FP7-ENV-2008-226873
http://www.mbkds.com/

D3.3 Interim Report on Data Validation Issues

WP3 – Quality Assurance

V1.2
Final

Lead beneficiary: UU
Date: 06/03/2014
Nature: Report
Dissemination level: RE
(Restricted to members of the Consortium and Commission Services)

TABLE OF CONTENTS

Document Information

Grant Agreement Number	ENV-226873	**Acronym**	MOBI-KIDS
Full title	Risk of brain cancer from exposure to radiofrequency fields in childhood and adolescence		
Project URL	http://www.mbkds.com/		
EU Project officer			

Deliverable	**Number**	3	**Title**	Interim Report on Data Validation Issues
Work package	**Number**	3	**Title**	Quality Assurance

Delivery date	**Contractual**	Month 53	**Actual**	06/03/2014
Status	Draft ☐		final ☑	
Nature	Demonstrator ☐ Report ☑ Prototype ☐ Other ☐			
Dissemination Level	Confidential ☐ Restricted ☑ Public ☐			

Authors (Partner)	(UU), (UU), (UU), (CREAL), (CREAL), (CREAL)		
Responsible Author		**Email**	
	Partner UU	**Phone**	

Document History

Name	Date	Version	Description
MOBI-KIDS_D3.3_version1.0_21.1.2014	21-1-2014	1.0	First draft version by
MOBI-KIDS_D3.3_version1.1_25.2.2014	25-2-2014	1.1	Second draft version seen by all authors
MOBI-KIDS_D3.3_version1.2_27.2.2014	27-2-2014	1.2	Final version, approved by all authors

Definitions

- Partners of the MOBI-KIDS Consortium are referred to herein according to the following codes:
  CREAL - Fundació Centre de Recerca en Epidemiologia Ambiental (Spain) – Coordinator
  UU - Universiteit Utrecht (Netherlands) – Beneficiary
  FT- France Telecom SA (France) – Beneficiary
  HPA - Health Protection Agency (United Kingdom) – Beneficiary
  LMU - Ludwig-Maximilians-Universitaet Muenchen (Germany)- Beneficary
  MUV - Medizinische Universitaet Wien (Austria)
  UNITO - Università degli Studi di Torino (Italy)
  ARECEA - Association pour la Recherche Epidémiologique dans les Cancers de l'Enfant et de l'Adolescent (France)
  UOA-SARG - National and Kapodistrian University of Athens (Greece)
  GERTNER INSTITUTE - Gertner Institute for Epidemiology & Health Policy Research (Israel)
  UOTTAWA - University of Ottawa (Canada)
  MONASH - Monash University (Australia)
  AUCKLANDUNI - The University of Auckland (New Zealand)

- **Grant Agreement**: The agreement signed between the beneficiaries and the European Commission for the undertaking of the MOBI-KIDS project (ENV-226873).
- **Project**: The sum of all activities carried out in the framework of the Grant Agreement.
- **Work plan**: Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.
- **Consortium**: The MOBI-KIDS Consortium, conformed by the above-mentioned legal entities.
- **Consortium Agreement**: Agreement concluded amongst MOBI-KIDS participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.
- **Deliverable review**: An evaluation procedure by one or more reviewers, which precedes the distribution of a deliverable (as defined in the Work plan) to the European Commission.
- **Quality assurance**: All the planned and systematic activities implemented to provide adequate confidence that an entity will fulfil requirements for quality.
- **Quality policy**: A set of principles on which quality assurance procedures are based.
- **Risk**: Uncertainty that may have a significant impact on the execution or outcome of the project, and which effect may be negative – a *threat* risk - or positive – an *opportunity* risk.
- **Foreground:** Means the results, including information, whether or not they can be protected, which are generated by activities in the Project. Such results include rights

related to copyright; design rights; patent rights; plant variety rights or similar forms of protection.

- **Background:** Means information which is held by participants prior to their accession to the Grant Agreement, as well as copyrights or other intellectual property rights pertaining to such information, the application for which has been filed before their accession to the Grant Agreement, and which is needed for carrying out the indirect action or for using the results of the indirect action.

Interim Report on Data Validation Issues

The protocol for the validation of questionnaire responses has been presented before in deliverable D.3.2. Part of the protocol was the table 'Validity checks of questionnaire data' (see Table I, Annex I); this table provides a list of checks to assess the quality of the questionnaires entered as part of the MOBI-KIDS study. Validity checks 1-6 (number of questionnaires entered) and 7-9 (distribution of responses for all variables) have already been covered in deliverables D5.5 and D.3.2, respectively; validity checks 10-18 (in-depth distribution of specific variables) will be covered in this report (D.3.3). Additionally, a validity check of the data from the occupational sections will be described.

