

Data documentation for the Norwegian version of the MacArthur-Bates Communicative Development Inventories (CDI)

In the following, the following abbreviations will be used:

W&G	CDI: Words and gestures
W&S	CDI: Words and sentences
TKU	Bleses, D., Vach, W., Wehberg, S., Faber, K. & Madsen. T. O. (2007). <i>Tidlig kommunikativ udvikling. Et værktøj til beskrivelse af sprogtilegnelse baseret på CDI-forælder rapportundersøgelser af danske normalhørende og hørehæmmede børn</i> . Syddansk Universitetsforlag

Contents

1. Dataset description

- 1.1 Extraction of datasets
- 1.2 Selection rules
- 1.3 Description of W&G CDI datasets
- 1.4 Description of W&S CDI datasets
- 1.5 Description of socio-demographic datasets

2. Output for W&G: Tables and figures

3. Output for W&S: Tables and figures

4. Open issues and recommendations

Some general comments

- For the programming, Stata version 9 was used
- Variable, programs, and dataset names are in italics
- Strings have a maximal length of 244 characters in the dataset, hence, some information may be cut off arbitrarily.
- All datasets are available both as .csv and .dta (Stata) datasets (version 9).

1. Dataset description

1.1 Extraction of datasets

The following .csv datasets have been extracted from the database on the 8th of March 2009:

File name	Number of observations	Number of variables
<i>scl_replies_080309.csv</i>	7,577	46
<i>cdi_1_replies_080309.csv</i>	3,408	498
<i>cdi_2_replies_080309.csv</i>	5,844	881

The identification variable *id_scl* is there for all observations (that is, no observations have to be excluded due to missing *id*). Data has been collected in two (closed) waves¹: 552 children have two W&G-CDI observations, 718 children have two W&S observations and 367 have both an W&G and an W&S observation. In all cases, only the first observation per child was included and the second (later) observations are excluded². This yields 2,699 W&G observations and 4,856 W&S observations in the RAW datasets (overall 7,555) (Stata logfile: LOG_INdata_140309.log).

22 children have only registered socio-demographic data without CDI data (see also footnote 1); these are not considered further (Stata logfile: LOG_socio_data_140309.log).

For W&G, 2,017 observations come from the first wave (data until 01-01-09) and 682 observations from the second wave (data until 28-02-09). For W&S, 3,298 observations come from the first wave (data until 01-01-09) and 1,558 observations from the second wave (data until 28-02-09). In the following, data from both waves will be treated equally.

1 16 W&G observations and 44 W&S observations from the (open) wave 28-02-10 are excluded at once.

2 Note that more than one observation per child could be included in the same wave at the same monthly age at the same filled-out date: then W&G is chosen before W&S. Note also that there are children where an W&S observations comes before an W&G observation.

1.2 Selection rules

The following general selection rules for inclusion into the FINAL datasets (both W&G and W&S) have been applied:

1. Contact to other languages has to be less than “frequent”³
2. Birth has occurred at full term (> 36 weeks)
3. Combined hospital stay should be less than 4 weeks
4. There has to be no serious, well-founded concern of the parents regarding their child's “normal” language development
5. At least one question has to be answered in *scl_replies_080309.csv*
6. At least one question has to be answered in either *cdi_1_replies_080309* or *cdi_2_replies_080309*
7. Age has to be between 8-20 months for W&G and between 16-36 months for W&S
8. CDI data must be appropriate (for example, there seems to be a database mix-up for *id* = 6405 and *id* = 9991)

In total, 340 W&G_children are excluded, leaving 2,359 children in the FINAL W&G dataset. In total, 641 W&S_children are excluded, leaving 4,215 children in the FINAL W&S dataset. Details are described in the following (corresponding Stata logfile: LOG_socio_data_140309.log).

