WP6

Harmonization Manual

Authors: Mònica López Vicente, Jordi Júlvez, Raquel Garcia, Silvia Fernandez and Jordi Sunyer

Date: 18/07/2019

Contents

Aim

The aim of Work Package 6 of LifeCycle is: "To study mental developmental trajectories and diseases throughout the life course and to examine the associations of early-life stressors during preconception, pregnancy, infancy and early childhood with mental health during childhood, adolescence and adulthood."

We have developed a list of cognitive and mental health domains that are of interest for the EU Child Cohort Network. The definitions for these have been obtained by different members of the WP6 Working Group and are mainly based on the available data in participating cohorts, but also on previous harmonization in other studies, scientific literature, and expert knowledge.

The aim of this protocol is to guide the cohorts in the harmonization process to generate variables that will be comparable across cohorts and measurements. The harmonization strategy is based on an adaptation of the DataSHaPER guidelines [1], aimed to facilitate a rigorous, transparent and effective harmonization, made in the framework of the MeDALL project [2].

Variables and instructions

The Appendix Table 1 (page 32) provides a description of the cognitive and mental health domains classified in this WP6 based on cohort data availability.

Variable priorization

We have assigned two different levels of priority to the core variables: highest priority has been given to main LifeCycle WP6 outcomes (shown in **bold red** in Table 1), based on availability in most cohorts and relevance for the project; second priority has been given to secondary core outcomes (shown in **bold blue** in Table 1). All the cohorts with available data on main outcomes will be required to harmonize these variables by default. Table 1 would also be useful to identify the cohort-specific instruments that measure the main outcomes. The rest of the variables not described in Table 1 are not going to be harmonized, by now, in this protocol.

The following are the highest priority variables:

- Internalizing problems,
- externalizing problems,
- ADHD symptoms and diagnosis,
- Autism spectrum disorder (ASD) symptoms and diagnosis.

We suggest that you begin with the highest priority variables. Harmonization of the highest priority variables should be completed by **September 2019**. Harmonization of the second priority variables should be completed by **February 2020**.

Table 1. Codes and Abbreviation of instruments used for measuring the main domains

Code	Abbreviation	Instrument	Internalizing problems	Externalizing problems	ADHD	Autism spectrum	Gross motor	Fine motor	Non-verbal intelligence	Working memory	Language
1	ADBB	Alarm Distress Baby Scale				X					
2	ADRIYC	Australian Developmental Record for Infants and Young Children					X	X	X		X
3	AQ	Autism Quotient Questionnaire				X					
4	ASQ	Ages and stages Questionnaires					X	X	X		
5	BAS	British Ability Scales							X		X
6	BDIST	Battelle Developmental Inventory Screening Test									X
7	BLS	Brunet-Lezine psychometric scale					X	X			
8	BPM	Brief Problem Monitor	X	X							
9	BPVS	British Picture Vocabulary Scale									X
10	BSID	Bayley Scales of Infant Development					X	X			
11	BST	Bus Story Test									X
12	CAST	Childhood Autism Spectrum Test				X					
13	CBCL	Child Behaviour Checklist	X	X	X						
14	CCC	Children's Communication Checklist									X
15	CDI	Child Development Inventory					X	X			
16	CELF	Clinical Evaluation of Language Fundamentals									X
17	CFIT	Culture Fair Intelligence Test							X		
18	CIIS	Cattell Infant Intelligence Scale							X		
19	CKAT	Clinical Kinematic Assessment Tool						X			
20	CPRS-R	Revised Conners' Parent Rating Scale			X						
21	CST	Counting Span Test								X	X
22	DBLR	Dale and Bishop language rating									X
23	DCDQ	Developmental Coordination Disorder Questionnaire					X	X			
24	DDST	Denver Development Screening Test					X	X			X
25	DISC-IV/DSM	Diagnostic Interview Schedule for Children			X						
26	ELM	Early Language Milestone Scale									X
27	ELOLA	European oral language test battery of aphasic children									X
28	ESAT	Early Screening of Autistic Traits Questionnaire				X					
29	EYFSP	Early Years Foundation Stage Profile									X
30	GDO-R	Gesell Developmental Observation-Revised							X		
31	GMDS	Griffiths Mental Development scales					X	X	X		X
32	GPT	Grooved Pegboard Test						X			
33	HELP	Hawaii Early Learning Profile					X				X
34	ITSEA	Infant-Toddler Social and Emotional Assessment	X	X							
35	M-ABC	Movement Assessment Battery for Children					X	X			
36	MCDI	MacArthur Communicative Development Inventory									X
37	M-CHAT	Modified Checklist for Autism in Toddlers				X					
38	MoCa	Montreal Cognitive Assessment							X		
39	MSCA	McCarthy Scales of Children's Abilities					X	X	X	X	X
40	N-Back	Working Memory Test								X	
41	NEPSY-II	Developmental Neuropsychological Assessment, Second Edition									X
42	NVCC	Non-Verbal Communication Checklist				X					
43	ONA	Object Naming Assessment									X
44	PMT	Peg Moving Task						X			
45	RPM	Raven's Progressive Matrices							X		
46	SCQ	Social Communication Questionnaire				X					
47	SDQ	Strengths and Difficulties Questionnaire	X	X	X						
48	SLAS	Speech and Language Assessement Scale									X
49	SON-R	Snijders-Oomen Non-Verbal Intelligence Test							X		
50	SRS	Social Responsiveness Scale				X					
51	STYCAR	Children's developmental progress from birth to five years					X	X			X
52	SVF	Semantic Verbal Fluency test									X

No.	Abbr.	Name										
53	SWAN	Strengths and Weaknesses of ADHD symptoms and Normal behavior			■							
54	TMT	Trail Making Test							■			
55	TowenNE	Touwen Infant Neurological Examination						■				
56	TRF	Teacher Report Form	■	■								
57	WAIS	Wechsler Abbreviated Scale of Intelligence								■		■
58	WDSC	Woodside Developmental Screening Chart						■	■			■
59	WISC	Wechsler Intelligence Scale for Children								■		■
60	WPPSI	Wechsler Preschool and Primary Scale of Intelligence								■		■
61	YSR	Youth Self Report	■	■								

Harmonization

When harmonizing the variables, please use the cleanest variables available within your cohort. Harmonization is done with data in the wide format. When creating variables, if no data exist within your cohort for a given variable at any age or at specific ages, then skip these and do not create the variable. For example, if there are no medical records for ADHD diagnosis, then do not create any ADHD registry variables. If internalizing problems assessment was done once at age 7.5 years, then you only need to create the variables int_raw_7, int_age_7, int_instr_7, int_eval_7, int_pro_7, int_avg_7, and int_pc_7. For repeated measures, please use the actual age at time of assessment (as opposed to the average age of the cohort at follow-up).

A brief explanation of the table heading is given below (Table 2). The required information about the cognitive and mental health variables needed to be harmonized in this protocol is shown in Table 3 (pages 10-24). Additionally in a separate excel file you will find the **data dictionary**.

