Core Variables Harmonisation Manual

CONTENTS

1. AIM

The aim of work package 1 of LifeCycle is:

"To harmonise core data related to stressors and outcomes of interest, including socio/economic, migration, urban environment, lifestyle related stressors and cardio-metabolic, respiratory and mental health related outcomes in the EU Child cohort studies"

We have developed a list of core variables related to stressors and outcomes of interest for the EU Child Cohort Network. The definitions for these are based on previous harmonisations in other studies, scientific literature, expert knowledge, international classification systems and, most importantly, what data are available in participating cohorts.

Since these are core variables, the list aims to be as inclusive as possible. In formulating the definitions for the core variables, a major criterion has been that as many cohorts as possible should be able to derive the variable. In some instances, this comes with the cost of losing some information. It will also be the case that not every cohort can derive/harmonise every variable, or that a cohort can only partially harmonise some variables.

2. THE CORE VARIABLES

2.1 The Core Variables Table

The core variables table provides a description of the LifeCycle core variables: their name, description and instructions for how to derive the harmonised LifeCycle variables. A brief explanation of the table headings is given below (Table 2). The actual core variables table follows on the following pages (Table 3).

When harmonising the variables, please use the cleanest variables available within your cohort. Create all variables. If no data exist within your cohort for a given variable, the variable remains empty. An example is provided in Table 1 below.

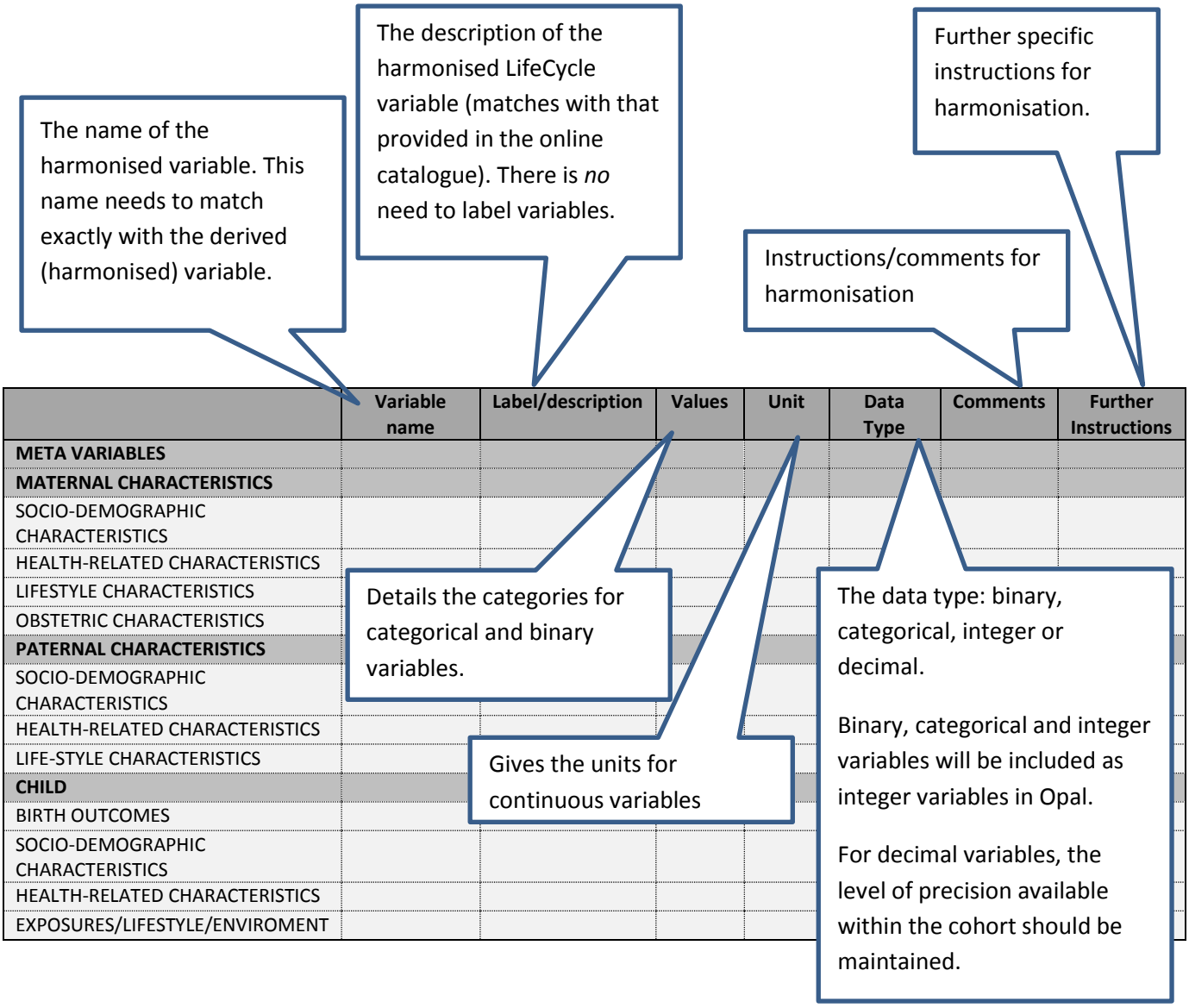
mother_id	preg_no	child_no	child_id	cohort_id	preg_thyroid	preg_fever
100025	1	1	1000250101	120	0	
100025	2	1	1000250201	120	0	
100026	1	1	1000260101	120	0	
100026	1	2	1000260102	120	0	
100027	1	1	1000270101	120	1	

Table 1. An example of an empty variable. The cohort is missing information on fever in the mother during pregnancy. The cohort has created the variable "preg_fever", but it remains empty.

We have assigned three different levels of priority to the variables: highest priority has been given to the variables that will be analysed during the DataSHIELD workshop at the Oulu meeting in June (highlighted in pink); second highest priority has been given to the variables we deem most critical to future analyses (highlighted in blue); third highest priority has been given to the remaining core variables. We suggest that you begin with the highest priority variables.

Please record a description of harmonisation, to be entered in the online catalogue. This includes a description of the source variables, a description of harmonisation and whether the variable is fully or partially harmonised. Where a variable is only partially harmonised, please provide an explanation for why the variable is partially harmonised in the harmonisation description. An outline of the Excel templates for the source variable descriptions and harmonisation descriptions are provided in Appendix I of this manual; the actual Excel templates have been sent to you and are also available to download from the LifeCycle intranet.

If you have any queries about harmonisation or the core variables list please contact Angela Pinot de Moira (anpi@sund.ku.dk) and Pernille Stemann Larsen (pernillelarsen@sund.ku.dk) from WP1.



	Variable name	Label/description	Values	Unit	Data Type	Comments	Further Instructions
META VARIABLES							
MATERNAL CHARACTERISTICS							
SOCIO-DEMOGRAPHIC CHARACTERISTICS							
HEALTH-RELATED CHARACTERISTICS							
LIFESTYLE CHARACTERISTICS							
OBSTETRIC CHARACTERISTICS							
PATERNAL CHARACTERISTICS							
SOCIO-DEMOGRAPHIC CHARACTERISTICS							
HEALTH-RELATED CHARACTERISTICS							
LIFE-STYLE CHARACTERISTICS							
CHILD							
BIRTH OUTCOMES							
SOCIO-DEMOGRAPHIC CHARACTERISTICS							
HEALTH-RELATED CHARACTERISTICS							
EXPOSURES/LIFESTYLE/ENVIROMENT							

Table 2. An explanation of the table headings in the core variables table.

	highest priority variables						
	2nd highest priority variables						
	3rd highest priority variables						
	Variable name	**Label/description**	**Values**	**Unit**	**Data Type**	**Comments**	**Further instructions**
META VARIABLES							
Mother identifier	mother_id	Unique identifier number for the mother				Either the original id or a new id generated by the cohort	
Pregnancy number	preg_no	Within-mother pregnancy number for each pregnancy included in the cohort				Increases with each subsequent pregnancy included in the cohort (i.e. '1' is allocated to the first pregnancy included, '2' is allocated to the second pregnancy included, etc.). For cohorts with only one pregnancy per a mother, this number will always be '1'.	
Child number	child_no	Within-pregnancy birth order				For single child pregnancies, assign '1' to the child. For multiple child pregnancies, each child is numbered in the order that they were born (i.e. the first-born child from a multiple pregnancy is assigned the number '1', the second-born child is assigned the number '2', etc..)	
Child identifier	child_id	Unique identifier number for the index child				Either the original id or a new id generated by the cohort	
Cohort id	cohort_id	Unique identifier number for the cohort	(101) Gen R (102) INMA (103) NINFEA (104) SWS (105) ALSPAC (106) DNBC (107) BIB (108) GECKO (109) RHEA (110) MOBA (111) ELFE (112) EDEN (113) NFBC66 (114) NFBC86 (115) HBCS (116) CHOP (117) RAINE		Categorical		
Recruitment age	recruit_age	Age of the child Indicative of time of enrolment	In days. Negative if prepartum, positive if postpartum	days	Integer		

Country of cohort	coh_country	Country of the cohort	(36) Australia (208) Denmark (246) Finland (250) France (276) Germany (300) Greece (380) Italy (528) Netherlands (578) Norway (724) Spain (826) United Kingdom		Categorical	Based on ISO 3166 numeric country codes	
MATERNAL CHARACTERISTICS:							
SOCIO-DEMOGRAPHIC CHARACTERISTICS							
Cohabitation status	cohab_0 cohab_1 cohab_2 cohab_3 … cohab_17	Cohabitation status of the mother: are her and her partner living together as a couple? Repeated measures: cohab_0: at birth or as near to birth as possible and within one year of birth (child's age ≥0 year and <1 year) cohab_1: measure when the child is aged between ≥1 and <2 years, if more measures declared within the period use the best measure or measure closest to child's first birthday cohab_2: measure when the child is aged between ≥2 and <3 years, if more than one declared use best measure or measure closest to child's second birthday cohab_3: measure when the child is aged between ≥3 and <4 years, if more than one declared, use best measure or measure closest to child's third birthday .... cohab_17: measure when the child is aged between ≥17 and <18 years, if more than one declared use best measure or measure closest to child's seventeenth birthday	1) Yes, living as a couple 2) No, not living as a couple		Binary	"Mother's partner" can be the biological partner, a new partner or a partner of the same gender.	

Maternal occupational status (core)	occup_m_0 occup_m_1 occup_m_2 …. occup_m_17	Occupational status of the mother. If more than one status is recorded within the defined time frame use the highest reported level. Repeated measures: occup_m_0: highest reported level within one year of birth (child aged between >-1 year and <1 year) occup_m_1: highest reported level when the child was aged between ≥1 and <2 years occup_m_2 highest reported level when the child was aged between ≥2 and <3 years occup_m_3: highest reported level when the child was aged between ≥3 and <4 years .... occup_m_17: highest reported level when the child was aged between ≥17 and <18 years	1) Employed 2) Self-employed 3) Unemployed 4) Student, apprentice, student 5) Domestic tasks (housewife etc.) 6) Inactive/other (Receiving benefits or pension etc.)		Categorical	Employed/self-employed includes maternity leave if she was employed before commencing maternity leave. If a cohort does not have data specifying whether the mother is employed or self-employed, categorise the mother as employed; the variable will be partially harmonised, which should be highlighted and described in the online catalogue	
Maternal occupational codes (core)	occupcode_m_0 occupcode_m_1 occupcode_m_2 …. occupcode_m_17	Occupation of the mother classified according to ISCO-88 1-digit codes. If more than one occupation is recorded within the defined time frame use the highest reported level. Repeated measures: occupcode_m_0: highest reported level within one year of birth (child aged between >-1 year and <1 year) occupcode_m_1: highest reported level when the child was aged between ≥1 and <2 years occupcode_m_2: highest reported level when the child was aged between ≥2 and <3 years occupcode_m_3: highest reported level when the child was aged between ≥3 and <4 years .... occupcode_m_17: highest reported level when the child was aged between ≥17 and <18 years	1) Legislators, senior officials, managers 2) Professionals 3) Technicians and associate professionals 4) Clerks 5) Service workers, and shop and market sales workers 6) Skilled agricultural, and fishery workers 7) Craft and related trades workers 8) Plant and machine operators, and assemblers 9) Elementary occupations 0) Armed forces occupations		Categorical	Classified according to ISCO-88 (International Standard Classification of Occupations 1988) 1-digit codes. A description of this classification scheme can be found here: http://www.ilo.org/public/english/bureau/stat/isco/isco88/index.htm A link to ISCO-88 codes and titles can be found here: http://ec.europa.eu/eurostat/documents/1978984/6037342/ISCO-88-COM.pdf.	