For each child the following variables were defined:

- *in_out_cl* (contact to other languages)
- *in_out_bo* (child lived abroad for more than 6 months)
- *in_out_pb* (premature birth)
- *in_out_hs* (hospital stay)
- *in_out_pw* (parental concern)
- *in_out_missing* (socio-demographic questions are not answered at all)
- *in_out_CDI* (CDI questions are not answered at all)
- *in_out_age* (age range does not fit)
- *in_out_add* (additional exclusions due to strange data)
- *in_out_comments* (exclusion criteria relevant information mentioned in parents' comments)
- *in_out* (overall)

All *in_out* variables are 0/1 variables, where *in_out_** == 1 means exclusion criteria fulfilled, that is, the child is excluded from the FINAL dataset.

3 We assume a frequent contact to another language if the child lived outside Norway for more than 6 months.

1.2.1 Contact to other languages and Living abroad

The leading question “1.a Er barnet regelmessig i kontakt med andre språk enn norsk?” is answered as follows:

i_3_1	Freq.	Percent	Cum.
Ingen kontakt	5,946	78.70	78.70
Noe kontakt	1,067	14.12	92.83
Hyppig kontakt	441	5.84	98.66
.	101	1.34	100.00
Total	7,555	100.00	

General rules:

Observations with “ingen kontakt” or “noe kontakt” are not excluded unless text in the corresponding text field “1.b Hvis det er krysset av for “Noe kontakt” eller “hyppig kontakt” under 1a, oppgi hvilket/hvilke språk” indicates contact to sign language (tegn to tale, norsk tegnsprog). Observations with “hyppig kontakt” are excluded. Observations with missing values are (conservatively) excluded if the text field was filled.

In detail:

For observations with “ingen kontakt”, the text field is filled for 5 children. The child *id* == 7200 is excluded (text “tegn til tale”). For observations with “noe kontakt”, the text field is filled for 1,054 children. 14 children receive *in_out_cl* == 1 because there is a reference to sign language (*id* == 28, 844, 2843, 4946, 6077, 8885, 9199, 10929, 12000, 12117, 14599, 15139, 15218, 18620). All 441 children with “hyppig kontakt” are excluded. Two observations with missing values (*id* = 7377, 15650) are excluded since the text field is filled.

In all, 458 receive the exclusion code *in_out_cl* == 1 (160 W&G, 298 W&S) (Stata logfile: LOG_socio_data_140309.log).

19 children (*id* = 1811, 4767, 6016, 7456, 8449, 10320, 10346, 11375, 12654, 12920, 13243, 15261, 15365, 16805, 17194, 18518, 18732, 18759, 20072) receive *in_out_bo* == 1, because they lived abroad for more than 6 months (4 W&G, 15 W&S) (Stata logfile: LOG_socio_data_140309.log).

.

1.2.2 Premature birth

The leading question “2. Er barnet for tidlig født?” is answered as follows:

i_3_3	Freq.	Percent	Cum.
	68	0.90	0.90
Ja	1,217	16.11	17.01
Nei	6,270	82.99	100.00
Total	7,555	100.00	

General rule:

Children who were born in or before week 36 are excluded (that is, information "36 + 6 days" leads to exclusion). In addition, all children who were born after week 44 (45+0 and thereafter) are excluded.

In detail:

For observations with “missing value”, the corresponding text field “2b. Hvis det er krysset av for “Ja”, oppgi i hvilken svangerskapsuke det er født” is filled for 5 children (child *id* = 13231 reads “36” and is excluded). 17 observations with answer “Nei” added text: one child is excluded after a manual check (*id* = 13930 reads “9”). 1,197 observations with answer “Ja” added text: a manual check resulted in exclusion of 444 children⁴. Thereof, information on child with *id* = 12083 reads “24-02-09”, but the same information in an older file reads “24-2”⁵. This child is excluded. Information on child with *id* = 10752 reads “50”. This child is excluded as well.

In all, 446 children receive the exclusion code *in_out_pb* == 1 (163 W&G, 283 W&S) (Stata logfile: LOG_socio_data_140309.log).

Note:

There are some twin children in the sample (who could be identified via variable *i_2_2*). There is a larger possibility to exclude twins with respect to premature birth.