Table 2. A brief explanation of the harmonization table headings

Variable name	Label/description	Values	Unit	Data Type	Further instructions
The name of the harmonised variable. This name needs to be match exactly with the derived (harmonised) variable.	The description of the harmonised LifeCycle variable (matches with that provided in the online catalogue). There is no need to label variables.	Details the categories for categorical and binary variables	Gives the units for continuous variables	The data type: decimal (continuous), and integer (binary or categorical)	Further specific instructions for harmonization including prompts to make notes on the catalogue.

Please record a description of harmonization, to be entered in the **online catalogue** (see the excel example files). This includes a description of the source variables and whether the variable is fully or partially harmonized. Where variable is only partially harmonized, please provide an explanation for why the variable is partially harmonized in the harmonization description.

If you have any queries about harmonization or the mental health variables list please contact Jordi Julvez (jordi.julvez@isglobal.org) from WP6.

Instructions to derive the harmonized LifeCycle variables.

Handling missing items

We propose treating the total scale score as <u>missing</u> whenever there are more than 25% of the items comprising the scale incomplete. Otherwise, you should also compute the prorated total scale score ([Appendix I](#)) using the following equation:

$$T_P = \frac{T_R}{N_R} * N_T \qquad if \ \ \frac{N_R}{N_T} \geq 0.75$$

Where:

T_P = prorated total scale score
T_R = total raw score based on completed items
N_R = number of items which have been responded to
N_T = the total number of items in the scale

Note that $\frac{T_R}{N_R}$ is the average score.

Percentiles

Assign the corresponding percentile to each individual for each domain and age ([Appendix II](#)).

The interpretation of this new harmonized variable will be "the percentile where each individual is in relation to his/her group (cohort and age specific) using a specific instrument". We have decided not to use clinical cut-off points to derive the harmonized variables because they depend on the reference population used to define these cut-off points. The use of country-specific cut-offs in international studies, such as LifeCycle, is not adequate.

Standardized variables

For **cognitive outcomes**, in addition of creating percentile variables, provide standardized variables. First, calculate the z-score by subtracting the mean from the raw score and dividing by the standard deviation. Then, to convert the z-score on an IQ measure, multiply the z-score by 15 and add 100. ([Appendix III](#)). This method has been used in previous multicenter studies to homogenize different scales [3–8].

See Appendix I, II,III for examples based on Stata and R instructions.

In summary, for a given domain **Y at age t**, you should record the following variables:

   i.   total raw score: **Y_raw_t**
   ii.   exact age of the child (in years) at assessment: **Y_age_t**
   iii.   cohort-specific instrument: **Y_instr_t**
   iv.   evaluator: **Y_eval_t**
   v.   prorated total scale score: **Y_pro_t**
   vi.   average score: **Y_avg_t**
   vii. percentile: **Y_pc_t**
   viii. only for cognitive domains, calculate also standardized score: **Y_std_t**

where t equals

0: assessed when child is aged between > 0 year and < 1 year,

1: assessed when child is aged between ≥ 1 year and < 2 year,

2: assessed when child is aged between ≥ 2 year and < 3 year,

…

17: assessed when child is aged between ≥ 17 year and < 18 year

> **NOTE:** ALL variables described above are set to missing if >25% of the items comprising a scale are missing.
>
> This condition also applies for scales based in short forms with fewer items.

Additionally, for ADHD and autism domains create medical record/reported doctor diagnosis variables.

It could be that the same domain is measured by more than one instrument at the same age (Example 1) or there is more than one outcome per instrument (Example 2). Similarly, it could be the same domain is responded by more than one informant at the same age. To fix this problem, follow the prioritization principles below:

(i) Measure that is most frequently used within the cohort at different time points.

(ii) Measure with the best psychometric properties; the cohort expert in psychometrics will have to decide the best judged and accurate outcome. For additional help, please see *Instrument priority* column from Priority Score Table 4 in order to select the best judged instrument.

(iii) Measure with the prioritized informant. Please see *Informant priority* column from Priority Score Table 4.

Table 3. Cognitive and mental health variables harmonization table

	Variable name	Label/description	Values	Unit	Data Type	Comments	Further instructions
highest priority variables							
2nd priority variables							
META-VARIABLES							
Child identifier	child_id	Unique identifier number for the index child				Either the original id or a new id generated by the cohort	Should already be created for the core variable list, please add here to make it possible to combine data
DOMAINS							
Internalizing problems	int_raw_0 int_raw_1 int_raw_2 … int_raw_17	Repeated measures of internalizing symptoms, total raw score: int_raw_0: assessed within one year of childbirth (>0 year and <1 year) int_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … int_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal	For SDQ, combine the emotional and peer items	The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized
	int_age_0 int_age_1 int_age_2 … int_age_17	Exact age of the child (in years) when internalizing symptoms were recorded for int_raw_0, int_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	
	int_instr_0 int_instr_1 int_instr_2 … int_instr_17	Cohort-specific instrument used to measure the internalizing symptoms for int_raw_0, int_raw_1, etc.	8) BPM 13) CBCL 34) ITSEA 47) SDQ 56) TRF 61) YSR		Categorical		

	int_eval_0 int_eval_1 int_eval_2 … int_eval_17	Who the test was administered/answered by for int_raw_0, int_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.	
	int_pro_0 int_pro_1 int_pro_2 … int_pro_17	Total internalizing symptoms score after prorating int_raw_0, int_raw1, etc.			Decimal		
	int_avg_0 int_avg_1 int_avg_2 … int_avg_17	Average of available items comprising the total raw score of int_raw_0, int_raw_1, etc.			Decimal		
	int_pc_0 int_pc_1 int_pc_2 … int_pc_17	internalizing symptoms, percentiles of int_pro_0, int_pro_1, etc.			Decimal		
Externalizing problems	ext_raw_0 ext_raw_1 ext_raw_2 … ext_raw_17	Repeated measures of externalizing symptoms, total raw score: ext_raw_0: assessed within one year of childbirth (>0 year and <1 year) ext_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … ext_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal	For SDQ, combine the conduct problems and hyperactivity/inattention items	The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized
	ext_age_0 ext_age_1 ext_age_2 … ext_age_17	Exact age of the child (in years) when externalizing symptoms were recorded for ext_raw_0, ext_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	

	ext_instr_0 ext_instr_1 ext_instr_2 … ext_instr_17	Cohort-specific instrument used to measure the externalizing symptoms for ext_raw_0, ext_raw_1, etc.	8) BPM 13) CBCL 34) ITSEA 47) SDQ 56) TRF 61) YSR		Categorical		
	ext_eval_0 ext_eval_1 ext_eval_2 … ext_eval_17	Who the test was administered/answered by for ext_raw_0, ext_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.	
	ext_pro_0 ext_pro_1 ext_pro_2 … ext_pro_17	Total internalizing symptoms score after prorating ext_raw_0, ext_raw1, etc.			Decimal		
	ext_avg_0 ext_avg_1 ext_avg_2 … ext_avg_17	Average of available items comprising the total raw score of ext_raw_0, ext_raw_1, etc.			Decimal		
	ext_pc_0 ext_pc_1 ext_pc_2 … ext_pc_17	internalizing symptoms, percentiles of ext_pro_0, ext_pro_1, etc.			Decimal		
ADHD symptoms	adhd_raw_0 adhd_raw_1 adhd_raw_2 … adhd_raw_17	Repeated measures of adhd scale, total raw score: adhd_raw_0: assessed within one year of childbirth (>0 year and <1 year) adhd_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … adhd_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal	For DISC-IV/DSM, combine all items rated on a 4-point scale (0 = never or rarely, 1 = sometimes, 2 = often, or 3 = very often).	The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized

