The Longitudinal Study of Australian Children's Child Health CheckPoint

Data Issues Paper

December 2018

murdoch children's research institute

Growing up in Australia

The Longitudinal Study of Australian Children

Abbreviations

BIA	Bioelectrical Impedance Analysis
CELF-4	Clinical Evaluation of Language Fundamentals Fourth Edition, Australian version
FrACT	Freiburg Visual Acuity and Contrast Test
HRQL	Health Related Quality of Life
ISCW	International Survey of Children's Wellbeing
LSAC	Longitudinal Study of Australian Children
NPVT	National Institute of Health Toolbox Picture Vocabulary Test
PDS	Pubertal Development Scale
PedsQL	Pediatric Quality of Life Inventory
pQCT	Peripheral quantitative computed tomography
REDCap	Research Electronic Capture
SOPs	Standard Operating Procedures

Suggested citation

Readers wishing to refer to this document should cite the following:

> Sarah Davies, Susan Clifford, Alanna Gillespie, Katherine Lange, Josh Muller and Melissa Wake. (2018). Longitudinal Study of Australian Children's Child Health CheckPoint Data Issues Paper – December 2018. Melbourne: Murdoch Children's Research Institute. doi: 10.25374/MCRI.5821230.

Contact information

For more information, contact:

Child Health CheckPoint team
Murdoch Children's Research Institute
The Royal Children's Hospital
50 Flemington Road
PARKVILLE VIC 3052
Website: checkpoint-lsac.mcri.edu.au
Email: lsac.childhealthcheckpoint@mcri.edu.au
Phone: +61 3 9936 6464

Table of Contents

Abbreviations **2**

Suggested citation **3**

Contact information **3**

1 Introduction **5**

2 Data collection **6**

2.1 Review of data collection comments 6

2.2 Pubertal development 8

2.2.1 Slight wording change 8

2.2.2 Missing data due to branching of sex-specific questions 8

2.3 Bone and muscle morphology, bone density – missing data due to delays in licensing 9

2.4 Blood pressure – measured using an alternative device 9

2.5 Expressive and receptive language – imputation of missing data 9

2.6 Physical activity, sedentary behaviour and sleep 9

2.6.1 Missing data due to equipment malfunction or unavailability 9

2.6.2 Variation in number of weekdays and weekend days recorded 10

2.7 Visual acuity – retrospective correction of non-calibrated data 10

2.8 Chronic physical pain – inclusion of additional questions and branching logic 11

2.9 Middle ear function – missing data and data transfer method 12

2.10 Biospecimens – time between sample collection and processing 12

3 Data management **14**

3.1 Height, weight and body composition - inconsistencies between datasets 14

3.2 Arterial stiffness– pulse wave analysis proprietary equations not validated in children 14

3.3 Physical activity, sedentary behaviour and sleep – removal of impossible values 15

3.4 Hearing threshold level – review of cases of no response to auditory tones 15

3.5 Receptive vocabulary – removal of potentially unreliable data 16

3.6 Wellbeing and quality of life – likely inconsistency between responses 16

3.7 Snack observation 17

3.7.1 Correction or removal of impossible values 17

3.7.2 Changes in snacks due to being discontinued or altered by manufacturer 18

3.8 Hospital admissions – data from two sources combined to overcome incorrect branching logic 18

3.9 Community participation – activities data reshaped and rearranged to categorise activities 19

3.10 Biospecimens – sample tracking 20

References 21

1 Introduction

The Child Health CheckPoint Data Issues paper provides a summary of data issues relating to the Longitudinal Study of Australian Children (LSAC)'s Child Health CheckPoint. It is designed to assist CheckPoint data users as they undertake research and analysis using the dataset. This paper summarises data issues that may impact on the use of particular variables, emphasises changes to the data, and highlights important information for interpretation of results.

This paper is to be read in conjunction with:

- The Child Health CheckPoint Data User Guide (available at checkpoint-lsac.mcri.edu.au), and
- Data collection and management Standard Operating Procedures (SOPs, in preparation, anticipated to be available at checkpoint-lsac.mcri.edu.au from late 2018).