4 All commas are substituted from text fields in the data, implying that original information as “37,5” now reads “37.5”.

5 Extracting .csv files from the database resulted in incorrect formats (time formats as 01-05-09 instead of 1-5 or 1.5) on at least the variables *i_3_4* and *i_3_6*.

1.2.3 Hospital stay

The leading question “3. *Har barnet vært innlagt på sykehus?*” is answered as follows:

i_3_5	Freq.	Percent	Cum.
	129	1.71	1.71
Ja	1,060	14.03	15.74
Nei	6,366	84.26	100.00
Total	7,555	100.00	

General rule:

Children who were in hospital (in total) 4 weeks or longer are excluded.

In detail:

For observations with “missing value”, the corresponding text field “3b. *Hvis det er krysset av for “Ja”, oppgi hvor lenge (ca.) i antall uker*” is filled for 3 children: none is excluded after a manual check. 4 observations with answer “Nei” added text : none is excluded after a manual check. 1,029 observations with answer “Ja” added text: manual check resulted in exclusion of 123 children⁶. Children are not excluded, if the answer was "Ja" but no text was added.

In all, 123 children receive the exclusion code `in_out_hs == 1` (32 W&G, 91 W&S) (Stata logfile: LOG_socio_data_140309.log).

6 Entries as "06-05-09" originated from "65" and are treated accordingly, see previous footnote.

1.2.4 Parental concern

The leading question “5. *Har du vært bekymret for ditt barns språkutvikling og tatt kontakt med fastlege eller helsestasjon om dette?*” was answered as follows:

i_3_8	Freq.	Percent	Cum.
	86	1.14	1.14
Ja	309	4.09	5.23
Nei	7,160	94.77	100.00
Total	7,555	100.00	

General rules:

Children are excluded if the parents answered “Ja” to the leading question and did not fill the text field. In addition, children are excluded when parents added text and this text refers to (1) physical or mental disadvantages in learning to speak (2) daycare personnel raising concern (3) deafness of the child (but not: one-sided deafness or limited hearing because of ear infections etc.), see complete list in Stata logfile LOG_socio_data_140309.

Children are **not** excluded for the following reasons (examples): frequent ear infections, late talking, limited hearing due to “væske i øret” or “tette øre”, “dren”, astma

In detail:

For observations with “missing value”, the corresponding text field “5b. *Hvis det er krysset av for ”Ja”, beskriv kort hva du har vært bekymret for*” is filled for 1 child (information reads “At han kan høre dårlig eller evt ha nedsatt hørsel eller mye ørevoks rett og slett”). This child is not excluded. 7 observations with answer “Nei” added text: one child (*id* == 16408) is excluded after a manual check (text reads “barnet er født med isolert ganespalte (...)). 301 observations with answer “Ja” added text: manual check resulted in exclusion of 31 children. The 8 children whose parents answered “Ja” to the leading question but did not add text, are excluded as well (*id* == 4197, 5895, 7148, 7917, 10348, 10914, 13667, 18462).

In all, 40 received the exclusion code *in_out_pw* == 1 (9 W&G, 31 W&S).

1.2.5 At least one question has to be answered in scl_replies_080309.csv

All children answer at least 1 question in *scl_replies_080309.csv*. No children receive the code *in_out_missing == 1* (Stata logfile: LOG_socio_data_140309.log).

1.2.6 At least one question has to be answered in either cdi_1_replies_080309 or cdi_2_replies_080309

For all children at least 1 question was filled in W&G (*cdi_1_replies_080309*). For 4 children nothing was filled in W&S (*cdi_2_replies_080309*), *id* = 11754, 13286, 15816, 17218. These children receive *in_out_CDI* = 1 (Stata Logfiles: LOG_cross1_data_140309 and LOG_cross2_data_140309).