Maternal education (core)	edu_m_0 edu_m_1 edu_m_2 … edu_m_17	Level of education based on the highest on-going or completed education. If more than one education level is reported within the defined time frame, use highest recorded education level. Repeated measures: edu_m_0: highest reported level of education within one year of birth (child aged between >-1 year and <1 year) edu_m_1: highest reported level of education when the child was aged between ≥1 and <2 years edu_m_2 highest reported level of education when the child was aged between ≥2 and <3 years edu_m_3: highest reported level of education when the child was aged between ≥3 and <4 years .... edu_m_17: highest reported level of education when the child was aged between ≥17 and <18 years	1) High 2) Medium 3) Low		Categorical	Classification according to International Standard Classification of Education 97/2011 (ISCED-97/2011) High: Short cycle tertiary, Bachelor, Masters, Doctoral or equivalent (ISCED-2011: 5-8, ISCED-97: 5-6) Medium: Upper secondary, Post-secondary non-tertiary (ISCED-2011: 3-4, ISCED-97: 3-4) Low: No education; early childhood; pre-primary; primary; lower secondary or second stage of basic education. (ISCED-2011: 0-2, ISCED-97: 0-2) Mapping tools for specific countries can be found here: http://uis.unesco.org/en/isced-mappings	
Maternal country of birth	cob_m	Maternal country of birth	0) Born in country of cohort 1) Born in EU country (outside cohort country) 2) Born in other country		Categorical	According to EU SILC classification	
Ethnicity	ethn1_m	(Ethnic) background of mother based on country of origin (of parents)	1) Western 1) Non-western 3) Mixed		Categorical	Western countries include European Union, Andorra, Australia, Canada, Iceland, Liechtenstein, Monaco, New Zealand, Norway, San Marino, Switzerland, USA and Vatican City. Non-western countries include all other countries.	
	ethn2_m	Ethnic background of mother based on colour	1) White (Caucasian) 2) Non-white (non-Caucasian) 3) Mixed		Categorical		
	ethn3_m	Best estimate of mother's ethnic background based on ethn1_m or ethn2_m and cohort's own discretion.	1) Western 2) Non-western 3) Mixed		Categorical		
Maternal age at birth (years)	agebirth_m_y	Mother's age at delivery in complete years.	Continuous in years	years	Integer		Unusual values should be checked (e.g. mother's age outside the range 15-50 years)
Maternal age at birth (days)	agebirth_m_d	Mother's age at delivery in days	Continuous in days	days	Integer		

Maternal death	death_m	Indicator variable for death of the child's mother	0) No 1) Yes		Binary		
Age of child at death of the mother	death_m_age	Age of the child at the time of mother's death	Continuous in days	days	Integer		
HEALTH-RELATED CHARACTERISTICS							
Pre-pregnancy weight	prepreg_weight	Pre-pregnancy weight		kg	Decimal	Use measured weight if available If pre-pregnancy weight is not available, use early pregnancy weight closest to conception, limited to 1st trimester (<12 weeks). If early pregnancy weight is used the variable is partially harmonised; partial harmonisation should be detailed in the online catalogue. If an early pregnancy weight is used in place of pre-pregnancy weight, gestational age at time of measurement should be provided in variable "prepreg_weight_ga".	
	prepreg_weight_mes	Reported vs. measured pre-pregnancy weight	0) Self-reported 1) Measured		Binary		
	prepreg_weight_ga	Gestational age of mother when early pregnancy weight measured if early pregnancy weight is used as a surrogate for pre-pregnancy weight	Continuous in days (gestational age at measurement)	days	Integer	Missing if prepreg_weight is pre-pregnancy weight (as opposed to early pregnancy weight)	
Late pregnancy weight	latepreg_weight	Latest measurement of weight before delivery		kg	Decimal	Use measured weight if available If a pregnancy weight within one week of delivery is not available, a late pregnancy weight, closest to delivery from 32 weeks gestation is acceptable. This will be partial harmonisation, which should be indicated and detailed in the online catalogue.	
	latepreg_weight_mes	Reported vs. measured late-pregnancy weight	0) Self-reported 1) Measured		Binary		
	latepreg_weight_ga	Gestational age when late-pregnancy weight measured	Continuous in days (gestational age at measurement)	days	Integer		

Maternal gestational weight gain	preg_gain	Maternal gestational weight gain; weight gain from conception to delivery.		kg	Decimal	Use measured weight if available Negative values for weight loss: positive values for weight gain. 0 for no change. Where gestational weight gain is calculated: take the measures perceived to be the **best measurements** and report in detail in the online catalogue exactly what measures have been used to derive the variable.	
	preg_gain_mes	Reported vs. measured gestational weight gain	0) Self-reported 1) Measured 2) Pre-pregnancy weight measured and late-pregnancy weight reported 3) Pre-pregnancy weight reported and late-pregnancy weight recorded		Binary		
Maternal height	height_m	Maternal height		cm	Decimal	Use measured height if available	
	height_mes_m	Reported vs. measured maternal height	0) Self-reported 1) Measured		Binary		
History of diabetes (all kinds)	prepreg_dia	Diagnosis of mother with diabetes before index pregnancy	1) No 2) Type I 3) Type II 4) Gestational diabetes 5) Other/unspecified		Categorical		
Gestational diabetes	preg_dia	Gestational diabetes during index pregnancy	0) No 1) Yes		Binary	Gestational diabetes: glucose intolerance with onset or first recognition during pregnancy continuing beyond 24–28 weeks of gestation.	The variable is fully harmonised if data cover the whole pregnancy
Thyroid disorders during pregnancy	preg_thyroid	Hyper or hypo thyroid disorder during pregnancy	0) No 1) Yes		Binary		The variable is fully harmonised if data cover the whole pregnancy
Fever during pregnancy	preg_fever	Any fever during pregnancy	0) No 1) Yes		Binary		The variable is fully harmonised if data cover the whole pregnancy

Preeclampsia/HELLP syndrome	preeclam	Preeclampsia or HELLP syndrome during pregnancy	0) No 1) Yes		Binary	Preeclampsia is defined as: i) elevated blood pressure after 20 weeks of gestation (≥ 140 mm Hg systolic or ≥ 90 mm Hg diastolic) ii) proteinuria (> 0.3 g/24 hours). HELLP syndrome (haemolysis, elevated liver enzymes, and low platelets) is a manifestation or complication of preeclampsia	The variable is fully harmonised if data cover from 20 weeks until the end of pregnancy
Gestational hypertension	preg_ht	Hypertension during pregnancy	0) No 1) Yes		Binary	Gestational hypertension is the new onset of hypertension after 20 weeks of gestation. The diagnosis requires that the patient have: i) Elevated blood pressure (systolic ≥ 140 or diastolic ≥ 90 mm Hg, the latter measured using the fifth Korotkoff sound) ii) Previously normal blood pressures iii) No protein in the urine iv) No manifestations of preeclampsia Gestational hypertension is diagnosed retrospectively when the patient does not develop preeclampsia and if blood pressure returns to normal by the 12-week postpartum visit	The variable is fully harmonised if data cover from 20 weeks until the end of pregnancy
History of asthma	asthma_m	Maternal history of asthma before pregnancy (of index child)	0) No 1) Yes		Binary	Where data are available, asthma should be doctor diagnosed. If no information is available on doctor diagnosis, the variable is partially harmonised.	
Pre-pregnancy psychiatric disorders	prepreg_psych	Maternal history of any psychiatric disorder before pregnancy (Self-reported or diagnosed)	0) No 1) Yes		Binary	If there are no details of *any* psychiatric disorder but there are details of, for example, depression and anxiety, these can be included as "any" psychiatric disorder; this will then be only partial harmonisation, which should be detailed in full under "match" in the harmonisation table in the online catalogue. Visits to a psychologists or psychiatrist should only be used as an indication of a psychiatric disorder if no other data are available. This being the case, the variable will be partially harmonised (which should be detailed in full in the online catalogue).	
Psychiatric disorders during pregnancy	preg_psych	Any type of maternal psychiatric disorder during pregnancy (reported or diagnosed)	0) No 1) Yes		Binary	If there are no details of any psychiatric disorder but there are details of, for example, depression and anxiety, these can be included as "any" psychiatric disorder; this will then be only partial harmonisation, which should be detailed in full under "match" in the harmonisation table in the online	

						catalogue. Visits to a psychologists or psychiatrist should only be used as an indication of a psychiatric disorder if no other data are available. This being the case, the variable will be partially harmonised (which should be detailed in full in the online catalogue).	
Post-partum depression	ppd	Postpartum depression	0) No 1) Yes		Binary	**Complete harmonisation:** ppd is identified using a defined scale (for e.g. the Edinburgh postnatal depression scale) or is clinically diagnosed. **Partial harmonisation:** reported depression or psychiatric disorder around the time of birth	
LIFESTYLE CHARACTERISTICS							
Pre-pregnancy smoking	prepreg_smk	Smoking before pregnancy. At any time pre-partum	0) No 1) Yes		Binary		
	prepreg_cig	Cigarettes per a day before pregnancy At any time pre-partum	0) None 1) < 10 per day 2) ≥ 10 per day		Categorical	Cigars, pipes, cheroots should be converted into number of cigarettes (1:3). 1 cigar, pipe, cheroot etc. is equivalent to 3 cigarettes. (OECD and The Danish Cancer Society). Non-smokers categorised as 0 - none	
Maternal smoking during pregnancy	preg_smk	Any smoking during pregnancy	0) No 1) Yes		Binary		
	preg_cig	Average number of cigarettes smoked per day during pregnancy	0) None 1) < 10 per day 2) ≥ 10 per day		Categorical	Cigars, pipes, cheroots etc. converted into cigarettes; see above. Non-smokers categorised under 0 - none	
	smk_t1	Smoking during first trimester	0) No 1) Yes		Binary		
	smk_t2	Smoking during second trimester	0) No 1) Yes		Binary		
	smk_t3	Smoking during third trimester	0) No 1) Yes		Binary		
Pre-pregnancy alcohol intake	prepreg_alc	Did the mother drink alcohol before pregnancy? (ever)	0) No (never drinking) 1) Yes		Binary		
	prepreg_alc_unit	Average units of alcohol mother drank per week before pregnancy	0) 0 (Never drinking alcohol) 1) Light (< 3 units per week) 2) Moderate (≥3 - <7 units per week) 3) Moderate to heavy (≥7 - <14 units per week) 4) Heavy (≥14 units per week)		Categorical	Alcohol consumption before pregnancy includes ≥ 14 units per week. During pregnancy this category is not included. For converting from days to weeks apply the following rule: one drink per day corresponds to 7–13 drinks weekly, two drinks per day to 14–20 drinks weekly, etc. Non-drinkers categorised as 0 - never drinking alcohol	

Maternal alcohol intake during pregnancy	preg_alc	Maternal alcohol intake during pregnancy	0) No 1) Yes		Binary		
	preg_alc_unit	Average units of alcohol mother drank per week during pregnancy	0) 0 (Not drinking alcohol) 1) Light (>0 and <3units per week) 2) Moderate (≥3 and <7 units per week) 3) Heavy (≥7 units per week)		Categorical	Non-drinkers categorised under 0 - not drinking alcohol	
	alc_t1	Any alcohol intake in first trimester	0) No 1) Yes		Binary		
	alc_t2	Any alcohol intake in second trimester	0) No 1) Yes		Binary		
	alc_t3	Any alcohol intake in third trimester	0) No 1) Yes		Binary		
Folic acids supplement pre-conception	folic_prepreg	Intake of Folic Acids (folate, vitamin B9) before conception	0) No 1) Yes		Binary		
Folic acids supplement week 0-12	folic_preg12	Intake of Folic Acids (folate, vitamin B9) during the period from conception to early pregnancy (12 weeks)	0) No 1) Yes		Binary		
Folic acids supplement week 12+	folic_post12	Intake of Folic Acids (folate, vitamin B9) after week 12 of pregnancy.	0) No 1) Yes		Binary		
OBSTETRIC CHARACTERISTICS							
Maternal parity	parity_m	Mothers parity based on previous born children (previous stillbirths included, abortions excluded)	0) 0 1) 1 2) 2 3) 3 4) ≥ 4		Categorical	Stillbirth is defined as the death of a foetus at or after 22 completed weeks of gestation. Definition according to the International Standard of stillbirths defined and recommended by WHO for international comparison (WHO, 2006) If both Birth medical registry data and self-reported data are available, prioritisation is as follows: 1) Medical Birth registry data 2) Self-reported	Check unusual values, for e.g. mother is aged 15 years or less with two or more previous pregnancies
Planned pregnancy	preg_plan	Whether the pregnancy was planned or not planned.	1) Yes (Planned, partly planned) 2) No (Not planned)		Binary	Partly planned includes cases where the mother has stopped using contraceptives in order to become pregnant	

