

Chapter 9. The final databases

9.1 Introduction

This chapter describes the official dataset for the *Danish Cross-sectional CDI Study*. The final dataset for the *Danish Cross-sectional CDI Study* (8 to 36 months of age) is based upon 2.398 children from the *CDI-Words and Gestures* (8 to 15: N= 1.537 and 16 to 20: N= 861) and 3.714 children from the *CDI-Words and Sentences* (16 to 30: N= 2.863 and 31 to 36: N= 851) which is a total of 6.112 children in the dataset.

9.2 Overview over all variables in the final dataset

The following section describes all variables in the final dataset for the *Danish Cross-sectional CDI Study*, presented here in the table, as an overview. Note that variables within a block are numbered from left to right; for example i_07_1 as the first item in Food/drink is “appelsin”, i_07_2 is “kaffe” and i_07_3 is “rosiner” etc.

Table 18: Overview of CDI variables in the final dataset¹⁷⁵

<i>CDI-Words and Gestures</i>			
Part and title	Variable names (first part)	Variable names (second part)	Labels and data types
Part 1 A First signs of understanding	i_00_*	01-03	1 yes, 2 no
Part 1 B Phrases	i_01_*	1-26	1 understood
Part 1 C Starting to talk	i_02_*	1,2	1 never, 2 sometimes, 3 often
Part 1 D <u>Vocabulary checklist:</u>		(n = 410)	
Sound effects and animal sounds	i_03_*	1-11	1 understood, 2 said
Animal names	i_04_*	1-36	“
Vehicles	i_05_*	1-10	“
Toys	i_06_*	1-8	“
Food and drink	i_07_*	1-28	“
Clothing	i_08_*	1-21	“
Body parts	i_09_*	1-20	“
Small household items	i_10_*	1-39	“
Furniture and rooms	i_11_*	1-24	“
Outside things	i_12_*	1-14	“
Places to go	i_13_*	1-14	“
People	i_14_*	1-30	“

¹⁷⁵ This table is based upon Appendix 1.39.

Games and routines	i_15_*	1-15	“
Action words	i_16_*	1-53 ¹⁷⁶	“
Descriptive words	i_17_*	1-36	“
Words about time	i_18_*	1-8	“
Pronouns	i_19_*	1-11	“
Question words	i_20_*	1-6	“
Prepositions and locations	i_21_*	1-16	“
Quantifiers	i_22_*	1-10	“
Part 2 A			
First communicative gestures	i_23_*	1-12	1 not yet, 2 sometimes, 3 often
Part 2 B			
Games and routines	i_24_*	1-6	1 yes, 2 no
Part 2 C			
Actions with objects	i_25_*	1-17	“
Part 2 D			
Pretending to be a parent	i_26_*	1-13	“
Part 2 E			
Imitating other adult actions	i_27_*	1-15	“
Part 2 F			
Pretend objects: examples			(string)
<i>CDI_parents_comment</i>			(string), Comments given by the parents
<i>CDI_data_comments</i>		•	• (string), comments given by the student assistants during data entry

<i>CDI-Words and Sentences</i>		•	•
Part and title	Variables names	• Variable numbers	• Labels and data types
Part 1 A		•	•
<u>Vocabulary checklist:</u>			
Sound effects and animal sounds	i_01_*	1-12	1 said
Animal names	i_02_*	1-43	”
Vehicles	i_03_*	1-14	”
Toys	i_04_*	1-18	”
Food and drink	i_05_*	1-68	”
Clothing	i_06_*	1-30	”
Body parts	i_07_*	1-28	”
Small household items	i_08_*	1-50	”
Furniture and rooms	i_09_*	1-33	”
Outside things	i_10_*	1-31	”
Places to go	i_11_*	1-22	”
People	i_12_*	1-40	”
Games and routines	i_13_*	1-27	”
Action words	i_14_*	1-103	”
Descriptive words	i_15_*	1-63	”
Words about time	i_16_*	1-15	”
Pronouns	i_17_*	1-31	”
Question words	i_18_*	1-7	”

¹⁷⁶ In the CDI-parental report, it is incorrectly noted n = 52 (the correct number of items: n = 53).