This paper is structured as follows:

- The first section, 'Data collection', describes data issues relating to variations in data collection protocols.
- The second section, 'Data management', describes data issues that occurred during the data preparation processes i.e. during raw data extraction, data combining, data cleaning or when creating derived variables.

2 Data collection

2.1 Review of data collection comments

The Research Electronic Data Capture (REDCap) application (Harris et al., 2009) was used to capture much of the Child Health CheckPoint data. Data entry forms included a free-text comment field for CheckPoint team members to record notes. These included notes that were relevant to other team members during the visit or that could have impacted on data quality.

Post-data collection, the CheckPoint data team systematically reviewed free-text comments for most measures and biological samples, and the relevant investigator advised on the most suitable actions to take.

For measures, actions fell into four categories:

1. Where the comment was deemed to have no impact on the reported data, it was ignored (e.g. “Study child’s sibling in attendance” or “Participant had good technique” in lung function assessment).
2. Where the assessment data were recorded in the free-text comment field instead of the appropriate data entry field, the data were transferred to the appropriate field (e.g. “NPVT result did not save on iPad but recorded in comments”).
3. Where the comment indicated a data collection protocol change or participant condition which data users might need to be aware of, a flag variable was created to indicate which data points were affected (e.g. “Visual aids worn during test” or “Injury”).
4. Where the data were deemed to be incorrect, they were removed from the dataset (e.g. “No resistance on the bike. Results will not be valid"). As per missing data standards (see CheckPoint Data User Guide), incorrect values were recoded as ‘.e’.

For biological samples, actions fell into three categories:

1. Where the comment was deemed to have no impact on the reported data or collected sample, it was ignored (e.g. “Only one tube collected”).
2. Where the assessment data were recorded in the free-text comment field instead of the appropriate data entry field, the data were transferred to the appropriate field (e.g. barcode entered into the free-text comment field).
3. Where the comment indicated a data or sample collection protocol change or participant condition which data users might need to be aware of, this information was used in validation checks (e.g. “serum blood tube filled before EDTA blood tube”). Validation checks involved verifying data and biospecimens were correctly linked to participants. Where data was incorrectly linked but the correct link could be identified, the data was updated for the incorrect and correct participants to reflect this. Where data was incorrectly linked and the correct participant could not be identified, or was uncertain, data were removed (coded unexpected missing, ‘.e’).

Free-text comments regarding 2D and 3D photos, the handwritten story, and the saliva, buccal swab, hair and toenail samples have not yet been reviewed or actioned.

Table 1 summarises the actions taken for measures and biological samples. More detailed information is provided in the relevant Data Management SOPs (available at checkpoint-lsac.mcri.edu.au).

Table 1. Summary of actions taken from data collection comments (n=number of comments), study children and attending parents combined

Measure	No change to data required	Data moved to appropriate field	Protocol change/ condition flagged	Data removed; replaced with missing code[1]
Height, weight, body composition and waist circumference	215	51	10	18
Bone and muscle morphology, bone density	117	0	0	13
Carotid intima-media thickness and other attributes	68	5	32	0
Arterial stiffness and blood pressure	594	2	32	20
Microvasculature structure	155	0	117	0
Lung function	2239	49	13	15
Expressive and receptive language	133	3	115	37
Receptive vocabulary	64	59	0	0
Hearing threshold level	133	0	190	0
Middle ear function	158	222	1	1
Speech reception threshold	141	3	0	0
Food choices	541	180	47	14
Physical activity, sedentary behaviour and sleep	3	4	0	0
Time use	123	0	0	0
Large muscle power	53	0	9	3
Cardiorespiratory fitness	184	0	88	31
Visual acuity	853	9	4	4
Biological sample	**No change to data required**	**Data moved to appropriate field**	**Protocol change / issue used in validation checks**	
Venous blood sample	293	53	49	
Finger prick blood sample	50	17	12	
Urine sample	62	26	49	

[1]The entire set of variables were recoded to missing for affected participants, except for bone and muscle morphology, bone density, arterial stiffness and blood pressure where only a subset of variables were recoded to missing.