ART (Assisted reproductive treatment)	art	Did the mother become pregnant using ART	0) No 1) Yes		Binary	ART refers to treatments used to assist people in achieving a pregnancy	
In vitro-fertilisation (Fertility treatment)	ivf	Did the mother become pregnant using IVF or ICSI	0) No 1) Yes		Binary		
Birth outcome	outcome	The child´s condition at delivery	1) Live-born 2) Stillborn 3) Spontaneous abortion 4) Induced abortion 5) Unspecified abortion 6) Other (e.g. Molar + extrauterine pregnancy)		Categorical	A complete match is achieved when data on all categories (live born, still born, spontaneous abortion etc.) are available. When detailing the match of the variable in the catalogue it is important to include the approximate age of recruitment; in many cases this will explain a partial match. Where the match is only partial, also indicate whether there are complete data on live-born vs. stillborn. Stillborn defined according to WHO recommendations (22 completed weeks)	
Mode of delivery	mode_delivery	Mode of delivery	1) Vaginal (normal) 2) Forceps, vacuum 3) Elective caesarean 4) Emergency caesarean 5) Caesarean unspecified		Categorical	Complete match achieved when data on all categories are available, otherwise the match is partial. When detailing the match in the catalogue, include details of what data are available.	
Placental abruption at delivery	plac_abrup	Placental abruption	0) No 1) Yes		Binary		
PATERNAL CHARACTERISTICS:							
SOCIO-DEMOGRAPHIC CHARACTERISTICS							
Paternal occupational status, primary father (core)	occup_f1_0 occup_f1_1 occup_f1_2 … occup_f1_17	Occupational status of primary/main father-figure. If more than one status is recorded within the defined time frame, use the highest reported level. Repeated measures: occup_f1_0: highest reported level within one year of birth (child aged between >-1 year and <1 year) occup_f1_1: highest reported level when the child was aged between ≥1 and <2 years occup_f1_2 highest reported level when the child was aged between ≥2 and <3 years occup_f1_3: highest reported level when the child was aged between ≥3 and <4 years .... occup_f1_17: highest reported level when the child was aged between ≥17 and <18 years	1) Employed 2) Self-employed 3) Unemployed 4) Student, apprentice, student 5) Domestic tasks (housewife etc.) 6) Inactive/other (Receiving benefits or pension etc.)		Categorical	If a cohort does not have data specifying whether the father is employed or self-employed, categorise the father as employed; the variable will be partially harmonised, which should be highlighted and described in the online catalogue	

Type of father (biological father vs. social father vs. social mother), primary father's occupational status (core)	occup_f1_fath0 occup_f1_fath1 occup_f1_fath2 … occup_f1_fath17	Variables indicating whether the occupational statuses recorded in occup_f1_0, occup_f1_1, occup_f1_2 etc. are for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
Paternal occupational status, secondary father (core)	occup_f2_0 occup_f2_1 occup_f2_2 … occup_f2_17	Occupational status of secondary father-figure. If more than one status is recorded within the defined time frame, use the highest reported level. Repeated measures: occup_f2_0: highest reported level within one year of birth (child aged between >-1 year and <1 year) occup_f2_1: highest reported level when the child was aged between ≥1 and <2 years occup_f2_2: highest reported level when the child was aged between ≥2 and <3 years occup_f2_3: highest reported level when the child was aged between ≥3 and <4 years .... occup_f2_17: highest reported level when the child was aged between ≥17 and <18 years	1) Employed 2) Self-employed 3) Unemployed 4) Student, apprentice, student 5) Domestic tasks (housewife etc.) 6) Inactive/other (receiving benefits or pension etc.)		Categorical	If a cohort does not have data specifying whether the father is employed or self-employed, categorise the father as employed; the variable will be partially harmonised, which should be highlighted and described in the online catalogue	
Type of father (biological father vs. social father vs. social mother), secondary father's occupational status (core)	occup_f2_fath0 occup_f2_fath1 occup_f2_fath2 … occup_f2_fath17	Variables indicating whether the occupational statuses recorded in occup_f2_0, occup_f2_1, occup_f2_2 etc. are for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		

Paternal occupational codes, primary father (core)	occupcode_f1_0 occupcode_f1_1 occupcode_f1_2 …. occupcode_f1_17	Occupation of the primary father figure classified according to ISCO-88 1-digit codes. If more than one occupation is recorded within the defined time frame use the highest reported level. Repeated measures: occupcode_f1_0: highest reported level within one year of birth (child aged between >-1 year and <1 year) occupcode_f1_1: highest reported level when the child was aged between ≥1 and <2 years occupcode_f1_2 highest reported level when the child was aged between ≥2 and <3 years .... occupcode_f1_17: highest reported level when the child was aged between ≥17 and <18 years	1) Legislators, senior officials, managers 2) Professionals 3) Technicians and associate professionals 4) Clerks 5) Service workers, and shop and market sales workers 6) Skilled agricultural, and fishery workers 7) Craft and related trades workers 8) Plant and machine operators, and assemblers 9) Elementary occupations 0) Armed forces occupations		Categorical	Classified according to ISCO-88 (International Standard Classification of Occupations 1988) 1-digit codes. A description of this classification scheme can be found here: http://www.ilo.org/public/english/bureau/stat/isco/isco88/index.htm A link to ISCO-88 codes and titles can be found here: http://ec.europa.eu/eurostat/documents/1978984/6037342/ISCO-88-COM.pdf.	
Type of father (biological father vs. social father vs. social mother), primary father's occupational codes (core)	occupcode_f1_fath0 occupcode_f1_fath1 occupcode_f1_fath2 …. occupcode_f1_fath17	Variables indicating whether the occupation codes recorded in occupcode_f1_0, occupcode_f1_1, occupcode_f1_2 etc. are for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		

Paternal occupational codes, secondary father (core)	occupcode_f2_0 occupcode_f2_1 occupcode_f2_2 …. occupcode_f2_17	Occupation of the secondary father figure classified according to ISCO-88 1-digit codes. If more than one occupation is recorded within the defined time frame use the highest reported level. Repeated measures: occupcode_f2_0: highest reported level within one year of birth (child aged between >-1 year and <1 year) occupcode_f2_1: highest reported level when the child was aged between ≥1 and <2 years occupcode_f2_2: highest reported level when the child was aged between ≥2 and <3 years .... occupcode_f2_17: highest reported level when the child was aged between ≥17 and <18 years	1) Legislators, senior officials, managers 2) Professionals 3) Technicians and associate professionals 4) Clerks 5) Service workers, and shop and market sales workers 6) Skilled agricultural, and fishery workers 7) Craft and related trades workers 8) Plant and machine operators, and assemblers 9) Elementary occupations 0) Armed forces occupations		Categorical	Classified according to ISCO-88 (International Standard Classification of Occupations 1988) 1-digit codes. A description of this classification scheme can be found here: http://www.ilo.org/public/english/bureau/stat/isco/isco88/index.htm A link to ISCO-88 codes and titles can be found here: http://ec.europa.eu/eurostat/documents/1978984/6037342/ISCO-88-COM.pdf.	
Type of father (biological father vs. social father vs. social mother), secondary father's occupational codes (core)	occupcode_f2_fath0 occupcode_f2_fath1 occupcode_f2_fath2 …. occupcode_f2_fath17	Variables indicating whether the occupation codes recorded in occupcode_f2_0, occupcode_f2_1, occupcode_f2_2 etc. are for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		

Paternal education, primary father (core)	edu_f1_0 edu_f1_1 edu_f1_2 … edu_f1_17	Primary father figure's level of education based on the highest on-going or completed education. If more than one education level is reported within the defined time frame, use highest recorded education level. Repeated measures: edu_f1_0: highest reported level of education within one year of birth (child aged between >-1 year and <1 year) edu_f1_1: highest reported level of education when the child was aged between ≥1 and <2 years edu_f1_2 highest reported level of education when the child was aged between ≥2 and <3 years .... edu_f1_17: highest reported level of education when the child was aged between ≥17 and <18 years	1) High 2) Medium 3) Low		Categorical	Classification according to International Standard Classification of Education 97/2011 (ISCED-97/2011) High: Short cycle tertiary, Bachelor, Masters, Doctoral or equivalent (ISCED-2011: 5-8, ISCED-97: 5-6) Medium: Upper secondary, Post-secondary non-tertiary (ISCED-2011: 3-4, ISCED-97: 3-4) Low: No education; early childhood; pre-primary; primary; lower secondary or second stage of basic education. (ISCED-2011: 0-2, ISCED-97: 0-2) Mapping tools for specific countries can be found here: http://uis.unesco.org/en/isced-mappings	
Type of father (biological father vs. social father vs. social mother), primary father's education	edu_f1_fath0 edu_f1_fath1 edu_f1_fath2 …. edu_f1_fath17	Variables indicating whether the education levels recorded in edu_f1_0, edu_f1_1, edu_f1_2 etc. are for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		

Paternal education, secondary father (core)	edu_f2_0 edu_f2_1 edu_f2_2 … edu_f2_17	Secondary father figure's level of education based on the highest on-going or completed education. If more than one education level is reported within the defined time frame, use highest recorded education level. Repeated measures: edu_f2_0: highest reported level of education within one year of birth (child aged between >-1 year and <1 year) edu_f2_1: highest reported level of education when the child was aged between ≥1 and <2 years edu_f2_2 highest reported level of education when the child was aged between ≥2 and <3 years .... edu_f2_17: highest reported level of education when the child was aged between ≥17 and <18 years	1) High 2) Medium 3) Low		Categorical	Classification according to International Standard Classification of Education 97/2011 (ISCED-97/2011) High: Short cycle tertiary, Bachelor, Masters, Doctoral or equivalent (ISCED-2011: 5-8, ISCED-97: 5-6) Medium: Upper secondary, Post-secondary non-tertiary (ISCED-2011: 3-4, ISCED-97: 3-4) Low: No education; early childhood; pre-primary; primary; lower secondary or second stage of basic education. (ISCED-2011: 0-2, ISCED-97: 0-2) Mapping tools for specific countries can be found here: http://uis.unesco.org/en/isced-mappings	WP3 will provide more detailed variables
Type of father (biological father vs. social father vs. social mother), secondary father's education	edu_f2_fath0 edu_f2_fath1 edu_f2_fath2 …. edu_f2_fath17	Variables indicating whether the education levels recorded in variables edu_f2_0, edu_f2_1, edu_f2_2 etc. are for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
Paternal country of birth	cob_p	Paternal country of birth	1) Born in country of cohort 2) Born in EU country (outside cohort of country) 3) Born in other country		Categorical	According to EU SILC classification	
Type of father, paternal country of birth	cob_p_fath	Variable indicating whether country of birth recorded in variable cob_p_fath is for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
Paternal ethnicity	ethn1_p	Father's (ethnic) background based on country of origin (of father's parents)	1) Western 2) Non-western 3) Mixed		Binary	Western countries include European Union, Andorra, Australia, Canada, Iceland, Liechtenstein, Monaco, New Zealand, Norway, San Marino, Switzerland, USA and Vatican City. Non-western countries include all other countries.	
	ethn2_p	Father's ethnic background based on colour	1) White (Caucasian) 2) Non-white (non-Caucasian) 3) Mixed		Binary		

	ethn3_p	Best estimate of father's ethnic background based on ethn1_p or ethn2_p and cohort's own discretion.	1) Western 2) Non-western 3) Mixed		Binary		
Type of father, paternal ethnicity	ethn_p_fath	Variable indicating whether the ethnicities recorded in variables ethn1_p, ethn2_p, ethn3_p are for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
Paternal age at birth (years)	agebirth_p_y	Father's age at birth in complete years	Continuous in years	years	Integer		
Paternal age at birth (days)	agebirth_p_d	Father's age at birth in days	Continuous in days	days	Integer		
Type of father, paternal age	agebirth_p_fath	Variable indicating whether the age recorded in variables agebirth_p_y and agebirth_p_d is for the biological father, the social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
Paternal death	death_p	Indicator variable for death of the child's father	0) No 1) Yes		Binary		
Age of death of the father	death_p_age	Age of the child at the time of father's death	Continuous in days	days	Integer		
Type of father, paternal death	death_p_fath	Variable indicating whether the paternal death recorded in variable death_p, is the death of the biological father, a social father/mother or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
HEALTH-RELATED CHARACTERISTICS							
Paternal weight, primary father	weight_f1	Primary/main father figure´s weight reported in first paternal questionnaire		kg	Decimal	Use measured weight if available. Note in the catalogue when it was recorded, i.e., which follow-up	
Weight measurement, primary father	weight_mes_f1	Reported vs. measured weight	0) Self-reported 1) Reported by others 2) Measured		Binary		
Type of father, primary father's weight	weight_f1_fath	Variable indicating whether the weight recorded in weight_f1 is for the biological father, the social father, or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
Paternal height, primary father	height_f1	Primary/main father figure´s height reported in first paternal questionnaire		cm	Decimal	Use measured height if available. Note in the catalogue when it was recorded, i.e., which follow-up	
Height measurement, primary father	height_mes_f1	Reported vs. measured height	0) Self-reported 1) Reported by others 2) Measured		Binary		
Type of father, primary father's height	height_f1_fath	Variable indicating whether the height recorded in height_f1 is for the biological father, the social father, or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
History of diabetes (biological father)	dia_bf	Diabetes diagnosis (biological father)	1) No 2) Type I 3) Type II 4) Other/unspecified		Categorical	State in the online catalogue whether the type of father was specified at time of data collection (i.e. whether these data definitely relate to the biological and not the social father)	
History of asthma (biological father)	asthma_bf	Paternal history of asthma (biological father)	0) No 1) Yes		Binary	State in online catalogue whether the type of father was specified at time of data collection (i.e. whether these data	