1.2.7 Children's age lies in the appropriate range (8-20 months for W&G, 16-36 for W&S)

3 children lie outside the age range for W&G (*cdi_1_replies_080309*), *id* = 8727, 8890, 9636, 10426. These children receive *in_out_age* = 1. All children lie within the appropriate age range in W&S (Stata Logfiles: LOG_cross1_data_140309 and LOG_cross2_data_140309).

1.2.8 Strange mix-up with data observations

1 child seems to have a W&S observation although reported in W&G (*cdi_1_replies_080309*), *id* = 6405. This child receives *in_out_add* = 1 (Stata Logfile: LOG_cross1_data_140309).

1.2.9 Exclusion due to parents' comments (only available in W&S)

5 W&S children receive *in_out_comments* = 1 due to their parents' comments (*cdi_2_replies_080309*), *id* = 8722, 9218, 15523, 15660, 18800. Here only children were manually checked who had not been given *in_out* = 1 before (Stata Logfile: LOG_WScomments_140309).

1.3 Description of W&G CDI datasets

There are two W&G CDI datasets: *WG_CDI_RAW_140309* and *WG_CDI_FINAL_140309*.

WG_CDI_RAW_140309 is based on *all* W&G observations with identifiable ID (identification *id*) and age (*months*). It includes the variables *id* (identification), *months* (age), *sex* etc., all original CDI variables, some derived CDI variables and all new variables *in_out_** which identify children who are not included in the final W&G CDI dataset (see below).

In comparison to *WG_CDI_RAW_140309*, some observations (children) are excluded in *WG_CDI_FINAL_140309*. The *in_out* variables are not available here.

Both datasets contain the variables from the first three table blocks below. Only *WG_CDI_RAW_140309* contain the 10 *in_out* variables (block 4 below) which are relevant for exclusion from the final dataset *WG_CDI_FINAL_140309*.

Number of observations in *WG_CDI_RAW_140309*: 2,699 (557 variables)

Number of observations in *WG_CDI_FINAL_140309*: 2,359 (547 variables)

Note that variables within a block are numbered from left to right, not column-wise.

Block 1:

Part	Name	Variable names	Variable numbers	Labels /values
		<i>id</i>		id number (integer)
		<i>months</i>		8-20
		<i>sex</i>		0 boys, 1 girls
		<i>id_child</i>		id number (integer)
		<i>id_cdi</i>		id number (integer)
		<i>wave</i>		(string)
		<i>cdi diversion</i>		INFANT
		<i>birthday</i>		time variable, fx. 07dec2008
		<i>filledout</i>		time variable, fx. 07dec2008

Block 2⁷:

Part	Name	Variable names	Variable numbers	Labels /values
Part 1A	De første tegn på forståelse	<i>i_1_*</i>	1-3	0 no , 1 yes
Part 1B	Faste uttrykk	<i>i_2_*</i>	1-26	1 understood
Part 1C	Begynner å snakke	<i>i_3_*</i>	1-2	1 never, 2 sometimes, 3 often
Part 1D	Sjekkliste over ordforrådet		(n=395)	
	1. Lydeffekter og dyrellyder	<i>i_4_*</i>	1-11	1 understood, 2 said
	2. Navn på dyr (virkelige eller lekedyr)	<i>i_5_*</i>	1-36	“
	3. Kjøretøy (virkelige eller leketøy)	<i>i_6_*</i>	1-9	“
	4. Leker	<i>i_7_*</i>	1-8	“
	5. Mat og drikke	<i>i_8_*</i>	1-28	“
	6. Klær og tilbehør	<i>i_9_*</i>	1-20	“
	7. Kroppsdeler	<i>i_10_*</i>	1-20	“
	8. Små gjenstander i huset	<i>i_11_*</i>	1-35	“
	9. Rom og møbler	<i>i_12_*</i>	1-23	“
	10. Ting utendørs	<i>i_13_*</i>	1-14	“
	11. Steder å dra til	<i>i_14_*</i>	1-12	“
	12. Mennesker	<i>i_15_*</i>	1-22	“
	13. Leker og rutiner	<i>i_16_*</i>	1-18	“
	14. Verb	<i>i_17_*</i>	1-53	“
	15. Beskrivende ord	<i>i_18_*</i>	1-36	“
	16. Ord om tid	<i>i_19_*</i>	1-8	“
	17. Pronomen	<i>i_20_*</i>	1-11	“
	18. Spørreord	<i>i_21_*</i>	1-6	“
	19. Preposisjoner og stedsangivelser	<i>i_22_*</i>	1-16	“
	20. Mengdeord og bestemmerord	<i>i_23_*</i>	1-9	“
Part 2A	De første kommuniserende faktene /gestene	<i>i_24_*</i>	1-12	1 not yet, 2 sometimes, 3 often
Part 2B	Leker og rutiner	<i>i_25_*</i>	1-6	0 no, 1 yes
Part 2C	Handling med objekter	<i>i_26_*</i>	1-17	“
Part 2D	Å leke at man er mor eller far	<i>i_27_*</i>	1-13	“
Part 2E	Herme etter det de voksne gjør	<i>i_28_*</i>	1-15	“

