NEPS
National Educational Panel Study

Research Data

Starting Cohort 6: Adults (SC6)
SUF-Version 1.0.0
Data Manual [Supplement B]
Merging Data
Jan Skopek

master file \ using file	pTarget	Reduced pTarget* (1 row per ID_t)	spSchool	spVocPrep	spMilitary	spVocTrain	spEmp	spUnemp	spParLeave	spGap	spPartner	spChild	spCourses	spFurtherEdu1	spFurtherEdu2	spFurtherEdu3	Methods	Biography	Education	FurtherEducation	Children	MaritalStates	Basics	RegioInfas
pTarget	▨		1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:1 ID_t wave						n:1 ID_t	
Reduced pTarget* (1 row per ID_t)		▨	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:1 ID_t	1:n ID_t
spSchool	n:1 ID_t wave	n:1 ID_t	▨														n:1 ID_t wave	n:1 ID_t splink	n:1 ID_t splink				n:1 ID_t	
spVocPrep	n:1 ID_t wave	n:1 ID_t		▨													n:1 ID_t wave	n:1 ID_t splink	n:1 ID_t splink				n:1 ID_t	
spMilitary	n:1 ID_t wave	n:1 ID_t			▨								1:1 ID_t splink subspell				n:1 ID_t wave	n:1 ID_t splink		n:m ID_t splink			n:1 ID_t	
spVocTrain	n:1 ID_t wave	n:1 ID_t				▨							n:1 ID_t course				n:1 ID_t wave	n:1 ID_t splink	n:1 ID_t splink	n:1 ID_t splink			n:1 ID_t	
spEmp	n:1 ID_t wave	n:1 ID_t					▨						1:1 ID_t splink subspell				n:1 ID_t wave	n:1 ID_t splink		n:m ID_t splink			n:1 ID_t	
spUnemp	n:1 ID_t wave	n:1 ID_t						▨					1:1 ID_t splink subspell				n:1 ID_t wave	n:1 ID_t splink		n:m ID_t splink			n:1 ID_t	
spParLeave	n:1 ID_t wave	n:1 ID_t							▨			n:1 ID_t child subspell	1:1 ID_t splink subspell				n:1 ID_t wave	n:1 ID_t splink		n:m ID_t splink	n:1 ID_t child		n:1 ID_t	
spGap	n:1 ID_t wave	n:1 ID_t								▨			1:1 ID_t splink subspell				n:1 ID_t wave	n:1 ID_t splink		n:m ID_t splink			n:1 ID_t	
spPartner	n:1 ID_t wave	n:1 ID_t									▨						n:1 ID_t wave						n:1 ID_t	
spChild	n:1 ID_t wave	n:1 ID_t							1:n ID_t child subspell			▨					n:1 ID_t wave				n:1 ID_t child		n:1 ID_t	
spCourses	n:1 ID_t wave	n:1 ID_t			1:1 ID_t splink subspell		1:1 ID_t splink subspell	1:1 ID_t splink subspell	1:1 ID_t splink subspell	1:1 ID_t splink subspell			▨	1:1 ID_t course			n:1 ID_t wave	1:1 ID_t splink		1:1 ID_t course			n:1 ID_t	
spFurtherEdu1	n:1 ID_t wave	n:1 ID_t											1:1 ID_t course	▨	1:1 ID_t course		n:1 ID_t wave	n:1 ID_t wave		1:1 ID_t course			n:1 ID_t	
spFurtherEdu2	n:1 ID_t wave	n:1 ID_t				1:n ID_t course							1:1 ID_t course	1:1 ID_t course	▨		n:1 ID_t wave	n:1 ID_t wave		1:1 ID_t course			n:1 ID_t	
spFurtherEdu3	n:1 ID_t wave	n:1 ID_t														▨	n:1 ID_t wave						n:1 ID_t	
Methods	1:1 ID_t wave	n:1 ID_t	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	1:n ID_t wave	▨	1:n ID_t wave		1:n ID_t wave			n:1 ID_t	
Biography		n:1 ID_t	1:n ID_t splink	1:n ID_t splink	1:n ID_t splink	1:n ID_t splink	1:n ID_t splink	1:n ID_t splink	1:n ID_t splink	1:n ID_t splink			1:1 ID_t splink				n:1 ID_t wave	▨	1:1 ID_t splink	1:n ID_t splink			n:1 ID_t	
Education		n:1 ID_t	1:n ID_t splink	1:n ID_t splink		1:n ID_t splink													▨	1:1 ID_t splink			n:1 ID_t	
FurtherEducation		n:1 ID_t			n:m ID_t splink	1:n ID_t splink	n:m ID_t splink	n:m ID_t splink	n:m ID_t splink	n:m ID_t splink			1:1 ID_t course	1:1 ID_t course	1:1 ID_t course		n:1 ID_t wave	n:1 ID_t splink		▨			n:1 ID_t	
Children		n:1 ID_t							1:n ID_t child			1:n ID_t child									▨		n:1 ID_t	
MaritalStates		n:1 ID_t																				▨	n:1 ID_t	
Basics	1:n ID_t	1:1 ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	1:n ID_t	▨	1:n ID_t
RegioInfas		n:1 ID_t																					n:1 ID_t	▨

* e.g. pTarget reduced to have only last wave record for each person (ID_t)

Legend: While there are numerous ways of linking data files, the merging matrix lists the most important linkage scenarios between data files. Rows represent the "master" file (the data matrix that is currently being loaded) that is supposed to be merged with the "using" file (that is the data matrix that will be joined to the right of the master case by case). Cells specify the type of relationship between master and using file as well as a set of linkage variables. For instance, a "1:1" relationship with ID_t as linkage variable means that one record from the master can be uniquely linked by ID_t to one record of the using file. A "1:n" relationship with "ID_t splink" as linkage variables means, that one record (being uniquely identified via ID_t splink) of the master dataset is linked to multiple records in the using file (where one record is not uniquely identified via ID_t splink). Please note that albeit a merging-relationship indicated in the matrix might be technical possible it has not to be necessarily meaningful by a substantive point of view. Please treat n:m relationships with caution since there is no unique logical assignment; for those cases always try to reduce at least one dataset to achieve a unique record matching (e.g., in spell files keep only subspell=0 observations).