

Data Dictionary for dataset provided as part of course: Survivial Analysis in MCPS using R

For queries please contact the course organisers.

Data is 'MCPS-like' - it is a dummy dataset provided to facilitate the course practical, in a format similar to available study data. For ease of use, there is no missing data. Please see MCPS Data Showcase for full details of all available study data variables: <https://datashare.ndph.ox.ac.uk>

For coded variables with many levels of response, see footnote at the end of document for details.

History of disease is self-report only.
For the question on medication use, participants were asked to report medications that they had taken for more than a month. MCPS clinician-derived medication variables (variable names containing DRUG_) are based on participant-reported medication use and indicate the presence/absence of mention of particular classes of medicines.

Dataset	Variable name	Number of individuals with data	Description
PRACTICAL	PATID	20000	Participant ID
PRACTICAL	YEAR_RECRUITED	20000	Year participant recruited
PRACTICAL	MONTH_RECRUITED	20000	Month participant recruited
PRACTICAL	MALE	20000	Male sex (1=yes, 0=no)
PRACTICAL	CALENDAR_AGE	20000	Self-reported year of age (date of birth was not asked at the baseline interview; participant was just asked how old they were)
PRACTICAL	AGE	20000	True (non-integer) age, approximated as CALENDAR_AGE + 0.5 years
PRACTICAL	COYOACAN	20000	Resident in Coyoacan (1=yes, 0=no (ie, lives in Iztapalapa))
PRACTICAL	EDUGP	20000	Educational level (coded 1-14 - see data dictionary)
PRACTICAL	EDU_LEVEL	20000	Educational level (1=University/college, 2=High school, 3=Elementary, 4=Other)
PRACTICAL	SMOKEGP	20000	Smoking group (1=Never, 2=Ex [quit>=3 yrs ago], 3=Current [<daily], 4=Current [<10/day], 5=Current [>=10/day])
PRACTICAL	ALCGP	20000	Alcohol group (1:5 = {Never, Former, up to 3 times a month, up to 2 times a week, 3+ times a week})
PRACTICAL	PHYSGP	20000	Physical activity group (1=None; 2=Up to 2 times a week; 3=3+ times a week)
PRACTICAL	HEIGHT	20000	Height (cm)
PRACTICAL	WEIGHT	20000	Weight (kg)
PRACTICAL	WAISTC	20000	Waist circumference (cm)
PRACTICAL	HIPC	20000	Hip circumference (cm)
PRACTICAL	BASE_DIABETES	20000	Disease history at baseline (diabetes: 1=yes, 0=no)
PRACTICAL	DRUG_ANTIDIABETIC	20000	Self-reported medication use at baseline: Any drug for diabetes (1=yes, 0=no)
PRACTICAL	BASE_HBA1C	20000	Baseline glycated HbA1c (%)
PRACTICAL	STATUS	20000	Vital status at 30th Sept 2022 (A=Alive, D=Dead, U=Uncertain (typically removed from prospective analyses of mortality))
PRACTICAL	PERSON_YEARS	20000	Follow-up (years) - follow-up is to 30th September 2022
PRACTICAL	D000	20000	Death from any cause (1=yes, 0=no)
PRACTICAL	D015	20000	Death from a vascular cause (1=yes, 0=no)

Educational level at baseline is coded as follows: 1=Illiterate, 2=Can read, 3=Can read and write, 4=Incomplete elementary, 5=Complete elementary, 6=Technical studies with complete elementary, 7=Incomplete high school, 8=Complete high school, 9=Technical studies with complete high school, 10=College, 11=Technical studies with complete college level, 12=Incomplete university, 13=Complete university, 14=Post graduate degree.