

**Supplementary Document SD4:** Example discrepancies associated with NIH CTR provided Conditions file data

NCT ID	Condition Name	Discrepancies	Standardized WHO database term
NCT02172768	acute myeloid leucaemia	Leukemia is not spelled correctly.	acute myeloid leukaemia
NCT02658838	lovastatin/ticagrelor [va drug interaction]	Drug names are presented instead of diseases.	Does not exist
NCT01291602	healthy male and female japanese volunteers	Description of study participant is provided instead of diseases.	Does not exist
NCT04323956	ann arbor stage iii follicular lymphoma	Extra information provided on staging of the tumor.	follicular lymphoma
NCT02637531	adrenocortical carcinoma (part g)	Extra information presented inside parenthesis.	adrenal cortical carcinoma
NCT03712605	pathologic stage i merkel cell carcinoma ajcc v8	Extra information on staging of the tumor.	merkel cell carcinoma
NCT01780740	disorder; heart, functional, postoperative, cardiac surgery	Multiple clinical terms entered in a comma separated format.	Does not exist
NCT01963481	metastatic beast cancer	Breast cancer is spelled incorrectly.	invasive breast carcinoma of no special type (5th Edition WHO) , invasive breast carcinoma (3rd Edition WHO)
NCT01782235	primary sjögren's syndrome (pss)	Special characters present in disease name.	Does not exist
NCT01209195	locally advanced/metastatic or recurrent ovarian cancer, fallopian tube cancer,	Multiple tumor names	Does not exist
NCT01863108	tumor vaccines	Description of tumor intervention	Does not exist
NCT02018874	solid tumors and non-	Solid tumors is a vague term	Does not exist

	hodgkin's lymphoma	whereas non-Hodgkin's lymphoma is specific and a childhood tumor.	
NCT05050630	at least one positive lesion according to the 2014 lugano criteria for hodgkin's and non-hodgkin's lymphoma	Vague tumor description. It is not clear whether the tumor is Hodkin's or non-Hodgkin's tumor.	Does not exist