						definitely relate to the biological and not the social father)	
History of psychiatric disorder (biological father)	psych_bf	Paternal history (biological father) of any type of psychiatric disorders before birth of the child. (Reported or diagnosed)	0) No 1) Yes		Binary	If there are no details of *any* psychiatric disorder but there are details of, for example, depression and anxiety, these can be included as "any" psychiatric disorder; this will then be only partial harmonisation, which should be detailed in full under "match" in the harmonisation table in the online catalogue State in the online catalogue whether the type of father was specified at time of data collection (i.e. whether these data definitely relate to the biological and not the social father)	
LIFESTYLE CHARACTERISTICS							
Paternal smoking during pregnancy	smk_p	Father smoked during pregnancy	0) No 1) Yes		Binary		
	smk_cig_p	Average number of cigarettes father smoked per day during pregnancy	0) None 1) < 10 per day 2) ≥ 10 per day		Categorical	1 cigar, pipe, cheroots etc. is equivalent to 3 cigarettes.	
Type of father, paternal smoking during pregnancy	smk_fath	Variable indicating whether paternal smoking during pregnancy variables (smk_p and smk_cig_p) are for the biological father, the social father or whether this is unknown	1) Biological father 2) Social father 3) Social mother 4) Unknown		Categorical		
CHILD							
BIRTH OUTCOMES							
Birth month	birth_month	Birth month of the index child	month (1 - 12)	calendar month	Integer		
Birth year	birth_year	Birth year of the index child	year	calendar year	Integer		
Apgar score	apgar	5-minute Apgar score	Score 1-10		Integer	10-minute Apgar score can be used if 5-minute Apgar score is not available; this will be partial harmonisation and should be detailed in the online catalogue 1-minute Apgar score cannot be used in place of 5-minute Apgar score	
Transferred to neonatal unit	neo_unit	Was the child transferred to neonatal unit?	0) No 1) Yes		Binary	Any transfer to a neonatal unit within the first week of life	
Sex	sex	Sex of the child	1) Male 2) Female		Binary		
Plurality	plurality	Number of foetuses in pregnancy	1) Single 2) Twin 3) ≥ Triplets		Categorical		
Gestational age (last menstrual period)	ga_lmp	Gestational age based on last menstrual period in days	Continuous in days (integer value)	days	Integer	If GA is reported in weeks then recalculate:	Gestational age <315 days

Gestational age (ultrasound)	ga_us	Gestational age based on ultrasound	Continuous in days (integer value)	days	Integer	n (weeks) x 7 + 3.5. Example: 40 weeks x 7 + 3.5 = 283.5	
Gestational age (Maternal report)	ga_mr	Gestational age based on maternal report	Continuous in days (integer value)	days	Integer		
Gestational age (Best judgement)	ga_bj	Best clinical judgement decided by cohorts estimating the most accurately measure of GA obtained within the cohort	Continuous in days (integer value)	days	Integer	Priority: 1) Use LMP if difference from US is less than 7 days or if US is not available 2) Difference between LMP and US more than 7 days --> use US 3) If LMP is not available use US 4) Non of them available - use maternal report.	
Birth weight	birth_weight	Weight of the child at birth		gm	Integer		
Birth length	birth_length	Length of the child at birth		cm	Decimal		
Birth head circumference	birth_head_circum	Circumference of the head at birth		cm	Decimal		
Size for gestational age	weight_who_ga	Weight of the child for gestational age at birth, based on WHO reference growth curves, categorised into small for gestational age (SGA), appropriate for gestational age (AGA) and large for gestational age (LGA).	1) SGA2) AGA3) LGA		Categorical	Defined using the WHO foetal growth charts (Kiserud et al 2017) using the 5th and 95th percentiles as cut-off values for SGA and LGA respectively. The relevant tables can be found in Kiserud *et al* (2017) (http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002220) and Appendix IV of this manual. Table 4 is the reference table for females and males born >= 40 completed weeks. Tables 14 and 15 are the reference tables for females and males (respectively) born <40 completed weeks. A child is classified as SGA if their birth weight is <= 5th percentile for their gestational age (in completed weeks).A child is classified as LGA if their birth weight is >= 95th percentile for their gestational age (in completed weeks).For e.g., a female born at 32 weeks + 2 days weighing 1,530g will be classified as SGA. A male born at 37 weeks + 5 days weighing 3,598 will be classified as LGA.	
Placenta weight	plac_weight	Weight of placenta at delivery	continuous in grams	gm	Integer		
Any congenital anomaly	con_anomalies	Any congenital malformation	0) No 1) Yes		Binary		
Severe congenital anomaly	major_con_anomalies	Major congenital malformations	0) No 1) Yes		Binary	According to EUROCAT guide 1.4, Table 3.3 (http://www.eurocat-network.eu/content/Full%20Guide%201%204%20version%2008_Sept2017.pdf).	

Cerebral palsy	cer_palsy	Any subtype of cerebral palsy	0) No 1) Yes		Binary		
Sibling position	sibling_pos	Previous live births from the mother regardless of whether they are living in the index child's household or not, including the index child. E.g. if index child is first born, then = 1	1) 1 2) 2 3) 3 4) 4 5) ≥ 5		Categorical		
HEALTH-RELATED CHARATERISTICS							
Death of child	death_child	Indicator variable for death of the child	0) No 1) Yes		Binary		
Age of death of child	death_child_age	Age of death of the child	Number of days alive indicating date of death in days	days	Integer	Only relevant for deceased children	
Childs height*	height_0 height_1 … height_215	Repeated measures of child's height: height_0: child's height measured between the ages of 0 and <1 month. **This does not include the child's birth length.** height_1: child's height measured between the ages of ≥1 and <2 months height_2: child's height measured between the ages of ≥2 and <3 months height_3: child's height measured between the ages of ≥3 and <4 months .... height_215: child's height measured between the ages of ≥215 and <216 months		cm	Decimal		
Age when child's height recorded	height_age0 height_age1 …. height_age215	Exact age of the child (in days) when child's height was recorded for height_0, height_1, height_2 etc.		days	Integer		
Childs weight*	weight_0 weight_1 … weight_215	Repeated measures of child's weight: weight_0: child's weight measured between the ages of 0 and <1 month. **This does not include the child's birth weight.** weight_1: child's weight measured between the ages of ≥1 and <2 months weight_2: child's weight measured between the ages of ≥2 and <3 months weight_3: child's weight measured between the ages of ≥3 and <4 months .... weight_215: child's weight measured between the ages of ≥215 and <216 months		kg	Decimal		

Age when child's weight recorded	weight_age0 weight_age1 …. weight_age215	Exact age of the child (in days) when child's weight was recorded for weight_0, weight_1, weight_2 etc.		days	Integer		
EXPOSURES/LIFESTYLE/ENVIRONMENT							
Exclusive breast-feeding	breastfed_excl	Total duration of exclusive breastfeeding (in the index child), in months	Duration in months	months	Decimal	Upper limit of 6 months; durations greater than 6 months assigned the value 6 months. Children never breastfed will have a duration of 0 months	
Any breastfeeding	breastfed_any	Total duration of any breastfeeding (in the index child), in months	Duration in months	months	Decimal	Upper limit of 12 months; durations greater than 12 months assigned the value 12 months. Children never breastfed will have a duration of 0 months	
Ever breastfed	breastfed_ever	Child ever breast fed	0) never breastfed 1) ever breastfed		Binary		
Solid food introduction	solid_food	Age of the child when solid food was introduced	Age in months	months	Decimal	Upper limit of 6 months; durations greater than 6 months assigned the value 6 months	
Child care introduction	childcare_intro	At what age he/she started in child care	Age in months	months	Decimal		
Child care	childcare_0 childcare_1 childcare_2 childcare_3	Child is cared for by other care givers other than parents of the child. childcare_0: child cared for by other care givers within 1st year of life (age range ≥0 and <1 year) childcare_1: child cared for by other care givers within 2nd year of life (age range ≥1 and <2 years) childcare_2: child cared for by other care givers within 3rd year of life (age range ≥2 and <3 years) childcare_3: child cared for by other care givers within 4th year of life (age range ≥3 and <4 years)	0) no childcare 1) childcare		Binary		

Cared for by relatives, friends, nanny or au pair	childcarerel_0 childcarerel_1 childcarerel_2 childcarerel_3	Child cared for by relatives, friends, nanny, babysitter or au pair. childcarerel_0: child cared for by friends, relatives etc. within 1st year of life (age range ≥0 and <1 year) childcarerel_1: child cared for by friends, relatives etc. within 2nd year of life (age range ≥1 and <2 years) childcarerel_2: child cared for by friends, relatives etc. within 3rd year of life (age range ≥2 and <3 years) childcarerel_3: child cared for by friends, relatives etc. within 4th year of life (age range ≥3 and <4 years)	0) not cared for by friends, relatives etc. 1) cared for by friends, relatives etc.		Binary		
Cared for by professional child minder	childcareprof_0 childcareprof_1 childcareprof_2 childcareprof_3	Child cared for by a professional child minder. childcareprof_0: child cared for by a child care professional within 1st year of life (age range ≥0 and <1 year) childcareprof_1: child cared for by a child care professional within 2nd year of life (age range ≥1 and <2 year) childcareprof_2: child cared for by a child care professional within 3rd year of life (age range ≥2 and <3 years) childcareprof_3: child cared for by a child care professional within 3rd year of life (age range ≥3 and <4 years)	0) not cared for by a professional child minder 1) cared for by a professional child minder		Binary		
Centre-based day care/nursery/kindergarten	childcarecentre_0 childcarecentre_1 childcarecentre_2 childcarecentre_3	Child attending a day care centre. childcarecentre_0: day care attendance within the first year of life (age range ≥0 and <1 year) childcarecentre_1: day care attendance within the 2nd year of life (age range ≥1 and <2 year) childcarecentre_2: day care attendance within the 3rd year of life (age range ≥2 and <3 years) childcarecentre_3: day care attendance within 4th year of life (age range ≥3 and <4 years)	0) not attending a day care centre 1) attending a day care centre		Categorical		

Child´s exposure to passive smoking	smk_exp0 smk_exp1 smk_exp2 … smk_exp17	Any exposure to smoking (mother smoking, biological father smoking, social father smoking, any smokers close to the child, or exposure to smoke in the home). smk_exp0: exposure to passive smoke from birth up to 1st birthday (age range ≥0 to <1year) smk_exp1: exposure to passive smoke from 1st birthday up to 2nd birthday (age range ≥1 to <2 years) smk_exp2: exposure to passive smoke from 2nd birthday up to 3rd birthday (age range ≥2 to <3 years) ... smk_exp17: exposure to passive smoke from 17th birthday up to 18th birthday (age range ≥17 to <18 years)	0) No 1) Yes		Binary		
Pets at child`s home	pets_0 pets_1 pets_2 … pets_17	Furry pet (dogs, cats, rodents) ownership in child's household. Repeated measures: pets_0: pets in the child's home from ≥0 to <1year. pets_1: From ≥1 year to < 2 years etc.	0) No 1) Yes		Binary		
Child´s exposure to parental mental disorders	mental_exp0 mental_exp1 mental_exp2 … mental_exp17	Any exposure to relatives with **any** mental disorders (mother, biological father, social father, or any individuals close to the child)mental_exp0: exposure to mental disordered relatives from birth up to 1st birthday (age range ≥0 to <1year) mental_exp1: exposure to mental disordered relatives from 1st birthday up to 2nd birthday (age range ≥1 to <2 years) mental_exp2: exposure to mental disordered relatives from 2nd birthday up to 3rd birthday (age range ≥2 to <3 years) ... mental_exp17: exposure to mental disordered relatives from 17th birthday up to 18th birthday (age range ≥17 to <18 years)	0) No 1) Yes		Binary	If there are no details of "*any*" psychiatric disorder but there are details of, for example, depression and anxiety, these can be included as "any" psychiatric disorder; this will then be partial harmonisation, which should be detailed in full under "match" in the harmonisation table in the online catalogue. Visits to a psychologists or psychiatrist should only be used as an indication of a psychiatric disorder if no other data are available. This being the case, the variable will be partially harmonised (which should be detailed in full in the online catalogue).	
HOUSEHOLD CHARACTERISTICS							