The data presented in this report are based on all interviews performed until the end of 2012 (n=937). The distribution of responses for all variables in the main and parental questionnaire were extracted from the MOBI-KIDS database by at CREAL; all distributions were checked by UU.

1. Exact replies versus ranges

1.1. Percentage of exact replies

The questionnaires contain numerous questions in which participants are asked to provide a quantitative estimate, for example, amount of calls, duration of phone calls in minutes/hours per day/week/month, the age/year that one started to use a mobile phone, etc. These questions are open questions (no answers categories provided), that could be answered with an 'exact' reply or a range. A total of 113,394 of this type of question were checked in the interviews available for this deliverable; 92.9% were answered with an exact reply. The percentage of exact replies was highest in Japan (99.1%), and lowest in Korea (89.9%).

1.2. Wideness of ranges

The questions for which participants mostly provided a range rather than an exact reply were the following questions:

- B.5.1. "Number of calls - *at start*": 35.8% provided a range.
  Wideness of ranges: median: 1 call/day, Interquartile range [IQR]: 0.86 calls/day, max: 19 calls/day.
- B.5.2. "Length of time making and/or receiving calls - *at start*": 32.6% provided a range.
  Wideness of ranges: median: 1 min./day, IQR: 3.71 min./day, max: 380 min./day.
- B.6.1. "Number of calls - *recent use*": 37.7% provided a range.
  Wideness of ranges: median: 1 call/day, IQR: 1.29 calls/day, max: 20 calls/day.
- B.6.2. "Length of time making and/or receiving calls - *recent use*": 35.9% provided a range. Wideness of ranges: median: 2 min./day, IQR: 8.60 min./day, max: 1170 min./day.

During the final process of data cleaning, the widest ranges should be explored in depth to rule out erroneous ranges.

2. Residential history

2.1. Completeness

In question A.1. the participants are asked to provide a full residential history, i.e., where they lived from birth until their current address. We checked the completeness of the self-reported residential histories in the period from birth until the diagnostic date (as this is the period we are interested in). For each participant, we calculated the expected lifetime residential history (i.e., the time [in days] from birth until the date of diagnosis), and the actual reported time (in days) of residential history. In total, 6.4% of the expected lifetime residential history was missing (Table 1). This differed per country, with the highest percentage of missing lifetime observed in Germany (23.8%), and the lowest percentage in France (1.1%). Feedback will be provided to the individual countries, to improve future completeness of residential history.

Table 1: Residential history: % missing lifetime, by country

Country	% missing lifetime
Australia	2.9
France	1.1
Germany	23.8
Greece	9.9
Israel	6.5
Italy	6.1
Japan	2.5
Korea	4.5
Spain	3.0
The Netherlands	2.6
Total	**6.4**

2.2. Overlap

Furthermore, we checked how often participants reported more than one residence at a time (≥1 year overlap). This was seen 71 times among 69 (7.4%) participants (2 participants had multiple overlaps). Among these participants, the period of overlap was on average 2759.7 days (7.6 years), with a maximum of 8400 days (23.0 years). Most overlaps were reported in Greece, while no overlaps were reported in France, Korea, and The Netherlands. The observed overlaps could, for example, be explained by children splitting time between divorced parents.

2.3. Geocoding

Next, we wanted to know how complete the residential histories were with regard to the specific addresses provided. To geocode the addresses to be used for exposure modelling, it is crucial to have the street name, number, city, and postal code. In Spain, a first check was done with the addresses collected so far. A number of addresses were incomplete, e.g. missing the number or a wrong street name. The percentage of addresses that could be geo-localized differed per region: 82% in Cataluña, 69% in Andalucía, 65% in Valencia, and 52% in

Madrid. In the Netherlands, this process of geo-localization has recently been done for another project with a much wider period of residential history (from 1918 until now): 76% of the addresses could be localized to a specific building. Each country participating in the MOBI-KIDS study will be asked in the coming months to check the completeness of the specific addresses provided in their country, and to start the process of geocoding.

3. Country of origin

Question A.5. "Origin" asks: "In which country was your *relative* born?" *Relatives* that are asked for are: father, paternal grandfather, paternal grandmother, mother, maternal grandfather , and maternal grandmother. This question is country-specific, which means that each country could decide whether or not to ask this question in their country. Germany decided to not ask this question. In total, missing responses for country of origin were found for 2.4% to 2.7% of the interviews depending on the relative: the lowest percentage was found for the country of origin of the mother, maternal grandfather and maternal grandmother (2.4%). The percentage of missing responses also differed by country (Table 2): Australia, Italy, Korea, and The Netherlands had 0.0% missing responses, while Greece had the highest percentage of missing responses (8.9%).