7 Note that the child id = 6405 does not fit in the W&G CDI scheme. This child is included in the RAW dataset, but excluded from the FINAL (*in_out_add* = 1).

Block 3 (derived CDI variables; see LOG_cross1_data_140309):

Part	Name	Variable names	Variable numbers	Labels /values
Part 1B	Faste uttrykk	<i>score1B</i>		0-26
Part 1D	Sjekkliste over ordforrådet	<i>PCscore</i>	4-23	see total number in categories as above
Part 1D	Sjekkliste over ordforrådet	<i>PRscore</i>	4-23	see total number in categories as above
Part 1D	Sjekkliste over ordforrådet	<i>score1D_PC</i>		0-395
Part 1D	Sjekkliste over ordforrådet	<i>score1D_PR</i>		0-395
Part 2A-2E	Aktiviteter og gestik	<i>sum_</i>	24-28	0-12, 0-6, 0-17, 0-13, 0-15
Part 2A-2E	Aktiviteter og gestik	<i>score2</i>		0-63

Block 4:

Part	Name	Variable names	Variable numbers	Labels /values
		<i>in_out_bo</i>		0 no, 1 yes (=exclusion)
		<i>in_out_cl</i>		0 no, 1 yes (=exclusion)
		<i>in_out_pb</i>		0 no, 1 yes (=exclusion)
		<i>in_out_hs</i>		0 no, 1 yes (=exclusion)
		<i>in_out_pw</i>		0 no, 1 yes (=exclusion)
		<i>in_out_missing</i>		0 no, 1 yes (=exclusion)
		<i>in_out_CDI</i>		0 no, 1 yes (=exclusion)
		<i>in_out_age</i>		0 no, 1 yes (=exclusion)
		<i>in_out_add</i>		0 no, 1 yes (=exclusion)
		<i>in_out</i>		0 no, 1 yes (=exclusion)

1.4 Description of W&S CDI datasets

There are two W&S CDI datasets: *WS_CDI_RAW_140309* and *WS_CDI_FINAL_140309*.

WS_CDI_RAW_140309 is based on *all* W&S observations with identifiable ID (identification *id*) and age (*months*). It includes the variables *id* (identification), *months* (age), *sex* etc., all original CDI variables, some derived CDI variables and all new variables *in_out_** which identify children who are not included in the final W&S CDI dataset (see below).

In comparison to *WS_CDI_RAW_140309*, some observations (children) are excluded in *WS_CDI_FINAL_140309*. The *in_out* variables are not available here.

Both datasets contain the variables from the first three table blocks below. Only *WS_CDI_RAW_140309* contain the 10 *in_out* variables (block 4 below) which are relevant for exclusion from the final dataset *WS_CDI_FINAL_140309*.

Number of observations in *WS_CDI_RAW_140309* : 4,856 (926 variables)

Number of observations in *WS_CDI_FINAL_140309*: 4,215 (915 variables)

Note that variables within a block are numbered from left to right, not column-wise.