adhd_age_0 adhd_age_1 adhd_age_2 … adhd_age_17	Exact age of the child (in years) when ADHD was recorded for adhd_raw_0, adhd_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	
adhd_instr_0 adhd_instr_1 adhd_instr_2 … adhd_instr_17	Cohort-specific instrument used to measure ADHD for adhd_raw_0, adhd_raw_1, etc	13) CBCL 20) CPRS-R 25) DISC-IV/DSM 47) SDQ 53) SWAN		Categorical	For CBCL, use the DSM-Oriented Attention-Deficit Hyperactivity Problem (DSM-ADH) scale	
adhd_eval_0 adhd_eval_1 adhd_eval_2 … adhd_eval_17	Who the test was administered/answered by for adhd_raw_0, adhd_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.	
adhd_pro_0 adhd_pro_1 adhd_pro_2 … adhd_pro_17	Total ADHD score after prorating adhd_raw_0, adhd_raw_1, etc.			Decimal		
adhd_avg_0 adhd_avg_1 adhd_avg_2 … adhd_avg_17	Average of available items comprising the total raw score of adhd_raw_0, adhd_raw_1, etc.			Decimal		
adhd_pc_0 adhd_pc_1 adhd_pc_2 … adhd_pc_17	ADHD scale, percentile of adhd_pro_0, adhd_pro_1, etc.			Decimal		

ADHD Registry diagnosis	adhdR_0 adhdR_1 adhdR_2 … adhdR_17	Repeated measures of ADHD diagnosis: adhdR_0: assessed within one year of childbirth (>0 year and <1 year) adhdR_1: assessed when the child was aged between ≥ 1 year and < 2 years … adhdR_17: assessed when the child was aged between ≥ 17 years and < 18 years	0) No 1) Yes		Binary		
	adhdR_age_0 adhdR_age_1 adhdR_age_2 … adhdR_age_17	Exact age of the child (in years) at onset for ADHD diagnosis for adhdR_0, adhdR_1, etc.		years	Decimal	If the onset age is not known please use the follow-up age	The variable is partially harmonized if using age of the child at follow-up as age of onset
	adhdR_eval_0 adhdR_eval_1 adhdR_eval_2 … adhdR_eval_17	How was the medical diagnosis reported for adhdR_0, adhdR_1, etc.	1) Medical registries or medical records linkage 2) Parentally reported doctor-diagnosed disorder		Binary		
Autism spectrum disorder (ASD) symptoms	asd_raw_0 asd_raw_1 asd_raw_2 … asd_raw_17	Repeated measures of ASD, total raw score: asd_raw_0: assessed within one year of childbirth (>0 year and <1 year) asd_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … asd_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal		The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized
	asd_age_0 asd_age_1 asd_age_2 … asd_age_17	Exact age of the child (in years) when ASD was recorded for asd_raw_0, asd_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	

	Variable	Description	Values		Type	Notes	
	asd_instr_0 asd_instr_1 asd_instr_2 … asd_instr_17	Cohort-specific instrument used to measure ASD for asd_raw_0, asd_raw_1, etc	1) ADBB 3) AQ 12) CAST 28) ESAT 37) M-CHAT 42) NVCC 46) SCQ 50) SRS		Categorical		
	asd_eval_0 asd_eval_1 asd_eval_2 … asd_eval_17	Who the test was administered/answered by for asd_raw_0, asd_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.	
	asd_pro_0 asd_pro_1 asd_pro_2 … asd_pro_17	Total ASD score after prorating asd_raw_0, asd_raw_1, etc.			Decimal		
	asd_avg_0 asd_avg_1 asd_avg_2 … asd_avg_17	Average of available items comprising the total raw score of asd_raw_0, asd_raw_1, etc.			Decimal		
	asd_pc_0 asd_pc_1 asd_pc_2 … asd_pc_17	ASD scale, percentile of asd_pro_0, asd_pro_1, etc.			Decimal		
ASD Registry diagnosis	asdR_0 asdR_1 asdR_2 … asdR_17	Repeated measures of ASD diagnosis: asdR_0: assessed within one year of childbirth (>0 year and <1 year) asdR_1: assessed when the child was aged between ≥ 1 year and < 2 years … asdR_17: assessed when the child was aged between ≥ 17 years and < 18 years	0) No 1) Yes		Binary		

	asdR_age_0 asdR_age_1 asdR_age_2 … asdR_age_17	Exact age of the child (in years) at onset for ASD diagnosis for asdR_0, asdR_1, etc.		years	Decimal	If the onset age is not known please use the follow-up age	The variable is partially harmonized if using age of the child at follow-up as age of onset
	asdR_eval_0 asdR_eval_1 asdR_eval_2 … asdR_eval_17	How was the medical diagnosis reported for asdR_0, asdR_1, etc.	1) Medical registries or medical records linkage 2) Parentally reported doctor-diagnosed disorder		Binary		
Gross motor	gm_raw_0 gm_raw_1 gm_raw_2 … gm_raw_17	Repeated measures of gross motor, total raw score: gm_raw_0: assessed within one year of childbirth (>0 year and <1 year) gm_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … gm_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal		The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized
	gm_age_0 gm_age_1 gm_age_2 … gm_age_17	Exact age of the child (in years) when gross motor score was recorded for gm_raw_0, gm_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	

	gm_instr_0 gm_instr_1 gm_instr_2 … gm_instr_17	Cohort-specific instrument used to measure the Gross motor for gm_raw_0, gm_raw_1, etc.	2) ADRIYC 4) ASQ 7) BLS 10) BSID 15) CDI 23) DCDQ 24) DDST 31) GMDS 33) HELP 35) M-ABC 39) MSCA 51) STYCAR 55) TowenNE 58) WDSC		Categorical	
	gm_eval_0 gm_eval_1 gm_eval_2 … gm_eval_17	who the test was administered/answered by for gm_raw_0, gm_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.
	gm_pro_0 gm_pro_1 gm_pro_2 … gm_pro_17	Total gross motor score after prorating gm_raw_0, gm_raw_1, etc.			Decimal	If gross motor is measured using "computerized" test, create prorated scores equals to raw scores: gm_pro_0=gm_raw_0, gm_pro_1=gm_raw_1, etc.
	gm_avg_0 gm_avg_1 gm_avg_2 … gm_avg_17	Average of available items comprising the total raw score of gm_raw_0, gm_raw_1, etc.			Decimal	
	gm_pc_0 gm_pc_1 gm_pc_2 … gm_pc_17	Gross motor, percentile of gm_pro_0, gm_pro_1, etc.			Decimal	