Table 2: Country of origin: % missing responses, by country

Country	% missing responses
Australia	0.0
France	1.1
Greece	8.9
Israel	0.0-0.7*
Italy	0.0
Japan	6.7
Korea	0.0
Spain	2.6-3.0*
The Netherlands	0.0
Total	**2.4-2.7***

*% differed by relative

4. Mobile phone calendar

In section B.I. of the main questionnaire, the participant is asked in detail how many mobile phones he/she has used during lifetime, and during which periods these phones were used. The built-in validation checks of the CAPI [computer-assisted personal interview] makes it impossible to enter dates before the birth date or after the interview date. We checked how often participants reported more than one mobile phone at a time (≥1 year overlap). This was seen 63 times among 50 (.5.3%) participants (several participants had multiple overlaps). Among these participants, the period of overlap was on average 1099.9 days (3.0 years), with a maximum of 2375 days (6.5 years). Most overlaps were reported in Japan, while no overlaps were reported in Australia, Italy, Korea, and The Netherlands. Overlap can for example be explained by participants having multiple phones, e.g., a personal and a business mobile phone, which will – most likely – only be the case among young adults (52.4% was 18

years or older at date of diagnosis). Children using more than one phone could be using each of their parents' phones (we ask about phone use, whether or not it is their own phone), or their sibling's phone.

5. Mobile phone identification

5.1. Unknown mobile phone models

The participant is further asked in section B.I. of the main questionnaire to report the makes (brand names) and models of the mobile phones he/she had used or is currently using. The brand name of the phone (e.g., Samsung) was unknown for 64 out of 2759 mobile phones (2.3%) reported by the participants. The model/type of the phone (e.g., Galaxy SII) was unknown for 614 (22.3%) mobile phones (Table 3). The percentage of unknown mobile phone models was highest in Greece (55.3%), while lowest in Korea (0%). The CAPI contains a mobile phone database to search for mobile phone model in order to minimise the number of unknown phone models. Before this mobile phone database was used by the centres during the interviews (before June 2011), 38.8% of reported phones was of unknown model. After countries started to use this database to assist participants, the percentage of unknown phone models dropped to 20.7%. This percentage is expecting to drop even more as more centres use the mobile phone database.

Table 3: Identified and unknown mobile phone models

	Before database use		After start database use	
Identified	142	61,2%	2003	74,3%
Unknown	90	38,8%	524	20,7%
Total	**232**	100%	**2527**	100%

5.2. Prompts

In order to search in the mobile phone database, participants are asked for a number of prompts to narrow the search. For 13.4% of the unknown mobile phone models no prompts were entered (some centres asked for the prompts but did not record them on paper, so the information was lost when entering the paper questionnaire into CAPI; feedback has been given to those centres). For 5.1% of the unknown phone models all prompts were known, while for 6.4% all prompts were reported as "don't know". Annex II presents the "don't know" responses for the specific prompts. The prompt "What generation was your phone?" had the highest percentage (66.3%) of "don't know" responses.

6. Changes in mobile phone and Wi-Fi use

Participants are also asked in detail how often they use(d) their mobile phone for making and/or receiving phone calls, and how often they use(d) Wi-Fi at home, at school/work, and at other places. To get a complete picture of a participant's lifetime mobile phone and Wi-Fi use, the use is estimated in three phases: at start, recent use, and any changes in between. As described in Table I (Annex I), we planned to perform a check on the quality of these data. However, we were not able to derive any parameters from these data that would provide an

useful indication of their quality. Nevertheless, there is a validation check in the CAPI – if the use at start and recent use is different, the interviewer must proceed to the 'change of use' table.

7. Occupational sections

Coding of the occupational history of the participants (approximately 35% reported a job) and their parents (around pregnancy and longest held job) is essential to link potential workplace-related exposures to the cases, controls and their parents via job-exposure matrices (JEMs). We performed several checks on the quality of the collected occupational histories. After a first evaluation of the collected occupational histories up to then, the protocol for recording the occupational history was improved in early 2013.