Block 1:

Part	Name	Variable names	Variable numbers	Labels /values
		<i>id</i>		id number (integer)
		<i>months</i>		16-36
		<i>sex</i>		0 boys, 1 girls
		<i>id_child</i>		id number (integer)
		<i>id_cdi</i>		id number (integer)
		<i>wave</i>		(string)
		<i>cdi diversion</i>		TODDLER
		<i>birthday</i>		time variable, fx. 07dec2008
		<i>filledout</i>		time variable, fx. 07dec2008

Block 2⁸

Part	Name	Variable names	Variable numbers	Labels /values
Part 1A	Sjekkliste over ordforrådet		(n=731)	
	1. Lydeffekter og dyreløyer	<i>i_1_*</i>	1-12	1 said
	2. Navn på dyr (virkelige eller lekedyr)	<i>i_2_*</i>	1-44	"
	3. Kjøretøy (virkelige eller leketøy)	<i>i_3_*</i>	1-14	"
	4. Leker	<i>i_4_*</i>	1-18	"
	5. Mat og drikke	<i>i_5_*</i>	1-68	"
	6. Klær og tilbehør	<i>i_6_*</i>	1-30	"
	7. Kroppsdeler	<i>i_7_*</i>	1-27	"
	8. Små gjenstander i huset	<i>i_8_*</i>	1-50	"
	9. Rom og møbler	<i>i_9_*</i>	1-34	"
	10. Ting utendørs	<i>i_10_*</i>	1-31	"
	11. Steder å dra til	<i>i_11_*</i>	1-22	"
	12. Mennesker	<i>i_12_*</i>	1-36	"
	13. Leker og rutiner	<i>i_13_*</i>	1-27	"
	14. Verb	<i>i_14_*</i>	1-107	"
	15. Beskrivende ord	<i>i_15_*</i>	1-63	"
	16. Ord om tid	<i>i_16_*</i>	1-16	"
	17. Pronomen	<i>i_17_*</i>	1-31	"
	18. Spørreord	<i>i_18_*</i>	1-7	"
	19. Preposisjoner og stedsangivelser	<i>i_19_*</i>	1-41	"
	20. Mengdeord og bestemmerord	<i>i_20_*</i>	1-22	"
	21. Hjelpeverb og verb med generelt innhold	<i>i_21_*</i>	1-22	"
	22. Bindeord	<i>i_22_*</i>	1-9	"
Part 1B	Hvordan barn bruker ord	<i>i_23_*</i>	1-5	1 not yet, 2 sometimes, 3 often

8 Note that the child with id = 9991 does not fit the W&S CDI scheme. The child is included in the RAW dataset, but excluded from the FINAL (*in_out_add* = 1).

Part	Name	Variable names	Variable numbers	Labels /values
Part 2A	Bøyningsendelser del 1	<i>i_24_*</i>	1-6	1 not yet, 2 sometimes, 3 often
Part 2B	Ordformer - Substantiv (flertallsformer)	<i>i_25_*</i>	1-8	1 said
Part 2B	Ordformer - Verb (fortidsformer)	<i>i_26_*</i>	1-25	“
Part 2C	Bøyningsendelser del 2 - Substantiv (flertalls- og bestemthetsformer)	<i>i_27_*</i>	1-13	“
Part 2C	Bøyningsendelser del 2 - Verb (fortidsformer)	<i>i_28_*</i>	1-37	“
	Har barnet ditt begynt å sette sammen ord slik som "mer kake" eller pappa Lars" (pappa til Lars)?	<i>i_29_1</i>	1	1 not yet, 2 sometimes, 3 often
Part 2D	Eksempler	<i>i_30_*</i>	1-3	(strings)
Part 2E	Kompleksitet	<i>i_31_*</i>	1-42	1 simple form, 2 complex form
	Parents' comments	<i>i_32_1</i>		(string)

Block 3 (derived CDI variables; see LOG_cross2_data_140309):