Prepositions and locations	i_19_*	1-41	”
Quantifiers	i_20_*	1-21	”
Helping verbs	i_21_*	1-21	”
Connecting words	i_22_*	1-7	”
Part 1 B	i_23_*	1-5	1 yes, 2 no
How children use words:			
Part 2 A	i_24_*	1-3	1 not yet, 2 sometimes, 3 often
Endings of words (part 1:)			
Part 2 B	i_25_*	1-29	1 said
Word forms			
Part 2 C	i_26_*	1-62	1 said
Endings of words (part 2:)			
Part 2 D			
Questions concerning combination of words	i_27_1		
Examples	example_*	1-3	1 not yet, 2 sometimes, 3 often (strings)
Part 2 E			
Complexity	i_28_*	1-33	(1,2; 2=complex version)
<i>CDI_parents_comment</i>		•	(string), Comments given by the parents
<i>CDI_data_comments</i>		•	• (string), comments given by the student assistants during data entry

Table 19: Overview of the general questionnaire variables in the final dataset¹⁷⁷:

= derived variables. Some variables have been generated in addition to the variables in the original datasets (“derived variables”). These can be identified via the “content & comments field” since there is reference to their definition (“cf. step 4”).

Variable name	Labels, data types, or range	Content and comments
Essential variables		
<i>barn_id#</i>	(string), examples: I0001, UI0001, T0001, UT0001	cf. Step ¹⁷⁸ 1-4; Identity, identification variable
<i>months#</i>	8-36	cf. Step1-4; age (in months)
<i>monthdate#</i>	(Stata date format)	cf. Step1-4; Date of receiving filled-in CDI form
<i>months_diff#</i>	(float)	cf. Step1-4; Difference between <i>monthdate</i> and <i>birthdate</i>
<i>months_error#</i>	(byte), different variable labels for Infant and Toddler part (<i>errl</i>)	cf. Step1-4; Categorised difference of <i>months_diff</i>

¹⁷⁷ Note that variables annotated with a (!) are not in the official output version of the final datasets (PUBLICDATA).

¹⁷⁸ The step numbers mentioned in this table refer to Appendix 1.39.

Selection criteria		
<i>inout_all#</i>	<ul style="list-style-type: none"> 0 “OUT”, 1 “IN” 	cf. Step 8; Combination of following In/Out variables
<i>inout_ol#</i>	0 “OUT”, 1 “IN”	cf. Step 8; OL= Contact to other languages
<i>inout_sp#</i>	0 “OUT”, 1 “IN”	cf. Step 8; SP= Speech problems
<i>inout_hp#</i>	<ul style="list-style-type: none"> 0 “OUT”, 1 “IN” 	cf. Step 8; HP= Health problems
<i>inout_CDI#</i>	<ul style="list-style-type: none"> 0 “OUT”, 1 “IN” 	cf. Step 8; CDI=Completion of CDI
<i>inout_del#</i>	<ul style="list-style-type: none"> 0 “OUT”, 1 “IN” 	cf. Step 8; del= Completion of GQ
<i>ears_prob_sp#</i>	<ul style="list-style-type: none"> 0 “”, 1 “Ear problems” 	cf. Step 8; ear problems reported as speech problems
<i>ears_prob_hp#</i>	<ul style="list-style-type: none"> 0 “”, 1 “Ear problems” 	cf. Step 8; ; ear problems reported as health problems
<i>contact_other_languages</i>	(string: categorized)	
<i>comment_contact_ol</i>	(string)	
<i>speech_problems_child</i>	(string: categorized)	
<i>comment_speech_pc</i>	(string)	
<i>health_problems_child</i>	(string: categorized)	
<i>comment_health_pc</i>	<ul style="list-style-type: none"> (string) 	
Personal information on the child		
<i>sex</i>	1 “male”, 2 “female”	1 missing set to male (Emil), I0366
<i>region</i>	1 “Fyn”, 2 “Jylland”, 3 “Sjælland”	
<i>!name_child</i>	(string)	
<i>!home_child</i>	(string)	City
<i>!birthdate</i>	(Stata date format)	
Sibling information		
<i>sibling</i>	(byte)	Number of siblings
<i>siborder</i>	(byte)	Position of child (first, second etc. of siblings)
<i>sex_1_child sex_2_child sex_3_child sex_4_child</i>	<ul style="list-style-type: none"> 	Siblings may not always be sorted after age (as intended: sibling 1 as next youngest, etc.). Possible further siblings should have been mentioned in <i>deltager_comments_typist</i>
<i>age_1_child age_2_child</i>	<ul style="list-style-type: none"> 	see above

age_3_child age_4_child		
Personal information on parents		
birthyear_mother#	• (int)	cf. Step 5
hometown_mother	• (string)	only for 8-15 (INF), 16-30 (TOD)
homeland_mother	• (string)	only for 16-20 (INF), 31-36 (TOD)
!homeregion_mother	• (string)	only for 16-20 (INF), 31-36 (TOD)
other_country_mother	• (string)	only for 16-20 (INF), 31-36 (TOD)
!name_mother, !street_mother, !home_mother, !postcode_mother, !phone_mother, !phone_work_mother		
!birthdate_mother	(string)	
!birth_mother#	(Stata date format)	cf. Step 5