	gm_std_0 gm_std_1 gm_std_2 … gm_std_17	Gross motor, standardized score of gm_pro_0, gm_pro_1, etc			Decimal		
Fine motor	fm_raw_0 fm_raw_1 fm_raw_2 … fm_raw_17	Repeated measures of fine motor, total raw score: fm_raw_0: assessed within one year of childbirth (>0 year and <1 year) fm_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … fm_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal		The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized
	fm_age_0 fm_age_1 fm_age_2 … fm_age_17	Exact age of the child (in years) when fine motor score was recorded for fm_raw_0, fm_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	
	fm_instr_0 fm_instr_1 fm_instr_2 … fm_instr_17	Cohort-specific instrument used to measure the Fine motor for fm_raw_0, fm_raw_1, etc.	2) ADRIYC 4) ASQ 7) BLS 10) BSID 15) CDI 19) CKAT 23) DCDQ 24) DDST 31) GMDS 32) GPT 33) HELP 35) M-ABC 39) MSCA 44) PMT 51) STYCAR 54) TMT 58) WDSC		Categorical		

	fm_eval_0 fm_eval_1 fm_eval_2 … fm_eval_17	who the test was administered/answered by for fm_raw_0, fm_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.	
	fm_pro_0 fm_pro_1 fm_pro_2 … fm_pro_17	Total fine motor score after prorating fm_raw_0, fm_raw_1, etc.			Decimal	If fine motor is measured using "computerized" test, create prorated scores equals to raw scores: fm_pro_0=fm_raw_0, fm_pro_1=fm_raw_1, etc.	
	fm_avg_0 fm_avg_1 fm_avg_2 … fm_avg_17	Average of available items comprising the total raw score of fm_raw_0, fm_raw_1, etc.			Decimal		
	fm_pc_0 fm_pc_1 fm_pc_2 … fm_pc_17	Fine motor, percentile of fm_pro_0, fm_pro_1, etc.			Decimal		
	fm_std_0 fm_std_1 fm_std_2 … fm_std_17	Fine motor, standardized score of fm_pro_0, fm_pro_1, etc			Decimal		
Non-verbal intelligence	nvi_raw_0 nvi_raw_1 nvi_raw_2 … nvi_raw_17	Repeated measures of non-verbal intelligence, total raw score: nvi_raw_0: assessed within one year of childbirth (>0 year and <1 year) nvi_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … nvi_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal		The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized

	nvi_age_0 nvi_age_1 nvi_age_2 … nvi_age_17	Exact age of the child (in years) when non-verbal intelligence score was recorded for nvi_raw_0, nvi_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	
	nvi_instr_0 nvi_instr_1 nvi_instr_2 … nvi_instr_17	Cohort-specific instrument used to measure the non-verbal intellgence for nvi_raw_0, nvi_raw_1, etc.	4) ASQ 5) BAS 10) BSID 17) CFIT 18) CIIS 30) GDO-R 31) GMDS 33) HELP 39) MSCA 45) RPM 49) SON-R 57) WAIS 59) WISC 60) WPPSI		Categorical	Use the following specified scales: Problem solving for ASQ; Cognitive for BSID and HELP; Perceptual-performance for MSCA; Performance IQ for WPPSI; Fluid reasoning index for WISC; Perceptual reasoning for WAIS; Practical reasoning for GMDS; Visual/spatial for GDO-R	
	nvi_eval_0 nvi_eval_1 nvi_eval_2 … nvi_eval_17	who the test was administered/answered by for nvi_raw_0, nvi_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.	
	nvi_pro_0 nvi_pro_1 nvi_pro_2 … nvi_pro_17	Total non-verbal intelligence score after prorating nvi_raw_0, nvi_raw_1, etc.			Decimal	If non-verbal intelligence is measured using "computerized" test, create prorated scores equals to raw scores: nvi_pro_0=nvi_raw_0, nvi_pro_1=nvi_raw_1, etc.	
	nvi_avg_0 nvi_avg_1 nvi_avg_2 … nvi_avg_17	Average of available items comprising the total raw score of nvi_raw_0, nvi_raw_1, etc.			Decimal		
	nvi_pc_0 nvi_pc_1 nvi_pc_2 … nvi_pc_17	Non-verbal intelligence, percentile of nvi_pro_0, nvi_pro_1, etc.			Decimal		

	nvi_std_0 nvi_std_1 nvi_std_2 … nvi_std_17	Non-verbal intelligence, standardized score of nvi_pro_0, nvi_pro_1, etc			Decimal		
Working memory	wm_raw_0 wm_raw_1 wm_raw_2 … wm_raw_17	Repeated measures of working memory total raw score: wm_raw_0: assessed within one year of childbirth (>0 year and <1 year) wm_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … wm_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal		The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized
	wm_age_0 wm_age_1 wm_age_2 … wm_age_17	Exact age of the child (in years) when working memory score was recorded for wm_raw_0, wm_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	
	wm_instr_0 wm_instr_1 wm_instr_2 … wm_instr_17	Cohort-specific instrument used to measure the working memory for wm_raw_0, wm_raw_1, etc.	21) CST 40) N-Back		Categorical	For N-Back, use detectability (d')	For N-Back use the load and stimuli with the highest number of repeated measures available during the follow-up.
	wm_eval_0 wm_eval_1 wm_eval_2 … wm_eval_17	who the test was administered/answered by for wm_raw_0, wm_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.	
	wm_pro_0 wm_pro_1 wm_pro_2 … wm_pro_17	Total working memory score after prorating wm_raw_0, wm_raw_1, etc.			Decimal	If working memory is measured using "computerized" test, create prorated scores equals to raw scores: wm_pro_0=wm_raw_0, wm_pro_1=wm_raw_1, etc.	

	wm_avg_0 wm_avg_1 wm_avg_2 … wm_avg_17	Average of available items comprising the total raw score of wm_raw_0, wm_raw_1, etc.			Decimal		
	wm_pc_0 wm_pc_1 wm_pc_2 … wm_pc_17	Working memory, percentile of wm_pro_0, wm_pro_1, etc.			Decimal		
	wm_std_0 wm_std_1 wm_std_2 … wm_std_17	Working memory, standardized score of wm_pro_0, wm_pro_1, etc			Decimal		
Language	lan_raw_0 lan_raw_1 lan_raw_2 … lan_raw_17	Repeated measures of language total raw score: lan_raw_0: assessed within one year of childbirth (>0 year and <1 year) lan_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … lan_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal		The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized
	lan_age_0 lan_age_1 lan_age_2 … lan_age_17	Exact age of the child (in years) when language score was recorded for lan_raw_0, lan_raw_1, etc.		years	Decimal	Calculate age in years as (date of test assessment - date of birth)/365.2422	

	lan_instr_0 lan_instr_1 lan_instr_2 … lan_instr_17	Cohort-specific instrument used to measure the language for lan_raw_0, lan_raw_1, etc.	2) ADRIYC 5) BAS 6) BDIST 7) BLS 9) BPVS 10) BSID 11) BST 14) CCC 15) CDI 16) CELF 22) DBLR 24) DDST 26) ELM 27) ELOLA 29) EYFSP 30) GDO-R 31) GMDS 33) HELP 36) MCDI 38) MoCa 39) MSCA 41) NEPSY-II 43) ONA 48) SLAS 51) STYCAR 52) SVF 57) WAIS 58) WDSC 59) WISC 60) WPPSI		Categorical	Use the following specified scales: Expressive subscale for BAS, BDIST; Auditory expressive for ELM; Language development speaking = for EYFSP; Language/Comprehension for GDO-R; Verbal for MSCA; Verbal comprehension for WAIS, WISC; Verbal IQ for WPPSI; Expressive naming, verbal fluency for NEPSY-II; Speech and language achievements for STYCAR	
	lan_eval_0 lan_eval_1 lan_eval_2 … lan_eval_17	who the test was administered/answered by for lan_raw_0, lan_raw_1, etc.	1) teacher 2) parent 3) examiner 4) computerized 5) self-reported 6) other		Categorical	"computerized" (value label 4) only applies if the child perform the test with automatic results. Any computerized rating scale or questionnaire does not apply.	