2.2 Pubertal development

2.2.1 Slight wording change

The Child Questionnaire contained the Pubertal Development Scale (PDS) (Carskadon & Acebo, 1993; Petersen, Crockett, Richards, & Boxer, 1988). The PDS includes five questions with four-point Likert scale responses ranging from 'Has not started yet' to 'Seems complete'. The first three questions are asked of both males and females, followed by two male-specific or female-specific questions in male and female versions, respectively.

For the first five weeks of the study, question 2 of the PDS was asked twice, using slightly different wording:

> 2.1 "Would you say that your body hair growth (hair any place other than your head, such as under your arms)…" and
> 2.2 "And how about the growth of your body hair? Would you say that your body hair growth…"

During March 2015 these two questions were removed and replaced with one question which read:

> 2.3 "And how about the growth of your body hair? (Body hair means hair any place other than your head, such as under your arms.) Would you say that your body hair growth…"

The variable 'PDS Q2 Growth of body hair' (fch03c01b) in the dataset contains:

- Responses to question 2.1 for the 155 children who participated in the first five weeks of the study.
- Responses to question 2.3 for the remaining children who participated after the first five weeks of the study.

Responses to question 2.2 were discarded, on advice from study investigators knowledgeable in pubertal development.

The question wording was deemed similar enough to combine the responses into a single variable. This decision was made in order to ensure as much data as possible was available to users.

2.2.2 Missing data due to branching of sex-specific questions

Study child sex (male or female) was not asked of the participants, but instead imported into the database prior to participants attending their assessment. In 14 cases, sex data were not imported prior to the assessment, so these children were not presented sex-specific questions on puberty in the electronic questionnaire. These data are missing and are marked as '.e'. Apart from these 14 participants, there were a larger number (up to 130) of children who answered some or none of the puberty questions, resulting in a high amount of unexpected missing data (labelled '.e').

2.3 Bone and muscle morphology, bone density – missing data due to delays in licensing

The peripheral quantitative computed tomography (pQCT) scanner emitted a low dose of radiation during operation. The CheckPoint team required licences to operate the scanner and to acquire and dispose of a source of ionising radiation. Separate licences were required in each state or territory around Australia. The pQCT scanner was not used on a number of days because the licencing was not complete. The missing data code '.e' was used to indicate missing data due to equipment not being available.

2.4 Blood pressure – measured using an alternative device

Blood pressure was measured using an automated sphygmomanometer (blood pressure machine) for 12 participants because the SphygmoCor XCEL device was not available. Because it has not been possible to calibrate the two devices against each other, flag variables (fcbpmach, fabpmach) are included to identify participants with blood pressure measured using the sphygmomanometer.

2.5 Expressive and receptive language – imputation of missing data

Pearson's Clinical Evaluation of Language Fundamentals Fourth Edition (CELF-4), Australian version (Semel, Wiig, & Secord, 2006) is a validated clinical tool used for the 'identification, diagnosis, and follow-up evaluation of language and communication disorders in students 5-21 years old.' The Recalling Sentences subtest of the CELF-4 was administrated as intended by the CELF-4 publishers to assess participants' expressive and receptive language skills in a research setting. The test ended after sentence 32 or after three consecutive incorrect responses. For 110 participants, data for some sentences were missing and the recordings of these recalled sentences were either missing or faulty. If less than eight of the 32 sentence recordings were missing, scores for these sentences were imputed based on scores for the previous and subsequent sentences. If eight or more sentence recordings were missing, the participant's data were removed from the dataset. More details are provided in the Language (Expressive and receptive language) Data Management SOP.

