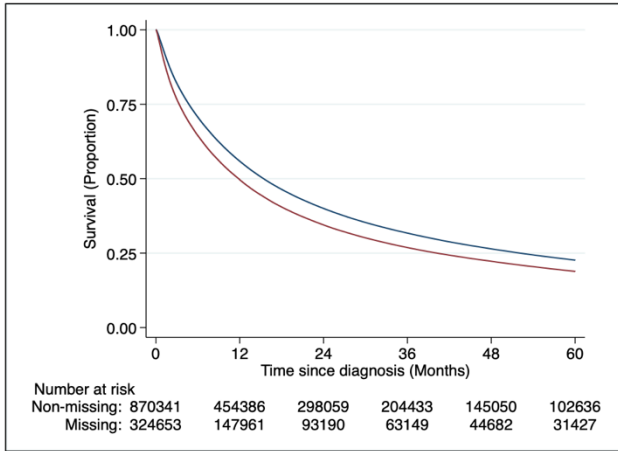


eFigure 9. Sensitivity analysis varying percentage of missing for variables of interest.

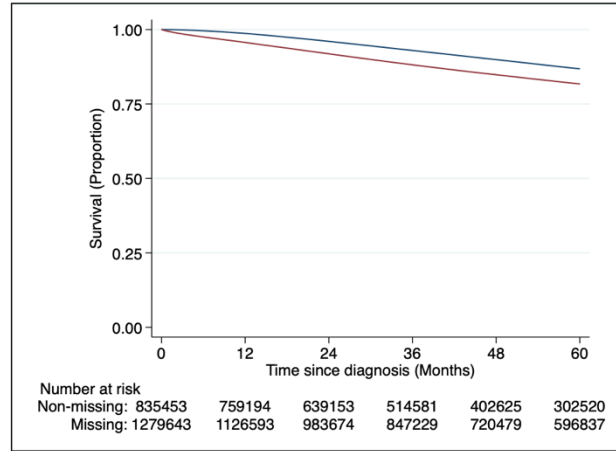
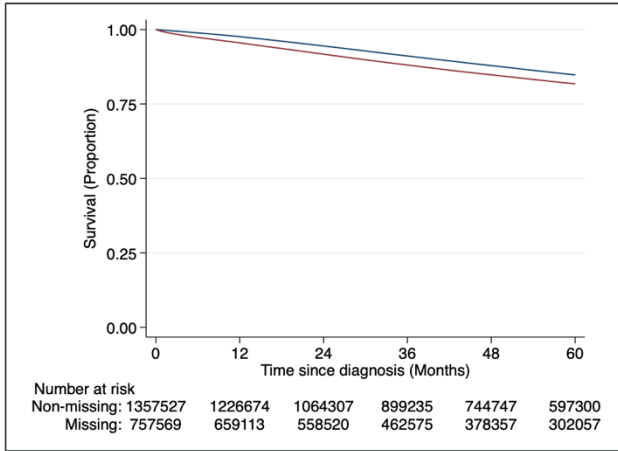
1-5% missing

>5-30% missing

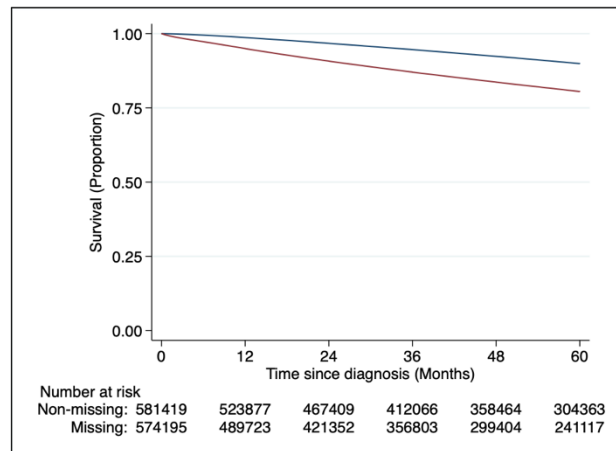
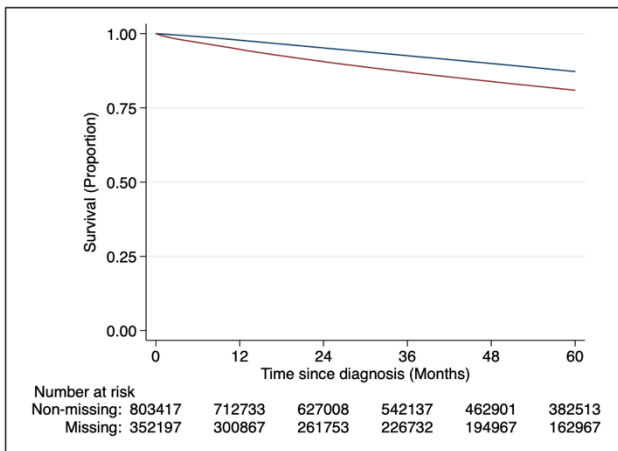
A. Non-small cell lung cancer



B. Breast cancer



C. Prostate cancer



— Non-missing data  
 — Missing data