		lan_pro_0 lan_pro_1 lan_pro_2 … lan_pro_17	Total language score after prorating lan_raw_0, lan_raw_1, etc.			Decimal	If language is measured using "computerized" test, create prorated scores equals to raw scores: lan_pro_0=lan_raw_0, lan_pro_1=lan_raw_1, etc.	
		lan_avg_0 lan_avg_1 lan_avg_2 … lan_avg_17	Average of available items comprising the total raw score of lan_raw_0, lan_raw_1, etc.			Decimal		
		lan_pc_0 lan_pc_1 lan_pc_2 … lan_pc_17	Language, percentile of lan_pro_0, lan_pro_1, etc.			Decimal		
		lan_std_0 lan_std_1 lan_std_2 … lan_std_17	Language, standardized score of lan_pro_0, lan_pro_1, etc			Decimal		

Table 4. Priory Score Table for Psychometric Tests

Abbreviation	Instrument	Instrument Priority (1=high, 2=low)	Informant Priority (mother, father, examiner, teacher, self-reported, computer based…)
ADBB	Alarm Distress Baby Scale	2	examiner
ADRIYC	Australian Developmental Record for Infants and Young Children	2	examiner
AQ	Autism Quotient Questionnaire	2	mother
ASQ	Ages and stages Questionnaires	2	mother
BAS	British Ability Scales	1	examiner
BDIST	Battelle Developmental Inventory Screening Test	2	examiner
BLS	Brunet-Lezine psychometric scale	1	examiner
BPM	Brief Problem Monitor	2	mother
BPVS	British Picture Vocabulary Scale	2	teacher
BSID	Bayley Scales of Infant Development	1	examiner
BST	Bus Story Test	2	examiner
CAST	Childhood Autism Spectrum Test	1	examiner
CBCL	Child Behaviour Checklist	1	mother
CCC	Children's Communication Checklist	2	mother
CDI	Child Development Inventory	2	mother
CELF	Clinical Evaluation of Language Fundamentals	1	examiner
CFIT	Culture Fair Intelligence Test	1	examiner
CIIS	Cattell Infant Intelligence Scale	1	examiner
CKAT	Clinical Kinematic Assessment Tool	1	examiner / computer based
CPRS-R	Revised Conners' Parent Rating Scale	1	mother
CST	Counting Span Test	2	examiner
DBLR	Dale and Bishop language rating	1	mother
DCDQ	Developmental Coordination Disorder Questionnaire	2	mother
DDST	Denver Development Screening Test	1	examiner
DISC-IV/DSM	Diagnostic Interview Schedule for Children	1	examiner
ELM	Early Language Milestone Scale	1	examiner
ELOLA	European oral language test battery of aphasic children	1	examiner
ESAT	Early Screening of Autistic Traits Questionnaire	1	mother
EYFSP	Early Years Foundation Stage Profile	1	examiner
GDO-R	Gesell Developmental Observation-Revised	1	examiner
GMDS	Griffiths Mental Development scales	1	examiner

GPT	Grooved Pegboard Test	1	examiner
HELP	Hawaii Early Learning Profile	1	examiner
ITSEA	Infant-Toddler Social and Emotional Assessment	1	mother
M-ABC	Movement Assessment Battery for Children	1	examiner
MCDI	MacArthur Communicative Development Inventory	1	mother
M-CHAT	Modified Checklist for Autism in Toddlers	2	mother
MoCa	Montreal Cognitive Assessment	1	examiner
MSCA	McCarthy Scales of Children's Abilities	1	examiner
N-Back	Working Memory Test	1	examiner/computer-based
NEPSY-II	Developmental Neuropsychological Assessment, Second Edition	1	examiner
NVCC	Non-Verbal Communication Checklist	2	teacher
ONA	Object Naming Assessment	2	examiner
PMT	Peg Moving Task	1	examiner
RPM	Raven's Progressive Matrices	1	examiner
SCQ	Social Communication Questionnaire	1	mother
SDQ	Strengths and Difficulties Questionnaire	2	mother
SLAS	Speech and Language Assessement Scale	2	mother
SON-R	Snijders-Oomen Non-Verbal Intelligence Test	1	examiner
SRS	Social Responsiveness Scale	1	mother
STYCAR	Children's developmental progress from birth to five years	1	examiner
SVF	Semantic Verbal Fluency test	2	examiner
SWAN	Strengths and Weaknesses of ADHD symptoms and Normal behavior	1	mother
TMT	Trail Making Test	2	examiner
TowenNE	Touwen Infant Neurological Examination	2	examiner
TRF	Teacher Report Form	1	teacher
WAIS	Wechsler Abbreviated Scale of Intelligence	1	examiner
WDSC	Woodside Developmental Screening Chart	2	examiner
WISC	Wechsler Intelligence Scale for Children	1	examiner
WPPSI	Wechsler Preschool and Primary Scale of Intelligence	1	examiner
YSR	Youth Self Report	1	self-reported

Uploading data to OPAL

Please check WP1 for instructions on uploading harmonized tables to Opal. When you have created the harmonized WP6 dataset in wide format, this should be reshaped into a table with yearly-repeated measures. The reshaping scripts will be shared with all cohorts at a later stage. These scripts will create .csv files that can be uploaded in Opal/DataSHIELD, and contain the meta variable **child_id** to allow merging with the other tables that are uploaded online.

References

1. Fortier I, Burton PR, Robson PJ, Ferretti V, Little J, L'Heureux F, et al. Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. Int J Epidemiol. 2010;39:1383–93.

2. Antó JM, Pinart M, Akdis M, Auffray C, Bachert C, Basagaña X, et al. Understanding the complexity of IgE-related phenotypes from childhood to young adulthood: a Mechanisms of the Development of Allergy (MeDALL) seminar. J Allergy Clin Immunol. 2012;129:943–954.e4.

3. Levie D, Korevaar TIM, Bath SC, Dalmau-Bueno A, Murcia M, Espada M, et al. Thyroid Function in Early Pregnancy, Child IQ, and Autistic Traits: A Meta-Analysis of Individual Participant Data. J Clin Endocrinol Metab. 2018;103:2967–79.

4. Forns J, Torrent M, Garcia-Esteban R, Grellier J, Gascon M, Julvez J, et al. Prenatal exposure to polychlorinated biphenyls and child neuropsychological development in 4-year-olds: An analysis per congener and specific cognitive domain. Sci Total Environ. 2012;432:338–43.