2.6 Physical activity, sedentary behaviour and sleep

2.6.1 Missing data due to equipment malfunction or unavailability

Some accelerometers malfunctioned due to extended periods without battery recharge, suspected water damage or unknown reasons. In addition, some accelerometers were lost in the field, lost in transit or remained with participants for extended periods. Therefore, on occasion, there were not enough accelerometers available for participants at their visit. The CheckPoint team asked participants if they could send an accelerometer, with a reminder of how to wear and return the device, at a later date if accelerometers weren't available at the time of their visit. This meant that, on some occasions, the study child and their attending parent wore their accelerometer weeks or months apart from each other and from their CheckPoint visit. A flag variable (fcdaysaft and fadaysaft) alerts data users to the number of days between child and attending parent accelerometry recordings.

At the end of the study, 926 participants (492 children and 434 parents) had no accelerometery data:

- 62 refused this measure.
- 161 were missing all data due to lost accelerometers.
- 703 were missing all data due to accelerometers malfunctioning by failing to record (n=673) or recording only invalid data (n=30).

In addition, 183 participants had incomplete data due to accelerometers malfunctioning (i.e. recording on only some of the days the device was worn). Available data for these participants are included in the dataset (see section 2.6.2 below).

2.6.2 Variation in number of weekdays and weekend days recorded

Participants were asked to wear an accelerometer for eight days following their visit. For study children participating during the school term, data were collected on both week and weekend days. A large proportion of the children who participated during school holidays had data for weekend or holiday days only. In addition, some participants have less than eight days' data due to extended periods of non-wear or accelerometer malfunction.

Because activity patterns can differ substantially by type of day, particularly for children, we have flagged those with at least 3 weekdays and 1 weekend day as meeting 'minimum weekend/weekday day distribution criteria' (fcvalidwkwe, favalidwkwe). However, participants who had some data but did not meet this criterion were still included in the dataset. In summary:

- 1280 study children and 1378 attending parents meet 'minimum weekend/weekday day distribution criteria'.
- 102 study children and 62 attending parents have data for at least one day but do not meet the above criterion (e.g. accelerometer malfunction or insufficient wear time).

In cases where distribution of day types is not considered relevant, data users can optionally use the alternative flag variables, 'meets minimum number of days and valid data criteria' (fcvalidmin, favalidmin). These flag variables identify participants with at least 4 days of any type, after removing participants with impossible values for sleep and combined activity time without sleep (referred to as 'wear time' in the dataset and data dictionary. See data cleaning section 3.3 below for criteria of impossible values). For this criterion:

- 1317 study children and 1402 attending parents meet 'minimum number of days and valid data criteria'.
- 65 study children and 38 attending parents have data for at least one day but do not meet the above criterion (e.g. accelerometer malfunction or insufficient wear time).

2.7 Visual acuity – retrospective correction of non-calibrated data

Visual acuity was measured using the Freiburg Visual Acuity and Contrast Test (FrACT) and its software program (Bach, 1996, 2006). This program required re-calibration each time it was set up in a new location. This involved measuring the length of a line displayed on the screen and entering this length into the program. Entering a wrong value would cause a linear shift in the derived results. Recalibration was accidently omitted during short periods of 2-19 operating days at some Assessment Centres. The incorrect settings used during these periods were recorded. Calculations for retrospective correction of non-calibrated data were provided by the FrACT program designer, Professor Michael Bach, and applied to the data. After correction the data were are accurate. Therefore, a flag variable

was not necessary to mark these participants. Data for four participants has been removed from the dataset and recoded '.e' because the incorrect settings were not recorded.

2.8 Chronic physical pain – inclusion of additional questions and branching logic

The Child and Parent Questionnaires contained self-reported pain assessments. Attending parents were presented with questions asking about pain they had experienced. Study children were asked questions and also presented with a diagram to mark the location of any pain they had felt.

Attending parents were asked:

1. Thinking back over the past month, have you had any pain or pains, which have lasted for a whole day or longer? Yes/No

Those who responded "Yes" were then asked:

2. When did the pain start? Less than 3 months ago/ More than 3 months ago

Study children were initially shown only a 'pain manikin', i.e. pictorial representation of the front and rear aspects of the body (Jones, Watson, Silman, Symmons, & Macfarlane, 2003) and asked to select regions (up to 33 on each side) where they 'often experience pain'. Due to a higher than expected proportion of study children reporting pain, three additional questions were added to the Child Questionnaire in July 2015. After this date, the Pain Manikin and additional pain questions were only presented to those who reported pain for at least one day in the last month. The three additional questions were the two above (asked of attending parents), plus a third:

3. In the past month, how often did the pain get in the way of the normal things you do (like going to school, seeing your friends, playing sport)? 0=Not got in the way at all→10=Unable to carry out any activities

The 'pain manikin' was administered to children either electronically or on paper.