Birth and parental leave times of parents		
<i>birth_leave_class_mother#</i>	(float), in months, 0.5 = 14 days	cf. Step 5
<i>parental_leave_class_mother#</i>	(float), in months, 0.5 = 14 days	cf. Step 5
<i>start_age_leave_class#</i>	(float), in months, 0.5 = 14 days	cf. Step 5; age of child when leave started; only for 8-15 (INF), 16-30 (TOD); not separately for parents
<i>start_age_leave_class_mother#</i>	(float), in months, 0.5 = 14 days	cf. Step 5; only for 16-20 (INF), 31-36 (TOD); separately for parents
<i>birth_leave_mother</i>	(string)	
<i>birth_leave_date_mother</i>	(string: DD(MM/YYYY))	only for 16-20 (INF), 31-36 (TOD)
<i>parental_leave_mother</i>	(string)	
<i>parental_leave_date_mother</i>	<ul style="list-style-type: none"> (string: DD/M M/YY YY) 	only for 16-20 (INF), 31-36 (TOD)
<i>child_age_start_parental_leave</i>	<ul style="list-style-type: none"> (string) 	only for 8-15 (INF), 16-30 (TOD); not separately for parents
<i>child_age_start_parental_leave_mother</i>	<ul style="list-style-type: none"> (string) 	only for 16-20 (INF), 31-36 (TOD); separately for parents
Working status, education and occupation of parents ¹⁷⁹		
<i>status_class_mother#</i>	1 "Unemployed", 2 "Working home", 3 "Employed", 4 "Vacation / on leave", 5 "Retired / pension", 6 "Student", 7 "Illness", 998 "Unknown/other", 999 "Missing"	cf. Step 5; this variable is not checked in connection with the working hours (Step 6) or the daycare situation of the child (Step 7), thus the working status might be "employed", even if working hours are zero and the daycare situation of the child is "at home"
<i>edu_class_mother#</i>	1 "Grundskole", 2 "Grundskole - Gymnasial U", 3 "Grundskole – Kortere VU", 4 "Gymnasial U", 5 "Gymnasial U - Kortere VU", 6 "Gymnasial – Lang VU", 7 "Kortere VU", 8 "Kort - mellemlang VU", 9 "Kort - Lang VU", 10 "Mellemlang VU", 11 "Mellemlang - Lang VU", 12 "Lang VU", 998 "Unknown/other", 999 "Missing"	cf. Step 5
<i>edu_class_total#</i>	<ul style="list-style-type: none"> “ 	cf. Step 5; "highest" education of mother and father
<i>e1_mother e2_mother eclass_mother further_education_mother e3_mother e4_mother and eclass_further_mother</i>	<ul style="list-style-type: none"> (byte: <i>eX_mother</i>) (string: all other variables) 	Due to students' coding process
<i>status_mother</i>	(string)	
<i>education_mother</i>	(string)	Information from parents

¹⁷⁹ For more details, see chapter 4.4. See also footnote 83.