Part	Name	Variable names	Variable numbers	Labels /values
Part 1A	Sjekkliste over ordforrådet	<i>PRscore</i>	1-22	see total number in categories as above
Part 1A	Sjekkliste over ordforrådet	<i>score1A_PR</i>		0-731
Part 1B	Hvordan barn bruker ord	<i>new23_*</i>	1-5	0 no, 1 yes (=combined 2+3 before)
Part 2B+ C		<i>PRscore</i>	25-28	0-
Part 2C	Bøyningsendelser del 2	<i>PRscore2C</i>		combined PRscore27, PRscore28
Part 2E	Kompleksitet	<i>score2E</i>		0-42

Block 4:

Part	Name	Variable names	Variable numbers	Labels
		<i>in_out_bo</i>		0 no, 1 yes (=exclusion)
		<i>in_out_cl</i>		0 no, 1 yes (=exclusion)
		<i>in_out_pb</i>		0 no, 1 yes (=exclusion)
		<i>in_out_hs</i>		0 no, 1 yes (=exclusion)
		<i>in_out_pw</i>		0 no, 1 yes (=exclusion)
		<i>in_out_missing</i>		0 no, 1 yes (=exclusion)
		<i>in_out_CDI</i>		0 no, 1 yes (=exclusion)
		<i>in_out_age</i>		0 no, 1 yes (=exclusion)
		<i>in_out_add</i>		0 no, 1 yes (=exclusion)
		<i>in_out_comments</i>		0 no, 1 yes (=exclusion)
		<i>in_out</i>		0 no, 1 yes (=exclusion)

1.5 Description of socio-demographic datasets

The child datasets *WG_Child_data_140309* and *WS_Child_data_140309* include the child *id*, all *in_out* variables (block 4 above), a variable *part* (1 for cdi 1, 2 for cdi 2) and all variables as below:

Skjema for utfyllende informasjon : version 1

A. Opplysninger om barnet

- | | |
|--|---|
| i_1_1 | 1. Barnets nåværende bosted i postnummer: |
| i_1_2 | 2. Hvor har barnet bodd inntil i dag? |
| i_1_3 | 3. Hvis det er satt kryss i "Har tidligere bodd på et annet sted", før opp postnummer og sted |
| B. Opplysninger om barnets familie herunder de pågående barn | |
| i_2_1 | 1. Hvor mange barn er det i alt i familien (inkl. halvsøsken og helgebarn)? |
| i_2_2 | 2. Hvilket nummer i søskenflokket er barnet (inkl. halvsøsken og helgebarn)? |
| i_2_3 | 3. Hvis barnet har søsken, oppgi kjønn og alder (i hele år). Begynn med den yngste. (inkl. halvsøsken og helgebarn) |
| i_2_4 | 3.a søsken |
| i_2_5 | 3.a søsken alder |
| i_2_6 | 3.a søsken kjønn |
| i_2_7 | 3.b søsken |
| i_2_8 | 3.b søsken alder |
| i_2_9 | 3.b søsken kjønn |
| i_2_10 | 3.c søsken |
| i_2_11 | 3.c søsken alder |
| i_2_12 | 3.c søsken kjønn |
| i_2_13 | 3.d søsken |
| i_2_14 | 3.d søsken alder |
| i_2_15 | 3.d søsken kjønn |
| i_2_16 | 3.e søsken |
| i_2_17 | 3.e søsken alder |
| i_2_18 | 3.e søsken kjønn |
| i_2_19 | 3.f søsken |
| i_2_20 | 3.f søsken alder |
| i_2_21 | 3.f søsken kjønn |