At Main and Mini-Assessment centers, children selected body regions where they 'often experience pain' using the iPad touch screen ('electronic' administration). At home visits, children used a pencil to shade in body regions on a print-out of the pictogram ('pen-and-paper' administration) due to internet connectivity restraints. The three pain questions were administered via iPad, regardless of visit type.

A flag variable (fcpainqu) alerts data users to which pain questions were administered to each study child:

1. "Electronic Pain Manikin only" (696 study children) – administered at Main Assessment Centres visits prior to July 2015.

2. "Electronic pain questions and Pain Manikin" (808 study children) – administered at Main and Mini-Assessment Centre visits from July 2015.

3. "Electronic pain questions and paper Pain Manikin" (290 study children) – administered at Home Visits. Pain questions were administered electronically with appropriate branching, and participants who reported pain completed the paper version of the Pain Manikin.

4. "Electronic pain questions and paper Pain Manikin administered incorrectly" (80 study children) – administered incorrectly at Home Visits. Pain questions were administered electronically with appropriate branching, and either (i) the paper version of the Pain Manikin was not provided to children who reported pain, or (ii) the paper version of the Pain Manikin was incorrectly provided to children without pain. For more information, please refer to the Pain data management SOP.

2.9 Middle ear function – missing data and data transfer method

Tympanometry assessments were added to the *Listen Up* station assessments in May 2015, as the station become more streamlined and faster to complete. Therefore, 702 participants who were assessed prior to the introduction of this measure are missing these data (indicated by the missing data code '.d'). An additional 730 Home Visit participants do not have data, as this measure was only administered at Assessment Centres, and 92 participants are missing data for other reasons (e.g. blockage or infection of the ear canal).

Individual participants' tympanometry results were displayed on the device LED screen, and exported to USB. Occasionally the results could not be exported so were captured by the CheckPoint team member 1) taking a photograph of the tympanometry LED display screen or 2) directly entering the results into the free-text comments section of the REDCap data entry page. The photographs were later transcribed.

No data were lost due to alternative data capture methods as all of the variables included in the export were displayed on the LED screen. Tympanometry data were exported to USB for 1682 participants, transcribed from a photograph for 535 participants and directly entered into the data entry page for eight participants. Flag variables (fctymptr, fatymptr) alert data users to the data transfer method used.

2.10 Biospecimens – time between sample collection and processing

Most biological samples collected at Main and Mini-Assessment Centres were processed within two hours. For urine samples collected at Home Visits and Mini-Assessment Centres without laboratory facilities, there was an unavoidable delay in sample processing.

Some urine samples collected at Home Visits were refrigerated overnight before processing at the Assessment Centre the next day. Other urine samples were sent via express post (not refrigerated) to a MCRI laboratory for processing.

Buccal swabs collected by the non-attending biological parent were also returned via express post, using collection kits (Oracollect DNA OCR-100, The Hague, Netherlands) that sealed the swab within a preservation solution to keep the sample stable at room temperature for at least 60 days. If these kits were not available, simple collection kits (FloqSwab COPAN Flock Technologies, Brescia, Italy) without preservation liquid were used instead. FloqSwabs were frozen at -80 degrees on receipt, usually within days of sample collection.

The length of time between sample collection and the start of processing is provided in the dataset in hours for blood (fcserumhr, faserumhr, fcwholebhr, fawholebhr, fcbuffcehr, fabuffcehr, fcbufflhr, fabufflhr, fcplasmehr, faplasmehr, fcplasmlhr, faplasmlhr) and saliva (fcsalivahr, fasalivahr) samples, in days for buccal swabs (fcswabdy, faswabdy fbswabdy), and hours (if processed same day; fcurinehr, faurinehr) or days (if not processed same day; fcurinedy, faurinedy) for urine samples.