Household income (core)	hhincome_0 hhincome_1 hhincome_2 ... hhincome_17	Total yearly income of the household categorised into quartiles of low, medium-low, medium-high and high income levels based on the national yearly household income distribution. If more than household income is recorded within the defined time frame, use highest recorded household income Repeated measures: hhincome_0: highest reported household income within one year of birth (child aged between >-1 year and <1 year) hhincome_1: highest reported household income when the child was aged between ≥1 and <2 years hhincome_2 highest reported household income when the child was aged between ≥2 and <3 years hhincome_3: highest reported household income when the child was aged between ≥3 and <4 years .... hhincome_17: highest reported household income when the child was aged between ≥17 and <18 years	1) 4th quartile (highest) 2) 3rd quartile 3) 2nd quartile 4) 1st quartile		Categorical	Income categorised into quartiles (low, medium-low, medium-high, high) based on the national household income distribution in the year of follow-up.	
Family split up	fam_splitup0 fam_splitup1 fam_splitup2 …. fam_splitup17	Was the child exposed to a split up during the 1st year of life (age range ≥0 to <1 year), 2nd year of life (age range ≥1 to <2 years) etc.	0) No split up1) Split up		Binary	Any recorded split up. fam_splitup0 indicates that the index child was exposed to a split up when the child was ≥0 to <1 year fam_splitup1: From ≥1 year to < 2 years etc. If the mother was originally single (no partner) and remains single, fam_splitupX = 0 (no split up)	

Family size	famsize_child0 famsize_child1 famsize_child2 … famsize_child17	Total number of children in the household under 18 years old **including index child** (All children; biological, half-siblings, adopted, fostered etc.) Repeated measures: famsize_child0: maximum recorded family size in 1st year of life (≥0 to <1 year) famsize_child1: maximum recorded family size in 2nd year of life (≥1 to <2 years) famsize_child2: maximum recorded family size in 3rd year of life (≥2 to <3 years) … famsize_child17: maximum recorded family size in 18th year of life (≥17 to <18 years)	Total number of children	children	Integer	If the cohort has a different definition for child (e.g. <14 years), harmonisation is still possible, the variable will be a partial match (this should be detailed under the match in the catalogue) Household where the child lives the majority of the time	
	famsize_adult0 famsize_adult1 famsize_adult2 … famsize_adult17	Total numbers of adults in the same household. Preferably from same time point as when hh income recorded Repeated measures: famsize_adult0: number of adults in same hh in 1st year of life (≥0 to <1 year), preferably from same time point as when hh income recorded famsize_adult1: number of adults in same hh in 2nd year of life (≥1 to <2 years), preferably from same time point as when hh income recorded famsize_adult2: number of adults in same hh in 3rd year of life (≥2 to <3 years), preferably from same time point as when hh income recorded … famsize_adult17: number of adults in same hh in 18th year of life (≥17 to <18 years)	Total number of adults	adults	Integer		

Table 3. The Core Variables Table. The core variables table provides a description of the LifeCycle core variables: their name, description and instructions for how to derive the harmonised LifeCycle variables.

3. EXAMPLE HARMONISATIONS

3.1 Example of complete harmonisation using data from the DNBC

A. Definition of the harmonised LifeCycle variable (history of maternal asthma):

	Variable name	Label/description	Values	Unit	Data Type	Comments	Further Instructions
MATERNAL CHARACTERISTICS							
History of asthma	asthma_m	Maternal history of asthma before pregnancy (of index child)	0) No 1) Yes		Binary	Where data are available, asthma should be doctor diagnosed. If no information is available on doctor diagnosis, the variable is partially harmonised.	

B. Data (source variables): DNBC, interview 1:

A052. Did you ever have asthma?

1. yes
2. no ->A055
3. do not know->A055
4. do not want to answer->A055

A053. Was the asthma diagnosed by a doctor?

1. yes
2. no ->A055
3. do not know->A055
4. do not want to answer ->A055

C. Harmonisation: description

asthma_m = 1 if a053 = 1
asthma_m = 0 if a053 = 2 or 3 OR a052 = 2 or 3
asthma_m = missing otherwise

COMPLETE HARMONISATION: the harmonised variable matches the description provided in core variable table.

3.2 Example of partial harmonisation using data from the DNBC

A. Definition of the harmonised LifeCycle variable (fever during pregnancy):

	Variable name	Label/description	Values	Unit	Data Type	Comments	Further Instructions
MATERNAL CHARACTERISTICS							
Fever during pregnancy	preg_fever	Any fever during pregnancy	0) No 1) Yes		Binary		The variable is fully harmonised if data cover the whole pregnancy

B. Data (source variables): DNBC, interviews 1 (≈ gest. week 17) & 2 (≈ gest. week 30):

A100. Have you had any episode of fever during your pregnancy? [Asked around gest. week 17]

1. yes
2. no ->A106
3. do not know ->A106
4. do not want to answer->A106

B110. Have you had fever while pregnant? [Asked around gest. week 30]

1. yes
2. yes, but not since the last interview ->B116
3. no ->B116
4. do not know ->B116
5. do not wish to answer ->B116

C. Harmonisation: description

preg_fever = 1 if a100 = 1 OR b110 = 1 or 2
preg_fever = 0 if (a100 = 2 AND b110 = 3) OR (a100 = 2 AND b110 = missing) OR (a100 = missing AND b110=3)
preg_fever = missing otherwise

PARTIAL HARMONISATION: there is no information on fever during the last part of pregnancy, therefore the variable is partially harmonised.

3.3 Example of converting a categorical variable to a continuous variable (partial harmonisation) using data from Generation R

A. Definition of the harmonised LifeCycle variable (age at solid food introduction):

	Variable name	Label/description	Values	Unit	Data Type	Comments	Further Instructions
CHILD							
Solid food introduction	solid_food	Age of the child when solid food was introduced	Age in months	months	Decimal	Upper limit of 6 months; durations greater than 6 months assigned the value 6 months	

B. Data (source variables): Generation R:

intro_groep2. Age at solid food introduction

1. < 3 months
2. 3-6 months
3. > 6 months

C. Harmonisation: description

The categorical variable is converted to a continuous variable by taking the mid-point of each category:
< 3 months = 1.5 months
3 – 6 months = 4.5 months
> 6 months = 6 (the harmonised LifeCycle variable has an upper limit of 6 months

PARTIAL HARMONISATION: only a crude approximation to the LifeCycle variable can be made, therefore the variable is partially harmonised.

3.4 Example of converting from one categorical variable to another categorical variable (partial harmonisation) using data from Generation R

A. Definition of the harmonised LifeCycle variable (pre-pregnancy alcohol intake):

	Variable name	Label/ description	Values	Unit	Data Type	Comments	Further Instructions
MATERNAL CHARACTERISTICS:							
Pre-pregnancy alcohol intake	prepreg_alc_unit	Average units of alcohol mother drank per week before pregnancy	0) 0 (Never drinking alcohol) 1) Light (< 3 units) 2) Moderate (≥3 and <7 units) 3) Moderate to heavy (≥7 - <14 units) 4) Heavy (≥14 units)		Categorical	Alcohol consumption before pregnancy includes ≥ 14 units per week. During pregnancy this category is not included. For converting from days to weeks apply the following rule: one drink per day corresponds to 7–13 drinks weekly, two drinks per day to 14–20 drinks weekly, etc. Non-drinkers categorised as 0 - never drinking alcohol	

B. Data (source variables): Generation R:

F0500101. Pre-pregnancy alcohol intake

1. none
2. <1 per week
3. 1-3 per week
4. 4-6 per week
5. 1 glass per day
6. 1-3 glasses per day
7. >3 glasses per day

C. Harmonisation: description

prepreg_alc_unit = 0 if F0500101 = 1

prepreg_alc_unit = 1 if F0500101 = 2 or 3

prepreg_alc_unit = 2 if F0500101 = 4

prepreg_alc_unit = 3 if F0500101 = 5

prepreg_alc_unit = 4 if F0500101 = 6 or 7

PARTIAL HARMONISATION: the categories match only partially

4. APPENDIX

I. Online Catalogue Templates

cohort	variable	description	values	unit	datatype	collectionType	dependencies	dateOfUpdate	id	cohortLabel
	The source variable's name	*A brief description of the source variable*	*Details of the categories for categorical and binary variables*	*The units for continuous variables*	*E.g. binary, categorical, integer or decimal*	*How the data were collected*	*Dependencies of the variable, for e.g. within a questionnaire*	*Date the variable was last updated*	*Automatically asigned*	*Automatically asigned*

Table 4. An outline of the Excel template for the source variables descriptions

Cohort	Source variables	Harmonized variable	Description	Syntax	Match	ID	Harmonized variable	Source variables	Date of update
	A list of the source variables	*The harmonised, LifeCycle variable's name*	*The description of harmonisation*	*The syntax used to generate the variable*	*Whether the variable was fully or partially harmonised*	*Automatically asigned*	*The harmonised, LifeCycle variable's label*		*Date the variable was last updated*

Table 5. An outline of the Excel template for the harmonisation descriptions

II. Year, month, week, day conversions

1 year = 12 months

1 year = 52.1775 weeks

1 year = 365.2422 days

1 month = 0.0833 years

1 month = 4.3481 weeks

1 month = 30.4368 days

1 week = 0.0192 years

1 week = 0.2300 months

1 week = 7 days

1 day = 0.0027 years

1 day =0.0329 months

1 day = 0.1429 weeks

III. Imperial to metric conversions

1 inch = 2.54 centimetres (cm)

1 inch = 0.0254 metres (m)

1 foot (ft) = 0.3048 metres (m)

1 pound (lb) = 0.4536 kilograms (kg)

1 stone (st) = 6.3503 kilograms (kg)

IV. Size for Gestational Age: WHO foetal growth charts

The variable weight_who_ga categorises children into three groups: small (SGA), acceptable (AGA) and large for gestational age (LGA). The 5th and 95th percentiles from the WHO foetal growth charts (Kiserud *et al*. 2017 PLoS Med. 2017 Jan 24;14(1):e1002220) are used as cut-off values for SGA and LGA respectively.

A child is classified as SGA if their birth weight is <= 5th percentile for their gestational age (in completed weeks). A child is classified as LGA if their birth weight is >= 95th percentile for their gestational age (in completed weeks).

Table 4 from Kiserud *et al* 2017 is the reference table for females and males born >= 40 completed weeks. Tables 14 and 15 from Kiserud *et al* 2017 are the reference tables for females and males (respectively) born <40 completed weeks.

The full paper by Kiserud *et al* 2017 can be found here:
http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002220

The relevant tables are provided below.

Percentile	Birthweight (g) by Gestational Age (wk)											
	Female						Male					
	37	38	39	40	41	42	37	38	39	40	41	42
5	1,968	2,315	2,575	2,748	2,835	2,834	2,062	2,451	2,723	2,880	2,921	2,845
25	2,493	2,698	2,891	3,072	3,241	3,398	2,705	2,890	3,061	3,218	3,362	3,491
50	2,786	2,990	3,173	3,336	3,479	3,601	2,919	3,153	3,354	3,519	3,650	3,747
75	2,951	3,217	3,443	3,631	3,779	3,888	3,143	3,387	3,608	3,806	3,982	4,134
90	3,181	3,451	3,682	3,871	4,021	4,130	3,450	3,666	3,871	4,067	4,253	4,428
95	3,238	3,593	3,867	4,060	4,171	4,200	3,584	3,813	4,036	4,251	4,459	4,659

doi:10.1371/journal.pmed.1002220.t004

Table 6. Table 4 from Kiserud *et al* 2017: estimated birthweight percentiles for female and male neonates according to completed gestational week. This is the reference table for females and males born >= 40 completed weeks.

Gestational Age (Weeks)	Female Estimated Fetal Weight (g) by Percentile						
	5	10	25	50	75	90	95
14	73	77	82	89	96	102	107
15	92	97	104	113	121	129	135
16	116	122	131	141	152	162	170
17	145	152	164	176	189	202	211
18	180	188	202	217	233	248	261
19	221	231	248	266	285	304	319
20	269	281	302	322	346	369	387
21	324	339	364	388	417	444	466
22	388	405	435	464	499	530	557
23	461	481	516	551	592	629	660
24	542	567	608	649	697	740	776
25	634	663	710	758	815	865	907
26	735	769	823	880	946	1,003	1,051
27	846	886	948	1,014	1,090	1,156	1,210
28	967	1,013	1,083	1,160	1,247	1,323	1,383
29	1,096	1,150	1,230	1,319	1,418	1,505	1,570
30	1,234	1,296	1,386	1,489	1,601	1,699	1,770
31	1,379	1,451	1,553	1,670	1,796	1,907	1,984
32	1,530	1,614	1,728	1,861	2,002	2,127	2,209
33	1,687	1,783	1,911	2,060	2,217	2,358	2,445
34	1,847	1,957	2,101	2,268	2,440	2,598	2,690
35	2,008	2,135	2,296	2,481	2,669	2,846	2,943
36	2,169	2,314	2,494	2,698	2,902	3,099	3,201
37	2,329	2,493	2,695	2,917	3,138	3,357	3,462
38	2,484	2,670	2,896	3,136	3,373	3,616	3,725
39	2,633	2,843	3,096	3,354	3,605	3,875	3,988
40	2,775	3,010	3,294	3,567	3,832	4,131	4,247

doi:10.1371/journal.pmed.1002220.t014

Table 7. Table 14 from Kiserud *et al* 2017: growth chart for estimated fetal weight for female foetuses.
This is the growth reference table for females born <40 completed weeks.