In the fall of 2013, a first check on the codability was performed. First, the job histories were - if necessary - translated with Google Translate from the local language into English; they were then sent back to the centers with the question to check the accurateness of the translation and if necessary to update the translation. At the time of the coding exercise, 1,901 jobs from 13 countries were available. 668 jobs (representing roughly 20% of the unexposed jobs and 10% of the exposed jobs) were coded by (, UU) into ISCO 1968, the most detailed coding system available. (, Monash University) repeated the coding independently from for 651 jobs. Consequently, comparisons were made for the ISCO 1968 codes (at 5-, 4- and 3-digit level) and after linkage to a JEM for exposure to diesel motor exhaust, asbestos and crystalline silica. agreed at the 5-digit level for 55% of the jobs, for coding at 3-digit and at 2-digit level these percentages went up to respectively 69% and 78%. The percentage of codes resulting in the same exposure level was 97%, 95% and 97% respectively for diesel motor exhaust, asbestos and crystalline silica. The percentage of congruence for the exposed only was 81%, 49% and 59% respectively for the same three agents.

The overall conclusion was that the quality of the occupational histories was often poor, which made coding the jobs difficult. After the necessary adjustments to the protocol in early 2013, a clear improvement in data quality was apparent. However, special attention was still required for Japan, Korea and Taiwan.

Although the coding of the occupational histories appeared to be difficult and time-consuming, the quality of the coding seemed to be good. Given the relatively small number of job histories needed to be coded, it was decided to have all the coding done centrally by and . Additional consensus on non-congruent codes (leading to exposure) will be a necessity. At the local level, the data collected should be as specific and detailed as possible. The accuracy checks of the translation of the occupational histories into English will be carried out locally and should be as detailed as possible especially for blue collar jobs with a high potential of occupational exposures.

Annexes

Annex I – Protocol: Table 'Validity checks of questionnaire data'

Version: 08-02-2013

Table I: Validity checks of questionnaire data

	Validity check	Level
1	Number of complete, incomplete and pending main questionnaires	Overall, by centre
2	Number of complete, incomplete and pending parental questionnaires	Overall, by centre
3	Number of complete, incomplete and pending clinical questionnaires	Overall, by centre
4	Number of complete, incomplete, pending and refusals of non-respondent questionnaires	Overall, by centre
5	Number of biological samples collected	Overall, by centre
6	Number of subjects agreeing to participate in the operator validation study	Overall, by centre
7	Distribution of responses for all variables (frequencies, mean/median, minimum, maximum)	Overall, by question, by centre
8	Distribution of all "don't know" replies	Overall, by question, by centre
9	Distribution of all incorrectly missing replies	Overall, by question, by centre
10	Distribution of number of exact replies versus ranges	Overall, by question, by centre
11	Wideness of ranges reported	Overall, by question, by centre
12	Residence (QA.1): completeness of residential history (from birth until today)	Overall, by centre
13	Residence (QA.1): distribution of reporting more than one residence at a time (≥1 year overlap)	Overall, by centre
14	Country of origin (QA.5): distribution of missing replies for countries who are participating	Overall, by centre

15	Mobile phone calendar (QB.3.1/4): distribution of reporting more than one mobile phone at a time (≥1 year overlap)	Overall, by centre
16	Mobile phone prompts (QB.3.6-11): for unknown phone models/make, distribution of missing and "don't know" replies	Overall, by centre
17	Changes in phone use (section B.V/C.III): completeness of changes specified, i.e. how logical is the table of changes (e.g. if there is a huge gap between # of calls in the middle vs current date, is there a logical explanation?)	Overall, by centre
18	Changes in Wi-Fi use (QC.22.e/24.e/26.e): completeness of changes specified, i.e. how logical is the table of changes (e.g. if there is a huge gap between # of calls in the middle vs current date, is there a logical explanation?)	Overall, by centre

Annex II – Mobile phone prompts: "don't know" responses

B.I. Mobile phone prompts	**% "don't know" responses**
B.3.6 Was it a classic phone or other special shape?	11.2
B.3.7 Could you flip the phone open or slide it open?	11.2
B.3.8 Was your phone a smartphone?	16.9
B.3.9 Was there a visible antenna on your phone?	11.2
B.3.10 What generation was your phone?	66.3
B.3.11.1 Color	18.9
B.3.11.2 Internet	22.8
B.3.11.3 Wi-Fi	23.4
B.3.11.4 Radio	25.6
B.3.11.5 Touch screen	14.8
B.3.11.6 Bluetooth	23.2
B.3.11.7 PC connector	30.9
B.3.11.8 MP3	26.4
B.3.11.9 Camera	17.7
B.3.11.10 SMS	12.6
B.3.11.11 Changeable cover	33.1
B.3.11.12 Speaker	24.2
B.3.11.13 Full keyboard	19.7