5. Forns J, Fort M, Casas M, Cáceres A, Guxens M, Gascon M, et al. Exposure to metals during pregnancy and neuropsychological development at the age of 4 years. NeuroToxicology. 2014;40:16–22.

6. Vilahur N, Fernández MF, Bustamante M, Ramos R, Forns J, Ballester F, et al. In utero exposure to mixtures of xenoestrogens and child neuropsychological development. Environ Res. 2014;134:98–104.

7. Boucher O, Julvez J, Guxens M, Arranz E, Ibarluzea J, Miguel MS de, et al. Association between breastfeeding duration and cognitive development, autistic traits and ADHD symptoms: a multicenter study in Spain. Pediatr Res. 2017;81:434–42.

8. Villanueva CM, Gracia-Lavedan E, Julvez J, Santa-Marina L, Lertxundi N, Ibarluzea J, et al. Drinking water disinfection by-products during pregnancy and child neuropsychological development in the INMA Spanish cohort study. Environ Int. 2018;110:113–22.

EXAMPLES

1. **Example of the same domain measured by more than one instrument at the same age using data from INMA cohort.** Language was assessed using verbal subscale of McCarthy Scales of Children's Abilities (MSCA) (`pdp_verm_4y`) and Battelle Developmental Inventory Screening Test (BDIST) (`expresiva_4y`) at 4 years follow-up:

child_id	edMcCarthy	pdp_verm	expresiva
780101	4.235455	57	49

Following Priority Score Table, Verbal subscale of MSCA has priority over BDIST. So, the harmonized variables would be:

child_id	lang_raw_4	lang_age_4	lang_instr_4	lang_eval_4
780101	57	4.235455	39	3

2. **Example of more than one outcome per instrument allowed in the dictionary using data from INMA cohort.** Working memory was assessed using detectability from N-Back test at 7-, 9- and 11 years follow-ups. Three different stimuli (colours, letters, and numbers) and n-back loads (up to 3 back) were examined

row id	visit	dcolours1	dcolours2	dcolours3	dletters1	dletters2	dletters3	dnumbers1	dnumbers2	dnumbers3
417010107	7	1.28978	1.094968	1.28978	2.486475	2.213311	0.5882741	1.878055	0.7830862	1.028204
417010107	9	3.46105	3.628356	3.604818	.	.	.	3.919929	3.919929	3.919929
417010107	11	.	.	.	.	.	.	3.919929	3.919929	2.486475

2-back numbers detectability was the more proxy outcome decided by the expert in psychometrics as well as the most measured repeatedly. So, the harmonized raw score and age variables would be:

child id	wm raw 6	wm age 6	wm raw 9	wm age 9	wm raw 10	wm age 10
4170101	0.7830862	6.644764	3.919929	9.267625	3.919929	10.8063

APPENDIX:

I.   Stata syntax for prorating internalizing symptoms score

Strengths and Difficulties Questionnaire (SDQ) was the instrument used in INMA cohort for assessing internalizing symptoms at child's age 7 years. The syntax below creates the total raw score (`int_raw_7`), the average of completed items (`int_avg_7`) and the pro-rated score (`int_pro_7`) for internalizing problems if ≥ 75% of items completed:

```
generate int_raw_7 = 0
generate Nc = 0   /* number of completed items */
local vars "psomatic pworries punhappy pclingy pafraid ploner qfriend
qpopular pbullied poldbest"

local nvars = 0
foreach v of varlist `vars' {
    replace int_raw_7 = int_raw_7 + cond(missing(`v'), 0, `v')
    replace Nc = Nc + !missing(`v')
    local ++nvars
}
replace int_raw_7 = . if Nc /`nvars' < 0.75
gen int_avg_7 = int_raw_7 / Nc
gen int_pro_7 = round(`nvars' * int_avg_7)
drop Nc
```

II.   Syntax for obtaining percentiles

Creates language percentiles (`lan_pc_4`) that categorizes language total prorated scale score (`lan_pro_4`) by its percentiles:

Stata
```
xtile lan_pc_4 = lan_pro_4, nq(100)
```
R
```
# load required package
library(statar)
lan_pc_4 <-xtile(lan_pro_4, n=100)
```

III. Syntax for obtaining standardized variables

Creates the standardized values of language (`lan_std_4`), producing a variable with mean 100 and standard deviation 15.

Stata
```
egen lan_std_4 = std(lan_pro_4), mean(100) std(15)
```
R
```
std_m100sd15 <- function(x){
    res <- (x - mean(x, na.rm = T))/sd(x, na.rm = T)
    res <- (res * 15) + 100
    res
}
lan_std_4 <- std_m100sd15(lan_pro_4)
```

IV. 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

Table 1: Inventory of mental health domains in the LifeCycle cohorts

Domains	Dimensions	Definitions (different levels of complexity depending on age)
Mental health	Social behaviour/competency	Facial expression; eye contact; general level of activity; self-stimulation gestures; vocalizations; briskness of response to stimulation; relationship to the observer; attractiveness to the observer; social problems
	Depression	Measured through physical and emotional signs & symptoms
	Internalizing problems	Withdrawn; Somatic complaints; Anxious/depressed
	Externalizing problems	Attention problems; Delinquent behavior; Aggressive behavior
	Bipolar	Hyperactive, ambitious, exhibitionistic, feelings of euphoria, flights of thoughts
	Affectivity	Ability and degree to which negative versus positive emotions are experienced
	Social responsibility	Response to routine; response to the unfamiliar; following instructions; making explanations; sharing; helping others; initiating activities; giving direction to activities; reaction to frustration; accepting limits
	Autism spectrum disorder	Initiation and maintenance of communication; Social interaction; Play activities; Rigid or repetitive behaviours; Interests; Child responding & initiating; Qualitative abnormalities in reciprocal social interaction and communication, and restricted/repetitive/stereotyped patterns of behavior; Social Awareness, Social Cognition, Social Communication, Social Motivation, Restricted Interests and Repetitive Behavior
	Anxiety	Inertia, autonomic arousal, skeletal muscle effects, situational anxiety, subjective experience of anxious affect; Panic disorder, significant somatic symptoms, generalized anxiety disorder, separation anxiety disorder, social anxiety disorder, significant school avoidance
	Stress	Difficulty relaxing, nervous arousal, being easily upset/agitated, irritable/over-reactive and impatient; psychological well-being
	Misc. Psychiatric Conditions	Agoraphobia, Generalized anxiety, Overanxious, Obsessive-Compulsive, Panic, Post-traumatic Stress, Separation Anxiety, Social Phobia, Specific Phobia, Dysthymic disorder, major depressive episode, manic/hypomanic episode, ADHD, Oppositional-defiant, conduct disorder, alcohol/nicotine/marijuana/other drugs use/abuse/dependence, anorexia nervosa, bulimia nervosa, enuresis/encopresis, pica, schizophrenia, selective mutism, tourette's, trichotillomania
	Relationship quality	Dyadic Consensus; Dyadic Satisfaction; Dyadic Cohesion; Affection Expression
	Eating disorders	Restraint, eating concern, shape concern, weight concern
	Dysregulation	
	Maladaptive/antisocial behaviour	Physical aggression, social aggression, rule-breaking