Gestational Age (Weeks)	Male Estimated Fetal Weight (g) by Percentile						
	5	10	25	50	75	90	95
14	75	79	84	92	99	105	109
15	96	100	107	116	126	134	139
16	121	127	136	146	158	169	175
17	152	158	170	183	197	210	219
18	188	196	210	226	243	260	271
19	232	241	258	277	298	320	333
20	282	293	314	337	362	389	405
21	341	354	380	407	436	469	489
22	408	424	454	487	522	561	586
23	484	503	539	578	619	666	695
24	570	592	635	681	730	785	818
25	666	692	742	795	853	917	956
26	772	803	860	923	990	1,063	1,109
27	888	924	989	1,063	1,141	1,224	1,276
28	1,014	1,055	1,129	1,215	1,305	1,399	1,458
29	1,149	1,197	1,281	1,379	1,482	1,587	1,654
30	1,293	1,349	1,442	1,555	1,672	1,788	1,863
31	1,445	1,509	1,613	1,741	1,874	2,000	2,085
32	1,605	1,677	1,793	1,937	2,085	2,224	2,319
33	1,770	1,852	1,980	2,140	2,306	2,456	2,562
34	1,941	2,032	2,174	2,350	2,534	2,694	2,814
35	2,114	2,217	2,372	2,565	2,767	2,938	3,072
36	2,290	2,404	2,574	2,783	3,002	3,185	3,334
37	2,466	2,591	2,777	3,001	3,238	3,432	3,598
38	2,641	2,778	2,981	3,218	3,472	3,676	3,863
39	2,813	2,962	3,183	3,432	3,701	3,916	4,125
40	2,981	3,142	3,382	3,639	3,923	4,149	4,383

doi:10.1371/journal.pmed.1002220.t015

Table 8. Table 15 from Kiserud *et al* 2017: growth chart for estimated foetal weight for male foetuses.
This is the reference table for males born <40 completed weeks.

V. Link to EUROCAT Guide 1.4

Where ICD-10 codes are available, major congenital anomalies should be identified using the EUROCAT guide 1.4. The full document can be found here: http://www.eurocat-network.eu/content/Full%20Guide%201%204%20version%2008_Sept2017.pdf.

Table 3.3 (EUROCAT Subgroups of Congenital Anomalies) from EUROCAT guide 1.4 should be used to identify major anomalies; **all minor anomalies should be excluded when deriving the harmonised LifeCycle variable for major congenital anomalies.**

Guide 1.4 uses ICD10-BPA codes only. Table 3.3 from EUROCAT guide 1.4 also provides the ICD9-BPA codes and the minor anomalies pre-2005 for retrospectively making subgroups pre-2005 when this coding system was used.

VI. Stata syntax for placing growth measurements into monthly age intervals

This syntax is provided courtesy of SWS; it will need adapting to your cohort.

```
***Put ages into months for each wave of SWS and round them down to give the index for
the measurements
***(6m - prefix k, 12m - prefix l, 2y - prefix m, 3y -prefix n, 4y - prefix r, 6-7y -
prefix s, 8-9y - prefix t)

gen byte kagem = int(kage*0.23)
gen byte lagem = int(lage*0.23)
gen byte magem = int(mage*12)
gen byte nagem = int(nage*12)
gen byte ragem = int(rage*12)
gen byte sagem = int(sage*12)
gen byte tagem = int(tage*12)

***Put ages into days

gen kaged = round(kage*7)
gen laged = round(lage*7)
gen maged = round(mage*365.2422)
gen naged = round(nage*365.2422)
gen raged = round(rage*365.2422)
gen saged = round(sage*365.2422)
gen taged = round(tage*365.2422)


foreach num of numlist 0 / 215 {
gen height_age_`num' = .
gen weight_age_`num' = .
gen height_`num' = .
gen weight_`num' = .
}

capture program drop htwtage
program define htwtage
***Arguments are actual age in days, integer age month (index value), height and weight
args age agem height weight
foreach num of numlist 0 / 215 {
  replace height_age_`num' = `age' if `agem' == `num'
  replace height_`num' = `height' if `agem' == `num'
  replace weight_age_`num' = `age' if `agem' == `num'
  replace weight_`num' = `weight' if `agem' == `num'
  replace height_age_`num' = . if height_`num' ==.
  replace weight_age_`num' = . if weight_`num' ==.
    }
end

**Run the program for each wave of SWS data collection

htwtage kaged kagem kcrhl kwtkg
htwtage laged lagem lcrhl lwtkg
htwtage maged magem mht mwtkg
htwtage naged nagem nht nwtkg
htwtage raged ragem rhtcm rwtkg
htwtage saged sagem sht swtkg
htwtage taged tagem tht twtkg
```

LifeCycle

Instructions for Quality Control

of Harmonized Core Variables in LifeCycle

Content

Rationale

As the final part of harmonizing the core data in Work Package 1 of LifeCycle, the list of harmonized core variables needs a local validation. In this document, instructions for such local quality control are described.

Accordingly, each cohort in LifeCycle needs to perform quality checks to validate and evaluate their harmonization of the core variables, and align this information with that documented in the *Online Catalogue:* https://molgenis88.gcc.rug.nl/. The Quality Control must be in accordance with the *Core Variables List,* which has just been updated July 2019.

To this end, we ask each cohort to carefully read these instructions for a local validation, and follow and run the syntax from the Stata/SAS/R code to replicate the quality control on their own cohort data.

The aim of this Quality Control is to firstly check that each cohort's harmonized core variables match those described in the Core Variable list, and secondly ensure the quality of harmonization. If any unmatched or errors are found these should be amended accordingly. Each cohort should carry out the quality checks and any necessary corrections themselves.

If you have any queries about the quality control of the harmonized core variables please make contact to Johan Lerbech Vinther (johan.vinther@sund.ku.dk) or Angela Pinot de Moira (anpi@sund.ku.dk) from WP1.

An introduction to the quality control of the harmonized core variables was presented at the General Assembly Meeting in Copenhagen in May 2019. Slides from this presentation can be found at the LifeCycle Intranet.

Deadline for finalizing the Quality Control Checks is August 15th 2019.

Step 1: Verify list of variables and formats

Please, verify that your cohort-specific harmonized core variables completely match with the information provided in the Core Variables List.

Each cohort needs to check that the *name* and *data type* of each of the core variables correspond exactly to the Core Variable List (please, see table below). Also, please check the requirements for *type of harmonization*. Variables considered 'fully harmonised' must match the information provided in the comment's section. **Please, see the following examples of Step 1 in the Quality Control**:

Example: History of Asthma

Variable	Variable name	Label/description	Values	Unit	Data Type	Comments
History of asthma	asthma_m	Maternal history of asthma before pregnancy (of index child)	0) No 1) Yes		Binary	Where data are available, asthma should be doctor diagnosed. If no information is available on doctor diagnosis, the variable is partially harmonised.

Core Variable List

1. Check the variable name ('asthma_m')
2. Check that the numbers of categories match (two)
3. Check 'comments':
    a. Full Harmonization: if asthma is doctor diagnosed
    b. Partial Harmonization: if asthma is not doctor diagnosed

Example: Apgar Score

Variable	Variable name	Label/description	Values	Unit	Data Type	Comments
Apgar score	apgar	5-minute Apgar score	Score 1-10		Integer	10-minute Apgar score can be used if 5-minute Apgar score is not available; this will be partial harmonisation and should be detailed in the online catalogue 1-minute Apgar score cannot be used in place of 5-minute Apgar score

Core Variable List

1. Check the variable name ('apgar')
2. Check that values are integer
3. Check 'comments':
    a. Full Harmonization: if Apgar score is based on a 5-minute Apgar test.
    b. Partial Harmonization: if Apgar score is based on a 10-minute Apgar test.

If any mismatch is observed, please correct the errors accordingly.

Step 2: Check univariate distributions

Please, generate distributions for all core variables, and check for outliers and improbable values. Also, for variables that have been reported in papers/publications, verify that distributions or summary statistics of the harmonized LifeCycle variable match those of the reported/published variables.

For continuous variables, check that there are no outliers, i.e. values out of the minimum and maximum range defined in the Core Variable List.

Please, use your scientific knowledge and practical sense when making this quality check so as not to drop interesting outliers. Errors are probably caused by an error in your harmonization script, so please check this carefully and correct where required. In the code, we have highlighted guidelines and tips for specific variables of interest to help you.

Example: Mothers Age at Birth in years (agebirth_m_y)

For this variable, it is relevant to check for unusual values of mother's age at birth. A valid argument is that values outside the range 15-50 years need further considerations, especially if values are substantially low.

Variable	Variable name	Label/description	Values	Unit	Data Type
Maternal age at birth (years)	agebirth_m_y	Mother's age at delivery in complete years.	Continuous in years	years	Integer

Core Variables List

A concrete example (please, see output below) is mother's age at child's birth in the Danish National Birth Cohort (DNBC) where age range is 15-47 years, which seems fairly usual; hence no amendment is needed.

```
. sum agebirth_m_y

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
agebirth_m_y |    102,442    29.97417    4.348355        15         47
```

Output from STATA

For categorical variables, check that there are no improbable values, i.e. values not corresponding to the categories defined in the Core Variable List. Please, correct errors where relevant. In the code, we have highlighted the number and type of categories for each of the core variables to help you.

Example: History of diabetes (prepreg_dia)

For History of Diabetes, there are five possible categories following the Core Variable List :

Variable	Variable name	Label/description	Values	Unit	Data Type
History of diabetes (all kinds)	prepreg_dia	Diagnosis of mother with diabetes before index pregnancy	1) No 2) Type I 3) Type II 4) Gestational diabetes 5) Other/unspecified		Categorical

Core Variables List

Check that data is coded into a maximum of five categories, and falls into the categories 1-5 (see output):

```
. tab prepreg_dia

prepreg_dia |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    101,962       99.45       99.45
          2 |        292        0.28       99.73
          3 |         34        0.03       99.77
          4 |        233        0.23       99.99
          5 |          6        0.01      100.00
------------+-----------------------------------
```

Output from STATA

Example: Birth outcome (outcome)

For Birth Outcome, there are six possible categories following the Harmonization Protocol:

Variable	Variable name	Label/description	Values	Unit	Data Type	Comments
Birth outcome	outcome	The child´s condition at delivery	1) Live-born 2) Stillborn 3) Spontaneous abortion 4) Induced abortion 5) Unspecified abortion 6) Other (e.g. Molar + extrauterine pregnancy)		Categorical	A complete match is achieved when data on all categories (live born, still born, spontaneous abortion etc.) are available. When detailing the match of the variable in the catalogue it is important to include the approximate age of recruitment; in many cases this will explain a partial match. Where the match is only partial, also indicate whether there are complete data on live-born vs. stillborn. Stillborn defined according to WHO recommendations (22 completed weeks)

Core Variables List

Check that data is coded into a maximum of six categories. It might be that you don't have data in all six categories, but simply check that data falls into the categories 1-6 (see Stata-output below):

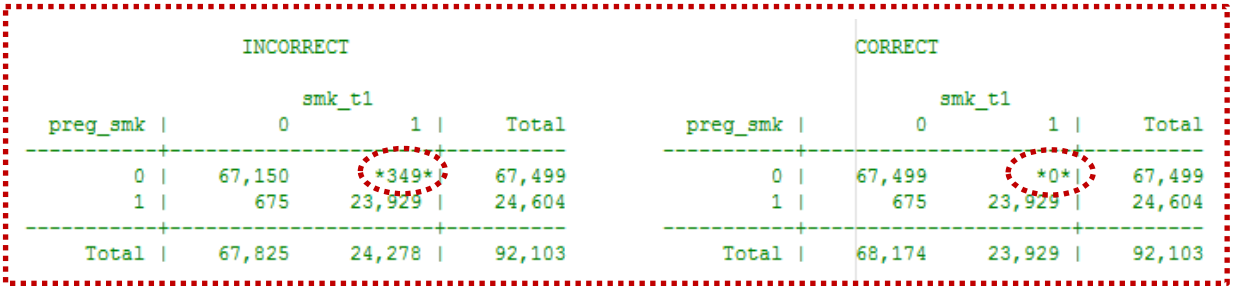
```
. tab outcome, nolabel

 RECODE of
  outcom_f
 (Udfald af
graviditet) |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |     96,834       94.51       94.51
          2 |        329        0.32       94.83
          3 |      4,739        4.63       99.46
          4 |        443        0.43       99.89
          6 |        114        0.11      100.00
------------+-----------------------------------
```

Output from STATA

Step 3: Check internal validity

Internal validation is an important part of the local quality control. Thus, <u>within reason</u> cross-tabulate the variables against other variables to check for consistency.

A non-related example yet perfectly illustrating internal validity-check is *prostate cancer and sex*. As prostate cancer only occur in men, cross-tabulating the two variables must only show values in men.
In the code, we have highlighted cross-tabulations for the variables where internal validation must be checked. To further assist you, we have visualized the examples showing when amendment is needed (see example below left) and when no amendment is needed (see example below right), respectively.