Data users should make their own decisions about excluding some samples from analyses due to delayed processing. Flag variables are also available to indicate where the collection or processing date or time was not recorded and had to be estimated (fcserumfl, fcbuffcefl, fcbufflfl, fcswabfl, fcplasmefl, fcplasmlfl, fcsalivafl, fcurinefl, fcwholebfl, faserumfl, fabuffcefl, fabufflfl, faswabfl, faplasmefl, faplasmlfl, fasalivafl, faurinefl, fawholebfl, fbswabfl).

3 Data management

3.1 Height, weight and body composition - inconsistencies between datasets

Weight and body composition were measured using bioelectrical impedance analysis (BIA) machines. Four-limb BIA scales were used in Assessment Centres. These scales displayed key variables on the LED screen, and a wider range of variables could be exported to USB. Portable two-limb BIA scales were used in most home visits. These scales displayed key variables on the LED screen; there was no option to export data to USB. The type of scale used to assess each participant is flagged in the dataset (fcbiasc, fabiasc).

Weight and total body fat were read off the four-limb BIA screen and entered into the '*Measure Up*' REDCap data entry page. This allowed CheckPoint team members at other stations to access height and weight data (a prerequisite for some other assessments) in real time. The four-limb BIA data were exported to USB regularly, usually at the end of each day.

Inconsistencies were discovered between some participants' data in REDCap and the BIA-exported data. Most inconsistencies were due to unreliable or incomplete participant ID in the BIA-exported dataset (e.g. the participant ID correctly identified the family, but not if the participant was the study child or attending parent). Participant ID was confirmed by matching other variables (e.g. height) in both datasets and other investigations (e.g. time stamps). Participant ID was investigated, and where appropriate, completed or corrected for approximately 600 participants.

Height was matched from BIA-exported data to the manually-recorded values in REDCap to validate this process; where height differed more than 1cm between the data sources these data are flagged in the dataset (fcrnderr, farnderr). BIA data were removed for these participants and given missing values '.e'. The anthropometry investigator advised that a height difference greater than 1cm would invalidate the calculated BIA measures (BIA data removed for 19 study children and 13 attending parents).

The height, weight and body fat (body fat kg, %, and fat mass index) variables in the dataset are from the BIA machine export, where available. If BIA-exported data were not available, data were exported from the REDCap data entry page. See the Anthropometry data management SOP for further details.

3.2 Arterial stiffness– pulse wave analysis proprietary equations not validated in children

The SphygmoCor XCEL device was used to measure arterial stiffness (e.g. pulse wave velocity and analysis) and blood pressure. Most of the child arterial stiffness and blood pressure variables (e.g. brachial blood pressure and pulse wave velocity) are established in the literature as robust measures.

Additionally, the SphygmoCor uses internal proprietary equations to generate estimated central aortic pressure and augmentation index values (i.e. components of pulse wave analysis). These equations have been validated in adults, but have not yet been validated in children. The CheckPoint dataset contains these data for completeness. The child pulse wave analysis variables are flagged within the 'notes' section of the Data Dictionary to ensure data users are aware to take 'caution in interpretation as not validated in children'. If validated child equations become available, these data will be updated in future releases of the CheckPoint dataset.

3.3 Physical activity, sedentary behaviour and sleep – removal of impossible values

During data cleaning, impossible values for sleep and combined activity time without sleep (referred to as 'wear time' in the dataset and data dictionary) were removed, as advised by the investigator overseeing accelerometry. Data were replaced as missing for approximately 20 participants who had less than 10 minutes of sleep or had more than 1350 minutes (22½ hours) of activity without sleep (wear time) recorded per day. These 'near-impossible' values were likely due to equipment or software malfunction. Possible yet implausible values, such as less than 200 minutes ($3^1/_3$ hours) of sleep per day, are flagged (fcflaccel, faflaccel) as potential data issues which may affect the calculated mean values.