	Family well-being	Problem Solving, Communication, Roles, Affective Responsiveness, Affective Involvement, Behavior Control and General Functioning
	Self efficacy	Social-emotional and behavioral functioning
	Schizophrenic body image aberration	Unclear boundaries of the body; feeling of unreality or estrangement of parts of one's body; feeling of deterioration of one's body; perception of change in the size, proportions, spatial relationship of one's body part; changes in the appearance of the body
	Quality of Life	Mobility, vision, hearing, breathing, sleeping, eating, speech, excretion, usual activities, mental function, discomfort & symptoms, depression, distress, vitality, sexual activity
	ADHD	Hyperactivity-impulsivity; Attention difficulties
	Physical/social Anhedonia	Ability to experience pleasure from typically pleasurable physical stimuli such as food, sex, and settings; ability to experience pleasure from non physical stimuli such as other people, talking, exchanging expressions of feelings
	Self-perception	Scholastic Competence, Social Competence, Athletic Competence, Physical Appearance, Job Competence, Romantic Appeal, Behavioral Conduct, Close Friendship, Global Self- Worth
	Adaptive functioning	Communication, self-help, self-direction
	Temperament/personality	Novelty seeking, harm avoidance, reward dependence, persistence, self-directedness, cooperativeness, self-transcendence; Activity level, rhythmicity, adaptability, approach to novelty, emotional intensity, quality of mood, sensory sensitivity, distractibility and persistence
	Alexithymia	Difficulty Describing Feelings, Difficulty Identifying Feelings, Externally-Oriented Thinking
	Coping mechanisms	Confrontation Coping, Distancing, Self-Controlling, Seeking Social Support, Accepting Responsibility, Escape-Avoidance, Planful Problem Solving, and Positive Reappraisal
	Risk behavior	Behaviors that contribute to unintentional injuries and violence, Sexual behaviors, Alcohol and other drug use, Tobacco use, Unhealthy dietary behaviors, Inadequate physical activity
	Obsessive compulsive disorder	Checking, touching, cleanliness/washing, repeating and exactness
	Emotional perception	Optimism, pessimism
	Schizoidia	Schizophrenia spectrum and personality disorders
School performance	Learning difficulties	School problems, special class, special educational need, repeated grade
	Performance assessments	National databases linkage and self-reported final grades and average/median test scores
	Mathematics/Science	Final grades, teacher-led assessments
	Language/Writting/Reading	Final grades, teacher-led assessments
	Perceived school performance	Parent's report of its own child school success

Psychomotor	Gross motor	Gross motor skills are involved in movement and coordination of the arms, legs, and other large body parts and movements. Gross motor skills can be further divided into two subgroups of locomotor skills and object control skills. Gross locomotor skills would include running, jumping, sliding, and swimming. Object control skills would include throwing, catching and kicking.
	Fine motor	Fine motor skills are involved in smaller movements that occur in the wrists, hands, fingers, and the feet and toes. They participate in smaller actions such as picking up objects between the thumb and finger, writing carefully, and even blinking.
Memory	Short-term verbal memory - immediate (Digit series recall)	
	Short-term verbal memory - immediate (Non word recall)	
	Short-term verbal memory - immediate (Narrative story recall)	
	Short-term verbal memory - immediate (List learning)	
	Short-term auditory memory - immediate (tapping sequence)	
	Short-term visual memory - immediate (Immediate object recall)	
	Short-term visual memory - immediate (Immediate face recall)	
	Short-term verbal memory - delayed (delayed story recall)	
	Short-term verbal memory - delayed (delayed list learning)	
	Short-term visual memory - delayed (delayed face recall)	
	Visual working memory (all n-back (except ALSPAC have a comparable visual working memory test))	
	Verbal working memory (reverse digit subtests)	

Language	Production / articulation	Repetition, phonology
	Comprehension / reception	Instructions, vocabulary
	Expression	Naming, fluency, speech
	Reading	Identify letters, word reading
	Pragmatics	Social communication, gestures
	Bilingualism	
Executive functions	Attention	Novelty preference, sustained attention, selective attention, attentional control, vigilance
	Inhibition	Delay prepotent responses
	Working memory	Temporarily storing and managing the information required to carry out cognitive operations
	Speed processing	Reaction time, time to do a mental task
	Mental flexibility	Shift attention from one pattern to the other, respond effectively to new and complex situations, adaptability
	Sequencing	Follow a given sequence of numbers, letters…
	Calculation	Perform mathematical operations
	Problem solving	Goal-oriented complex actions, such as puzzles
	Decision making	Hot executive functions, influenced by emotions, gambling
Intelligence	Non verbal IQ	IQ no dependent of verbal abilities (acultural)
	General cognitive index	IQ composed by the sum of different cognitive abilities such us: verbal, perceptive, quantitative, memory and performance abilities
	Cognitive development	Development quotient
MRI	Structural	
	DTI	
	Functional (resting state)	functional organization and activity when individual is not performing a task
	Functional (task-related)	changes in oxygenation concentration (blood oxygen level dependent, or BOLD contrast) when individual given a task (visual, auditory, or other stimulus) to induce different neural states.

Instructions for Quality control

of Harmonized WP6 Variables

Step1: Verify list of variables and formats

Please, verify that your cohort-specific harmonized wp6 variables completely match with the information provided in the WP6 Variable List.

Each cohort needs to check that the *variable name* and *data type* of each of the variables correspond exactly to the WP6 Variable List (please, see table below). Also, please check the requirements for *type of harmonization*. Variables considered 'fully harmonized' must match the information provided in the Further instructions section. **Please, see the following examples of Step 1 in the Quality Control**:

Example: Internalizing symptoms total raw score

Variable name	Label/description	Values	Unit	Data Type	Comments	Further instructions
int_raw_0 int_raw_1 int_raw_2 … int_raw_17	Repeated measures of internalizing symptoms, total raw score: int_raw_0: assessed within one year of childbirth (>0 year and <1 year) int_raw_1: assessed when the child was aged between ≥ 1 year and < 2 years … int_raw_17: assessed when the child was aged between ≥ 17 years and < 18 years			Decimal	For SDQ, combine the emotional and peer items	The variable is fully harmonized if the score is based on all required items of the respective instrument. In case of score calculated using a short form of the test, please inform about it (including the items used) and categorize it as partially harmonized

1. Check the variable name (i.e. 'int_raw_17')
2. Check that values are decimal
3. Check 'Further instructions':
    a. Full Harmonization: if internalizing symptoms total raw score is based on all required items of the respective instrument
    b. Partial Harmonization: if internalizing symptoms total raw score is based on a short form of the test

Example: Age of ADHD Registry diagnosis assessment

Variable name	Label/description	Values	Unit	Data Type	Comments	Further instructions
adhdR_age_0 adhdR_age_1 adhdR_age_2 … adhdR_age_17	Exact age of the child (in years) at onset for ADHD diagnosis for adhdR_0, adhdR_1, etc.		years	Decimal	If the onset age is not known please use the follow-up age	The variable is partially harmonized if using age of the child at follow-up as age of onset

1. Check the variable name (i.e. 'adhdR_age_17')
2. Check that values are expressed in years
3. Check 'Comments' and 'Further instructions'
    a. Full Harmonization: if age of the child is based on age of onset of ADHD
    b. Partial Harmonization: if age of the child is based on follow-up age

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

Step 2: Check univariates distributions:

Please, generate distributions for all 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 <u>continuous variables</u>, check that there are no outliers, i.e. values out of the minimum and maximum range <u>based on the instrument used in each case</u>. 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.