INCORRECT

preg_smk	smk_t1 0	smk_t1 1	Total
0	67,150	*349*	67,499
1	675	23,929	24,604
Total	67,825	24,278	92,103

CORRECT

preg_smk	smk_t1 0	smk_t1 1	Total
0	67,499	*0*	67,499
1	675	23,929	24,604
Total	68,174	23,929	92,103

<u>Example:</u> <u>Maternal Smoking during Pregnancy (preg_smk)</u>
A fine example used to check the internal validity is by cross-tabulating *maternal smoking during pregnancy* (preg_smk) and *average number of cigarettes smoked per day during pregnancy* (preg_cig):

Variable	Variable name	Label/description	Values	Unit	Data Type	Comments
Maternal smoking during pregnancy	preg_smk	Any smoking during pregnancy	0) No 1) Yes		Binary	
	preg_cig	Average number of cigarettes smoked per day during pregnancy	0) None 1) < 10 per day 2) ≥ 10 per day		Categorical	Cigars, pipes, cheroots etc. converted into cigarettes; see above. Non-smokers categorised under 0 - none

Mother's not smoking during pregnancy (preg_smk=0) are smoking no cigarettes (none) on average per day during pregnancy (preg_cig=0). Thus, for preg_smk=0, preg_cig=1 and preg_cig=2 the values should have no observations, as illustrated in the Stata output below (marked with red circles).

```
. tab preg_smk preg_cig
```

preg_smk	RECODE of __00000U 0	1	2	Total
0	70,498	0	0	70,498
1	0	16,868	8,152	25,020
Total	70,498	16,868	8,152	95,518

Please, amend any errors accordingly.

Step 4: Check consistency in repeated measures

In the Core Variable List, we have monthly and yearly repeated measures. Quality checking the repeated measures includes checking the consistency in time bands and order.

Check that repeated measures has the correct number of time bands

In LifeCycle, each cohort must upload all core variables to OPAL, incl. all time bands for the repeated measures, having the data or not. This is important as data cannot run in OPAL if variables are missing!
Thus, it is important for each cohort to check the existence of every repeated measure proposed in the Core Variable List. In the code, a script will check this.

Example: Cohabitation, e.g. cohab_0 (18 time bands, 0-17 years)

```
* Please, check for correct number of time bands [eightteen, 0-17]
foreach n of numlist 0 / 17 {
tab cohab_`n'
}
```

Example: Height, e.g. height_0 (216 time bands, 0-215 months)

```
* Please, check for correct number of time bands [twohundredandsixteen, 0-215]
foreach n of numlist 0 / 215 {
tab height_`n'
}
```

Example: Childcare, e.g. childcare_0 (4 time bands, 0-3 years)

```
* Please, check for correct number of time bands [four, 0-3]
foreach n of numlist 0 / 3 {
tab childcare_`n'
}
```

Check that repeated measures are placed in the correct time band

It is crucial that measures of repeated variables are placed in the correct time band. The logic of this quality check is the same for each cohort, however as each of the repeated variables are not measured exactly at the same time in each cohort, the cohort-specific quality control (in Stata, SAS or R) depends on date of collection within a cohort.

Example: Cohabitation in Danish National Birth Cohort (DNBC)

In DNBC, cohabitation is measured at 6 month follow-up (age_6m). Date of collection differed among the 70,000 subjects within the cohort, as age of the children when data were collected ranged from 0-2 years of age. Thus, in DNBC we have data for cohabitation in three time bands (cohab_0, cohab_1, cohab_2).

In theory, we check that in *cohab_0* only children at 0 years of age (at date of collection) are included. Likewise, in *cohab_1* that only children at 1 year of age (at date of collection), and in *cohab_2* there is only children at 2 years of age (at date of collection).

In practice, we check the consistency by cross-tabulating *cohabitation* (cohab_0, cohab_1 and cohab_2, respectively) with date of collection (age_6m). **See below for the example with cohab_0 and age_6m**:

INCORRECT

cohab_0	age_6m 0	1	2	Total
1	68,533	*63*	*1*	68,597
2	1,677	*2*	*0*	1,679
Total	70,210	*65*	*1*	70,276

CORRECT

cohab_0	age_6m 0	1	2	Total
1	68,533	*0*	*0*	68,533
2	1,677	*0*	*0*	1,677
Total	70,210	*0*	*0*	70,210

Left table (before qualify check):

A total of 66 children (shown with *) are mistakenly placed in *cohab_0*. In fact, 65 children had cohabitation measured at 1 year of age (at date of collection), while one child had cohabitation measured at the 2 years of age (at date of collection). These errors should be corrected accordingly.

Right table (after quality check)

In cohab_0 only children at 0 year of age (at date of collection) is placed in cohab_0. Children aged 1 and 2 (at date of collection) are placed in cohab_1 and cohab_2, respectively.

This step of the quality check in the repeated measures must be replicated for all monthly and yearly repeated measures, and importantly, within each measure for all the specific dates of collection!

Check that (specific) repeated measures follow a logical order

For a few specific repeated measures, the quality control involves checking that the variables follow a logical order being repeated, i.e. that if they change over time it only occurs in one specific direction. This is the case for maternal education (edu_m_) and child's height (height_), which values can only increase.

Example: Maternal Education (edu_m):

Is it impossible for a mother to have high level of education at year 0 (edu_m_0=1) and a medium level of education at year 5 (edu_m_5=2). Likewise, a mother can't have a medium level of education at year 10 (edu_m_10=2) and a low level of education at year 11 (edu_m_11=3).

In the code, a generic code check for consistency in maternal education (edu_m) over time

```
* NB! For this variable, it is important to check whether ID change category when not reasonable
findit missings // install this function in STATA (select the first listed option)
missings dropvars edu_m_0 - edu_m_17 //deletes variables with only missing data
keep child_id - coh_country edu_m_0 - edu_m_17 sex // keeps education variables
reshape long edu_m_, i(child_id) j(month) // reshapes to long

drop if edu_m_==.
by child_id, sort: gen n=_n

preserve
foreach n of numlist 0 / 17 {
bysort child_id (month), sort: gen diff_`n' = edu_m_[1] > edu_m_[_N]
list child_id edu_m_ month if diff_`n'==1
drop if n==`n'
}
```

Example: Child's height (height_):

It is impossible for a child to measure 50 cm at year 1 (height_1=50) and 45 cm at year 2 (height_2=45). Likewise, a child's height can't be 80 cm at year 5 (height_5=80) and 65 cm at year 10 (height_10=65).

In the code, a generic code check for consistency in child's height (height_) over time

```
* NB! For this variable, it is important to check whether ID change category when not reasonable
findit missings // install this function in STATA (select the first listed option)
missings dropvars height_0 - height_215 //deletes variables with only missing data
keep child_id - coh_country height_0 - height_180 sex // keeps height variables
reshape long height_, i(child_id) j(month) // reshapes to long

drop if height_==.
by child_id, sort: gen n=_n

preserve
foreach n of numlist 1 / 215 {
bysort child_id (month), sort: gen diff_`n' = height_[1] > height_[_N]
list child_id height_ month if diff_`n'==1
drop if n==`n'
}
```

In the case that either changes in level of education or height does not follow a logical order, please make thorough actions and correct the errors where required.

Step 5: Complete the Online Catalogue

As part of the Quality Control, each cohort must ensure that the Online Catalogue is aligned accordingly. This means that in the Online Catalogue description of harmonisation is complete and information in all three tabs ('description, 'variables used and script syntax' are completed in full.

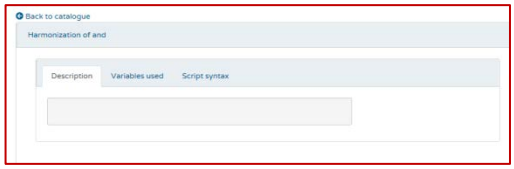
The checklist below is intended to help cohorts with this alignment process. It is important that each cohort check the list step-by-step to qualify the core fundament of The LifeCycle Project: THE DATA!

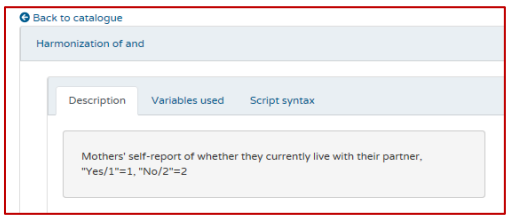
Checklist to align the Online Catalogue with the Quality Control

Each cohort has the responsibility to ensure that all cohort-specific information is correctly listed in the Online Catalogue. The Online Catalogue is a vital platform for future use of the LifeCycle data, hence it is very important that anyone outside the cohorts can understand the information reported and described in the Online Catalogue and are able to use the data provided by each cohort.

Check and verify 'Description'

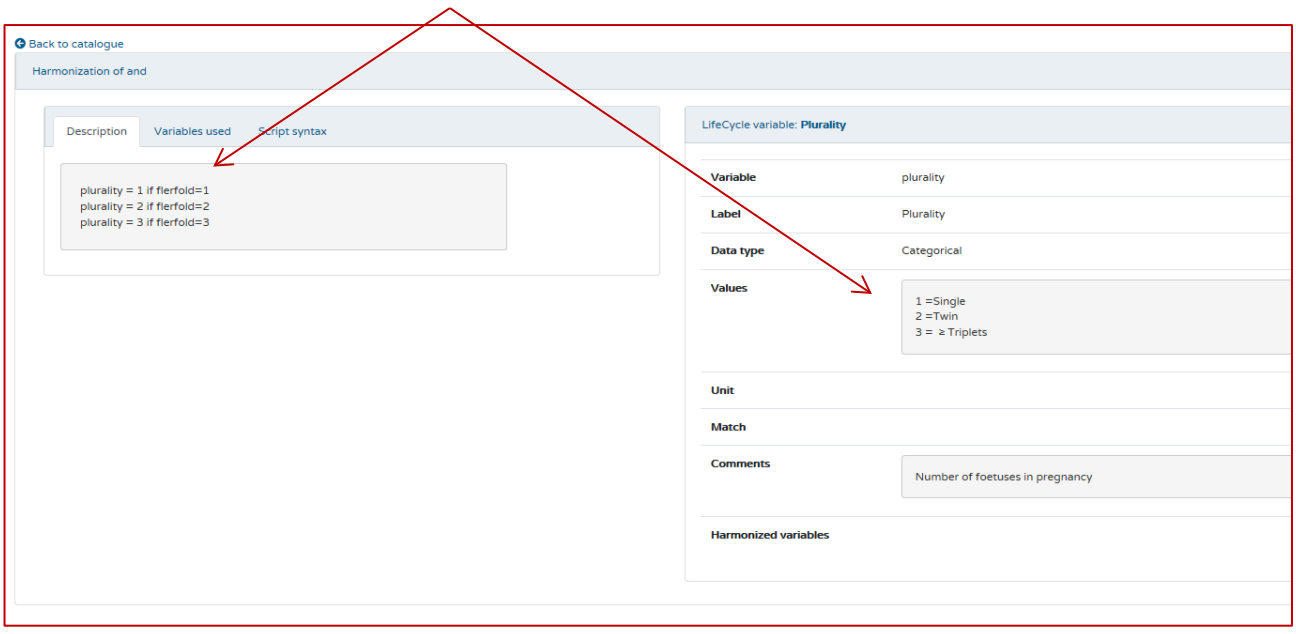
Ensure that a description is filled in – if no description is filled in, please fill in (see picture on the right for inadequate example)





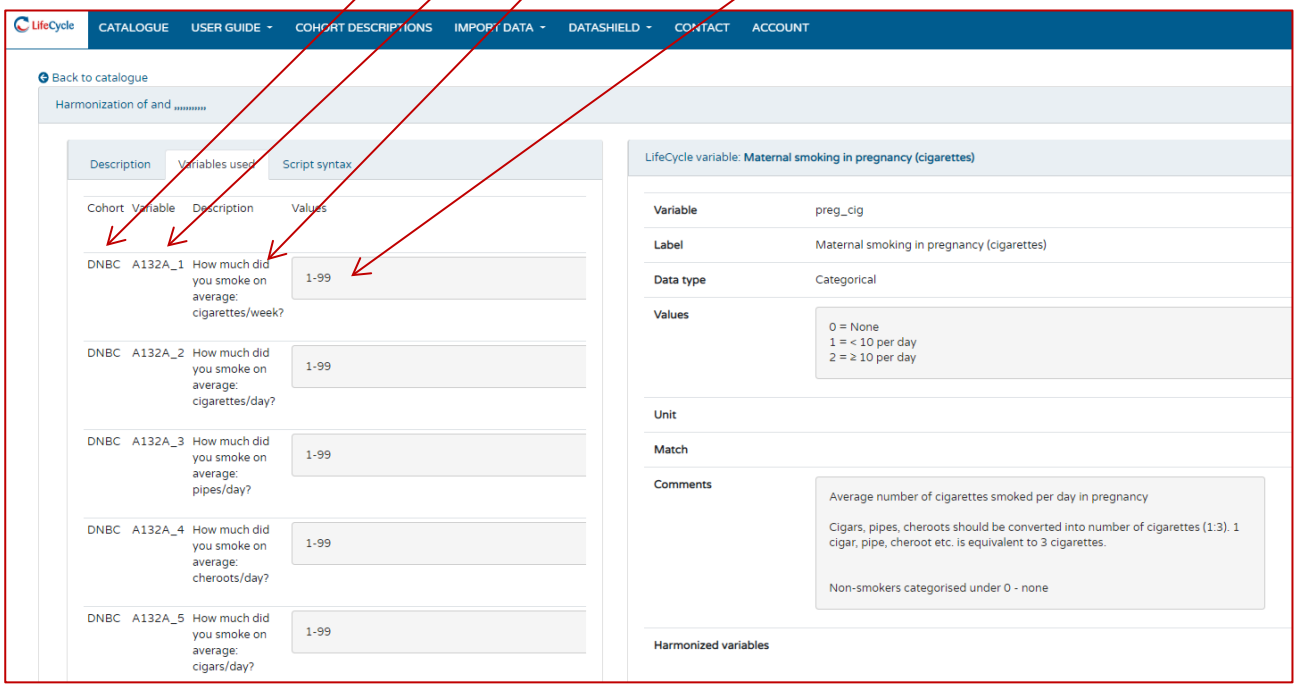
Ensure that the description can be understood by anyone out-side the cohort (see picture on the left for adequate example)