C. Andre opplysninger om barnet

- | | |
|-------|---|
| i_3_1 | 1.a Er barnet regelmessig i kontakt med andre språk enn norsk? |
| i_3_2 | 1.b Hvis det er krysset av for "Noe kontakt" eller "hyppig kontakt" under 1a, oppgi hvilket/hvilke språk |
| i_3_3 | 2 Er barnet for tidlig født? |
| i_3_4 | 2b. Hvis det er krysset av for "Ja", oppgi i hvilken svangerskapsuke det er født |
| i_3_5 | 3. Har barnet vært innlagt på sykehus? |
| i_3_6 | 3b. Hvis det er krysset av for "Ja", oppgi hvor lenge (ca.) i antall uker |
| i_3_7 | 4. Har barnet hatt kronisk ørebetennelse? |
| i_3_8 | 5. Har du vært bekymret for ditt barns språkutvikling og tatt kontakt med fastlege eller helsestasjon om dette? |
| i_3_9 | 5a. Hvis det er krysset av for "Ja", beskriv kort hva du har vært bekymret for |

D. Opplysninger om mor

- | | |
|-------|--|
| i_4_1 | 1. Mors bosted |
| i_4_2 | 1a. Hvis det er krysset for "Annen adresse enn barnet" |
| i_4_3 | 2. I hvilket land er mor oppvokst |
| i_4_4 | 2a. Hvis det er krysset av for "Norge", oppgi i hvilken landsdel mor primært er oppvokst |
| i_4_5 | 2b. Hvis det er krysset av for "Et annet land" |
| i_4_6 | 3. Mors utdanningsnivå |

E. Opplysninger om far

- | | |
|-------|--|
| i_5_1 | 1. Fars bosted? |
| i_5_2 | 1a. Hvis det er krysset for "Annen adresse enn barnet" |
| i_5_3 | 2. I hvilket land er far oppvokst |
| i_5_4 | 2a. Hvis det er krysset av for "Norge", oppgi i hvilken landsdel far primært er oppvokst |
| i_5_5 | 2b. Hvis det er krysset av for "Et annet land" |
| i_5_6 | 3. Fars utdanningsnivå |

Note: There were no consistent checks on the child datasets WG_Child_data_RAW_140309 and WS_Child_data_RAW_140309.

2. Output for W&G: Tables and figures

Output consists of files in 4 separate folders: GPH (stata graphs), EPS (eps graphs), CSV (comma separated table data), HTML (tables in HTML code). A *readme* file contains mapping of titles, comments etc to names of files / numbering. The output is based on WG_CDI_FINAL_140309 alone.

All files from Appendix B (TKU) (possibly with small changes) are stored in the folders. All text tables from chapter 5, section 1 (TKU) and chapter 6, section 1 (TKU) are provided as well.

For the Appendix tables, a single word file (.pdf) will be compiled, containing all figures and tables in the same order as in appendix B, TKU.

Note: tables/figures contained in section 6.3 (TKU): “Effect of demographic variables” will not be provided. These analyses need further data handling first (with respect to WG_Child_data_RAW_140309).

3. Output for W&S: Tables and figures

Output consists of files in 4 separate folders: GPH (stata graphs), EPS (eps graphs), CSV (comma separated table data), HTML (tables in HTML code). A *readme* file contains mapping of titles, comments etc to filenames/numbers. The output is based on WS_CDI_FINAL_140309 alone.

Almost all files from Appendix C (TKU) (possibly with small changes) are stored in the folders (tables/figures on pages 276-287 are excluded). All text tables from chapter 5, section 2 (except subsection 6; examples) (TKU) and chapter 6, section 2 (TKU) are provided as well.

Note: tables/figures referring to the analysis of “example sentences” of the children (MLU etc.) will not be provided. These analyses need further data handling first.

For the Appendix tables, a single word file (.pdf) will be compiled, containing all figures and tables in the same order as in appendix C, TKU (except for examples)

Note: tables/figures contained in section 6.3 (TKU): “Effect of demographic variables” will not be provided. These analyses need further data handling first (with respect to WS_Child_data_RAW_1740309).

4. Open issues and recommendations

- Analysis of the example sentences in W&S
- Data cleaning/handling of the child datasets (fx to prepare for demographic analyses)
- W&S Parents' comments are worth inspecting: fx many parents state their difficulty (and strategy) to deal with (the child's) dialect; twin parents were insecure which child they should report on, some parents had comments on the CDI schemes' instructions etc.