

Appendix C: Summary of clinical variables common in most cancer types.

cvars	Description	Values
yearstobirth	Age (at first diagnosis)	Integer
vitalstatus	Vital status	Binary; 1 - dead, 0 - alive/censored
daystodeath	Number of days to death (overall survival time)	Integer or NA if the vitalstatus is 0
daystolastfollowup	Number of days to the last follow-up (last known survival time)	Integer or NA (sometimes) if the vitalstatus is 1
gender	Gender	Categorical; "female" or "male"
race	Race	Categorical; "asian", "white", "black or african american", "american indian or alaska native", or "native hawaiian or other pacific islander"
ethnicity	Ethnicity	Categorical; "hispanic or latino" or "not hispanic or latino"
pathologicstage	Pathologic stage	Categorical; vary slightly based on cancer types but in general could range from Stage I to IV, and sub-stages such as Stage Ia, Ib, etc.
pathologyTstage	Tumor stage (in TNM staging system) describing the size and location of the tumor	Categorical; vary based on cancer types but in general could be "TX", "T0", "Tis", "T1", "T2", "T3", "T4", and substages like "T2a", "T2b" etc.
pathologyNstage	Lymph nodes status (in TNM staging system) describing if the cancer has spread into nearby lymph nodes	Categorical; vary based on cancer types but in general could be "NX", "N0", "N1", "N2", "N3", and substages such as "N1a", "N2a", etc
pathologyMstage	Metastasis status (in TNM staging system) describing if the cancer has spread to other parts of the body	Categorical; vary based on cancer types but in general could be "MX", "M0", "M1", and substages "M1a", "M1b", etc.