<i>occu_class_mother#</i>	1 "Selvstændig", 2 "Selvstændig - Højeste N", 3 "Selvstændig - Mellemste N", 4 "Selvstændig – Grundniveau", 5 "Selvstændig – Ukendt", 6 "Højeste N", 7 "Højeste N - Mellemste N", 8 "Højeste N – Grundniveau", 9 "Højeste N – Ukendt", 10 "Mellemste N", 11 "Mellemste N – Grundniveau", 12 "Mellemste N – Ukendt", 13 "Grundniveau", 14 "Grundniveau – Ukendt", 998 "Unknown/other", 999 "Missing"	cf. Step 5
<i>occu_class_total#</i>	“	cf. Step 5, "highest" occupation of mother and father
<i>work_hours_class_mother#</i>	0 "0 hours", 1 "1-10", 2 "11-20", 3 "21-30", 4 "31-40", 5 "41-50", 6 "51-60", 7 "61-70", 8 "71-80", 9 "81+", 998 "Unknown/other", 999 "Missing"	cf. Step 6; in <i>work_hours_class</i> , the current working status of the parents is recorded. For analysis purposes, it is recommended to use both <i>status_class_*</i> and <i>work_hours_class_*</i> .
<i>occupation_mother</i>	(string)	Information from parents
<i>work_hours_mother</i>	• (string)	
<i>o1_mother, o2_mother, and oclass_mother</i>	(byte: <i>oX_mother</i>) (string: all other variables)	Due to students' coding process
Daycare		
<i>daycare#</i>	0 "At home", 1 "At home: mother/father", 2 "At home: barnepige/au-pair", 3 "At home: grandparents", 4 "At home: daycare (mother herself)", 5 "Daycare (general)", 6 "Daycare: privat (dagpleje)", 7 "Daycare: Integrated Institution (0-6)", 8 "Daycare: Day Nursery (0-3)", 9 "Daycare: Kindergarden (3-6)", 998 "Unclear/other", 999 "Missing"	cf. Step 7, see "Notes to codes" there. The current status of the daycare situation is recorded, combined with <i>start_in_inst</i> , it should be possible to determine the length of the stay.
<i>daycare_original</i>	(string)	AD variable v76 is renamed v72 → 2 nd phase.
<i>daycare_checked#</i>	0/1	cf. Step 7.
<i>daycare_change#</i>	0 "still at home", 1 "In 1. institution", 2 "In 2. institution", 3 "From 1. institution back home", 4 "In 1. institution, change within home before", 5 "More than 2 changes", 998 "Unclear/other", 999 "Missing"	cf. Step 7. <i>daycare1, daycare_change_date1, hours_inst_class1, etc.</i>

Information on institution		
<i>start_in_institution</i>	(float)	
<i>hours_inst_class#</i>	1 "0-15", 2 "15-20", 3 "20-25", 4 "25-30", 5 "30-35", 6 "35-40", 7 "40-45", 8 "45-", 999 "Missing"	cf. Step 5; <i>hours_inst_class1</i> etc. for children with changing status, see Day-care (Step 7)
<i>size_inst_class#</i>	1 "0-25", 2 "25-50", 3 "50-75", 4 "75-100", 5 ">= 100", 999 "missing"	cf. Step 5; <i>size_inst_class1</i> etc. for children with changing status; see Day-care (Step 7). In different versions of the GQ, there are different categories e.g. "0-24" and "0-25"; solution: both "0-24" and "0-25" are summarized in the category "0-25" etc.
<i>size_inst_class#</i>	1 "0-25", 2 "25-50", 3 "50-75", 4 "75-100", 5 ">= 100", 999 "missing"	cf. Step 5; <i>size_inst_class1</i> etc. for children with changing status; see Day-care (Step 7). In different versions of the GQ, there are different categories e.g. "0-24" and "0-25"; solution: both "0-24" and "0-25" are summarized in the category "0-25" etc.
<i>size_group_class#</i>	1 "0-9", 2 "10-19", 3 ">= 20", 999 "Missing"	cf. Step 5; <i>size_group_class1</i> etc. for children with changing status; see Day-care (Step 7).
<i>size_group_additional#</i>	(string)	cf. Step 7; based on more specific information in <i>deltager_comments_* etc</i> (<i>size_group_additional1</i> etc. for children with changing status)
<i>place_inst_class#</i>	1 "little traffic", 2 "intermediate traffic", 3 "much traffic", 4 "other", 5 "comments", 999 "Missing"	cf. Step 5; before using this variable for analysis, the observations in "comments" should be classified in the existing groups, if possible.
<i>noise_class#</i>	1 "Yes", 2 "No", 999 "Missing"	cf. Step 5
<i>child_living</i>	(string: categorised)	
<i>passer_barn</i>	(string)	only in original data → 1st phase
<i>hours_institution</i>	(string: categorised)	
<i>size_of_institution</i>	(string: categorised)	
<i>size_of_group</i>	(string: categorised)	
<i>place_of_institution</i>	(string)	
<i>comment_place_i</i>	(string)	
<i>noise_policy</i>	(string: categorised)	
<i>noise_reduction</i>	(string: categorised)	
<i>comment_noise</i>	(string)	
<i>comment_unknown1</i>	(string)	Content not clear

Other		
<i>song_class#</i>	1 “daily”, 2 “3-4 times per week”, 3 “seldom”, 999 “Missing”	cf. Step 5
<i>song_and_reading</i>	(string: categorised)	
<i>deltager_comments_parents</i>	(string)	
<i>deltager_comments_typist</i>	(string: sorted after code words)	
<i>AD_v111-AD_v119</i>		From additional data sets (2 nd phase); content not clear