Example: Internalizing problems raw total score based on SDQ instrument
Check whether values of int_raw_11 are in the expected range.

At INMA 11-12 years follow-up, SDQ was used to assess child behavioral and emotional problems. The SDQ consists of 25 items covering five subscales relating to emotional problems, peer problems, behavioral problems, hyperactivity and prosocial behavior. Each subscale comprises five questions with 3-point response scales ('Not true' = 0, 'Somewhat true' = 1, 'Certainly true' = 2), with a subscale score range of 0–10. 'Internalizing' and 'externalizing' are two alternative ten-item subscales with ranges of 0–20.

```
. sum int_raw_11

    Variable |        Obs        Mean    Std. Dev.        Min        Max
-------------+--------------------------------------------------------
   int_raw_11 |        276     3.57971     2.81247          0         15
```
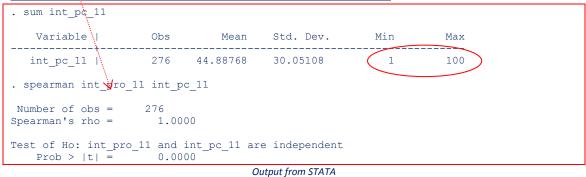*Output from STATA*

In addition to checking the distribution of <u>prorated total scale scores</u>, please check that Spearman correlation coefficient between prorated total scale score and its corresponding total raw score is ≈ 1.00.
Example: Internalizing problems prorated total scale score based on SDQ instrument

```
. spearman int_raw_11 int_pro_11

 Number of obs =      276
Spearman's rho =       1.0000

Test of Ho: int_raw_11 and int_pro_11 are independent
    Prob > |t| =       0.0000
```
Output from STATA

Note that for <u>percentiles</u>, the Spearmen correlation should be calculated using its corresponding prorated total scale score.
Example: Internalizing problems percentiles based on SDQ instrument

```
. sum int_pc_11

    Variable |        Obs        Mean    Std. Dev.        Min        Max
-------------+--------------------------------------------------------
    int_pc_11 |        276    44.88768    30.05108          1        100

. spearman int_pro_11 int_pc_11

 Number of obs =      276
Spearman's rho =       1.0000

Test of Ho: int_pro_11 and int_pc_11 are independent
    Prob > |t| =       0.0000
```
Output from STATA

** Remember, do not calculate the percentile variable if the number of observations with information available on prorated total scale score for a specific year is less than 10.

Example: ADHD assessed when the child was aged between ≥3 years and < 4 years

In the INMA cohort, there are only 7 children with prorated total scale score available between ages 3 and 4 years, so percentiles should not be calculated for year 3.

```
. sum adhd_pro_3 adhd_pc_3

    Variable |      Obs       Mean    Std. Dev.      Min       Max
-------------+--------------------------------------------------------
   adhd_pro_3 |       7    8.571429   10.79903        0        30
    adhd_pc_3 |       0
```
Output from STATA

For <u>categorical variables</u>, check that there are no improbable values, i.e. values not corresponding to the categories defined in the Variable List. Please, correct errors where relevant.

Example: ADHD Registry diagnosis evaluator

For ADHD Registry diagnosis evaluator, there are two possible categories following the WP6 Variable List:

Variable name	Label/description	Values	Unit	Data Type	Comments	Further instructions
adhdR_eval_0 adhdR_eval_1 adhdR_eval_2 … adhdR_eval_17	How was the medical diagnosis reported for adhdR_0, adhdR_1, etc.	1) Medical registries or medical records linkage 2) Parentally reported doctor-diagnosed disorder		Binary		

Check that data is coded into a maximum of two categories, and falls into the 1-2 (see output):

```
-> tabulation of adhdR_eval_4

        4 |
 adhdR_eval |      Freq.      Percent        Cum.
------------+-----------------------------------
        2 |      1,449       100.00       100.00
------------+-----------------------------------
    Total |      1,449       100.00
```
Output from STATA

Example: Internalizing problems instrument

For Internalizing problems instrument, there are six possible values following the WP6 Variable

Variable name	Label/description	Values	Unit	Data Type	Comments	Further instructions
int_instr_0 int_instr_1 int_instr_2 … int_instr_17	Cohort-specific instrument used to measure the internalizing symptoms for int_raw_0, int_raw_1, etc.	8) BPM 13) CBCL 34) ITSEA 47) SDQ 56) TRF 61) YSR		Categorical		

Check that data is coded in one of the above categories. SDQ was the instrument used at INMA 11-12 years follow-up:

```
. tab int_instr_11

       11 |
 int_instr |      Freq.      Percent        Cum.
------------+-----------------------------------
       47 |       276       100.00       100.00
------------+-----------------------------------
    Total |       276       100.00
```
Output from STATA

Step 3: Check internal validation

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.

For example, it is expected that male gender show more ADHD symptoms or child age is positively correlated with better cognitive performance. In addition to this, if more than one outcome were measured for a specific year and domain (for example, externalizing problems and ADHD), you could also check the spearman correlation between them, since we would expect a positive correlation.

Step 4: Check quality for repeated measures

In the WP6 Variable list, all variables are yearly repeated measures. Quality checking the repeated measures includes checking the consistency in time bands and order.

Note that the data is still in wide format. Check that *child_id* variable has no duplicates. It is important to check this, since sometimes, merging data you can create mistakes in this issue.

Check that exact age of assessment variables (*Y_age_t*) are in the corresponding age range, i.e. age interval *t* should include ages ≥ *t* and < *t+1*

<u>Example: Age of assessment of Internalizing symptoms in INMA cohort:</u>

```
. sum int_age_*, sep(0)

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   int_age_7 |        400    7.748864    .0701172    7.29911    7.994524
   int_age_8 |        385    8.656985    .2805814          8    8.999315
   int_age_9 |        477    9.321205     .230464    9.004791   9.982204
  int_age_10 |        504    10.62635    .2230522    10.00684   10.99795
  int_age_11 |        276    11.36488    .2621549    11.00068   11.98083
  int_age_12 |         23    12.20284    .1911812    12.00274   12.73648
```
Output from STATA

Check that the same instrument (Y_instr_t) is used for all children assessed within the same age range.

<u>Example: Internalizing symptoms were assessed using CBCL (code 13) at 7-8 and 9-10 years follow-ups; and SDQ (code 47) at 11-12 years follow-up.</u>

```
. sum int_instr_*, sep(0)

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 int_instr_7 |        400          13           0         13         13
 int_instr_8 |        385          13           0         13         13
 int_instr_9 |        477          13           0         13         13
int_instr_10 |        504          47           0         47         47
int_instr_11 |        276          47           0         47         47
int_instr_12 |         23          47           0         47         47
```
Output from STATA

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 harmonization is complete and information in all three tabs ('description, 'variables used and script syntax' are completed in full.

Please follow Step 5 of the instructions for quality control of WP1.