Ensure that cohort-specific information (left) exactly match the Harmonization Protocol (right)



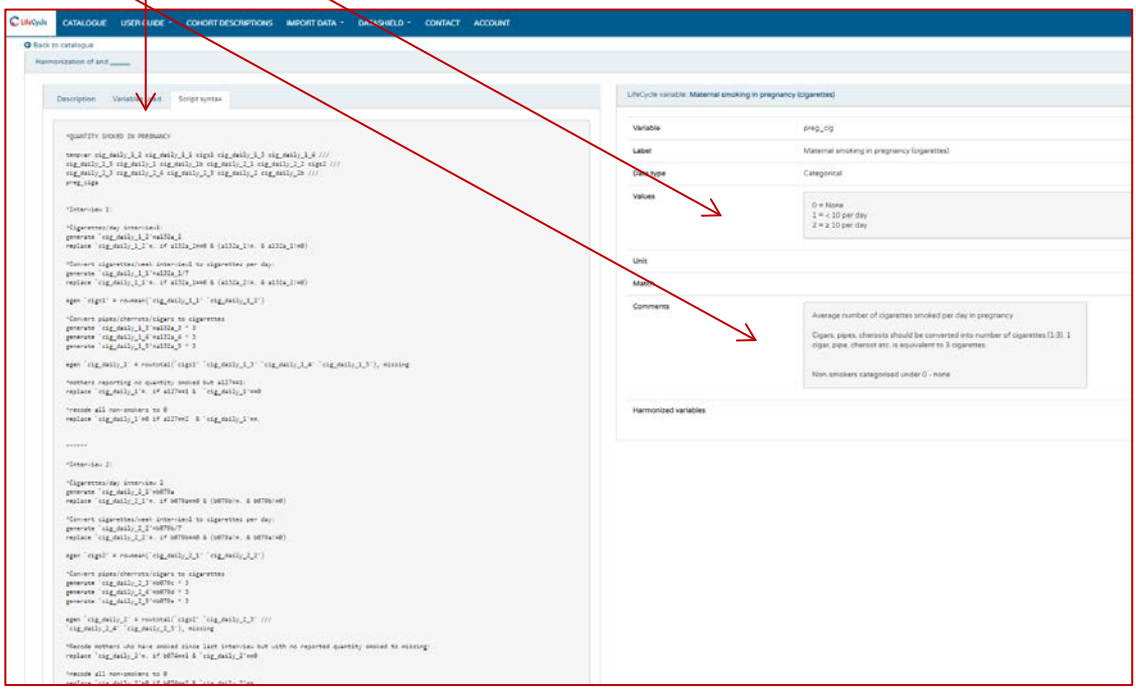
Check and verify ‘Variables used’

Ensure that information on *Cohort*, *Variable*, *Description*, and *Values* are completed in full.



Check and verify ‘Script syntax’

Check that the script syntax is filled in, and verify that the syntax reported is completed in full and matches the ‘*comments*’ and ‘*values*’ displayed on the right side. Also, ensure that any person outside the cohort can understand the syntax, and verify that any description in the syntax is in English.

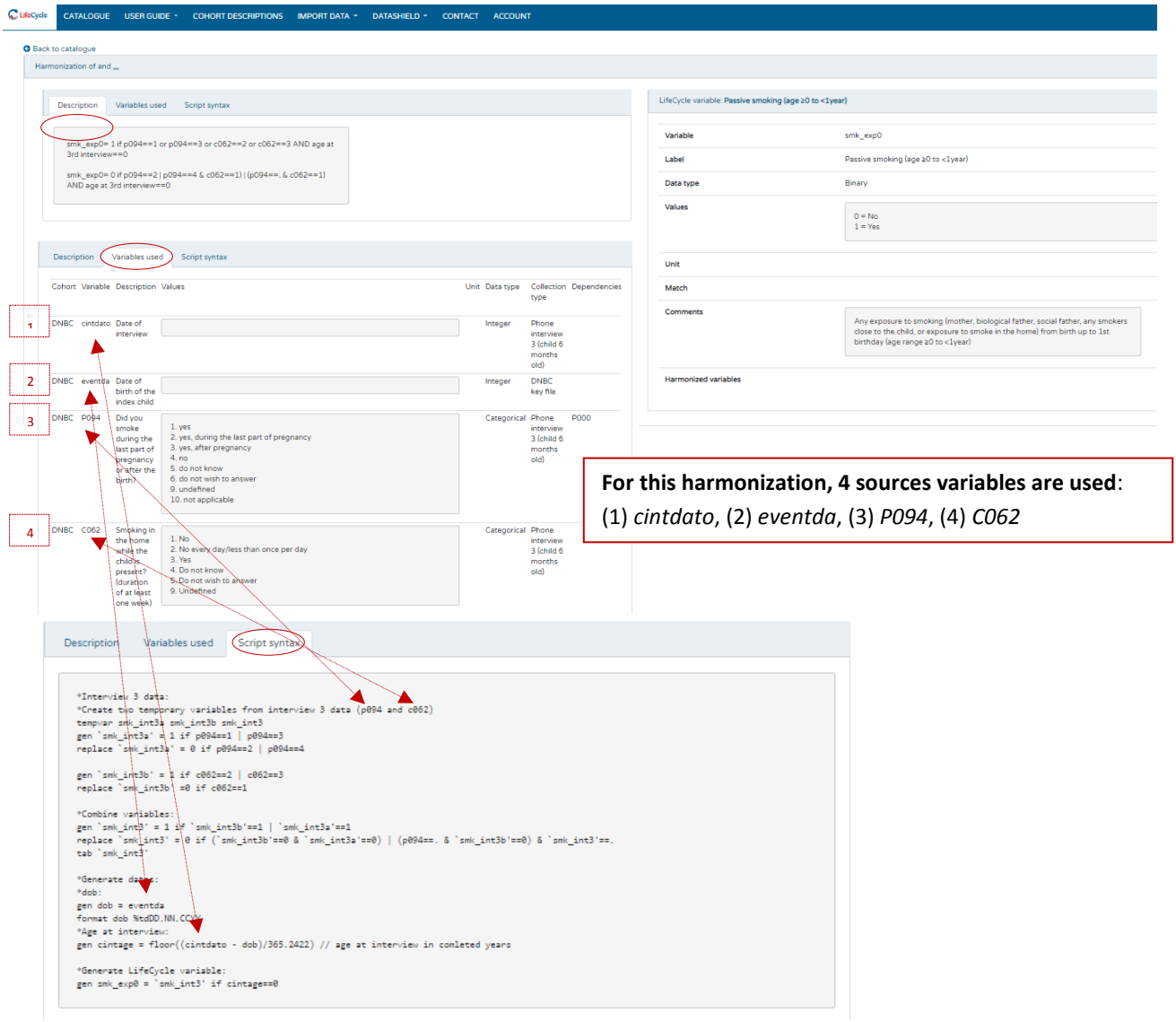


Check and verify consistency in Online Catalogue

As noted above, information in the Online Catalogue should be filled in by the cohorts themselves. Not only is it important that information in 'description', 'variables used' and 'script syntax' are filled in, but also information across the three boxes must be consistent.

Hence, the script syntax should include the variables used and match the variable description.

Please, <u>check and verify consistency in the core variables</u> incl. non-repeated and repeated measures:

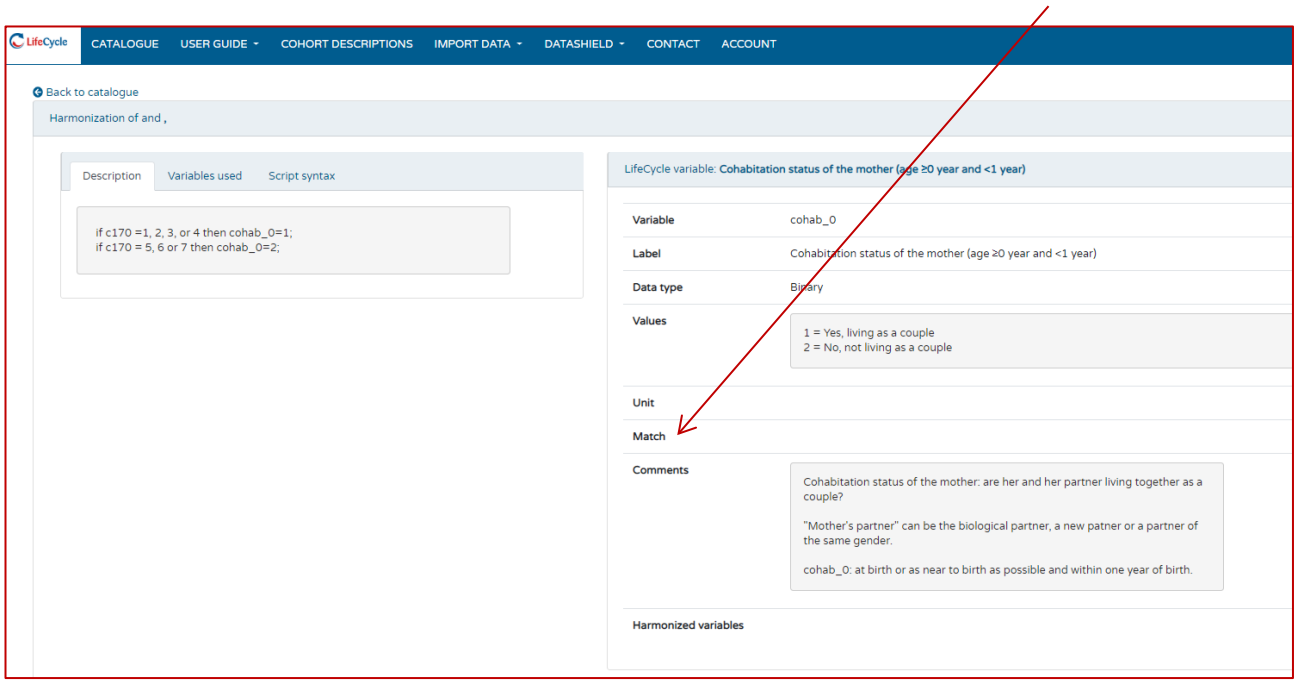


Check and verify consistency in repeated measures

Check that information (in '*description*', '*source variables*' and '*script syntax*') is filled in for each time band of the repeated measures. Each time band must have information on harmonization if variables are partial.

Check and verify the description of partially harmonized variables

Ensure that details of why a variable is only partially harmonized should be given in detail in ‘match’.



Quality Control Checklist

1. Check that all Core Variables are harmonized
2. Check that categories and formats for each variable match the Core Variable List
3. Check that repeated measures have the correct number of time bands
4. Check internal validity for each of the non-repeated and repeated measures
5. Check that repeated measures are placed into the correct time bands
6. Check that maternal education and child's height do not decrease over time
7. Check that all information in the Online Catalogue is completed, and is correct
8. Check that information in the Online Catalogue is understandable for anyone
9. Check that information on partial harmonization is explained in detail
10. Check that information in 'description', 'variables used' and 'syntax' match and are correct
11. Check that repeated measures are adequately described: correct number and information

Additional Comments R and SAS

SAS (from Trine Facius):

The code generates output tables that should be checked, as the STATA code. At a couple of places a table is generated in SAS.

These tables should contain the rows from the original dataset, where there might be problems with at specific variable. This is done for the variables

- *preg_no*; a table called preg_no_err is generated for those rows where there seems to be a problem with the pregnancy number
- *edu_m* ; a table called edu_m_err is generated for those rows, where the educational level seems to be decreasing over time
- *height*; a table called height_err is generated for those rows, where the height seems to be decreasing over time

R (from Sebastian Rauschert)

Running it on the Raine study data, it only throws errors when some variables are exclusively containing missing values, which makes sense and is not a problem.