3.4 Hearing threshold level – review of cases of no response to auditory tones

Detailed data were collected using the AudioConsole v3.3.4 software to determine hearing thresholds. This software presented data numerically in a table format and as a graphical presentation (audiogram). A review of participants with a 'no response' to any frequency audiogram (n=187) was undertaken. For these participants, audiograms were cross-checked with numeric results to check if the participant truly did not respond to any auditory tones, or if 'no response' was assigned incorrectly by the software. Two CheckPoint team members specialising in hearing reviewed the Audioconsole software data independently and classified the 'no response' classification as 'true' or 'false' according to the following criteria:

a. A **true** 'no response' classification was defined as when 'no response' appeared on the audiogram, no numeric data existed for the maximum possible stimulation level of 90 dB at 8 kHz, and responses over different stimulation levels corresponded with those expected for severe or profound (high-frequency) hearing loss.
b. A **false** 'no response' classification was defined as when responses were displayed on the audiogram and existed in the dataset at extreme stimulation (i.e. at very high or very low) levels, yet 'no response' was displayed on the audiogram for an intermediate stimulation level and numeric results existed for the same intermediate stimulation level.

Disagreements between the two assessors were resolved through discussion with the relevant study investigators.

The review revealed:

1. Most of the 'no response' classifications were false (n=174). For these participants, the numeric test results were included in the dataset.
2. True 'no response' classifications were confirmed for seven attending parents. The numeric results were replaced; the threshold was set at 5dB above the maximum stimulation level (95dB HL for 8kHz). This protocol has previously been followed with elderly participants (Homans et al., 2017) and with US children aged 6-19 years (Niskar et al., 1998).
3. Six participants' data were removed either because an error was found in the exported data (n = 1, multiple tests performed and it was not clear which tests belonged to the participant) or because no test data existed for this participant except for the 'no response' classification (n = 5).

3.5 Receptive vocabulary – removal of potentially unreliable data

The National Institute of Health Toolbox Picture Vocabulary Test (NPVT) was administered using an iPad and headphones. Participants heard a word and were asked to select the picture that best represented the meaning of the word. Data were exported and backed up each day, with cumulative files from all previous participants exported from each iPad. This included a 'detailed' raw data text file of all participant trials and a 'summary' CSV file of the single output score (theta score) for each participant for each iPad.

Quality checks on the raw data files revealed that some participants had data for multiple tests when only a single test had been completed. Data were removed if it was unclear which participant results belonged to (e.g. when multiple tests were exported under the same ID number; 95 participants), or if the test was administered incorrectly (e.g. a study child selecting the adult age group; 7 participants). The missing data code '.e' was used to reflect missing data due to 'unreliable value(s)'. Flag variables (fch10c01c, fch10a01c) alert data users to participants whose theta score was transferred from the REDCap free-text comment field.

Further details regarding data validation are provided in the Language (Receptive vocabulary) data management SOP.

3.6 Wellbeing and quality of life – likely inconsistency between responses

The following health-related quality of life (HRQL) measures were included in the Child Questionnaire:

- Pediatric Quality of Life (PedsQL) General Wellbeing Scale
- PedsQL Inventory
- Child Health Utility 9D, and
- International Survey of Children's Wellbeing (ISCW).

As part of the quality checks, all of the HRQL data were reviewed together with the assumption that an individual participant's responses should correlate with each other. Participants who provided very inconsistent responses were identified and their data reviewed, and the following actions were taken:

- One participant had selected the most extreme response (at either the lower or higher end) for almost every question on the whole questionnaire so that their responses systematically contradicted themselves; these responses were considered spurious and removed.
- Six participants provided responses that may be contradictory. For example, one study child provided opposite responses for subscales of the ISCW (i.e. 'not at all satisfied' for 'Your life as a whole' but 'totally satisfied' with other subscales of the instrument, as well as reporting difficulties with physical functioning on the PedsQL Inventory. These data were retained in the dataset with a flag variable (fcsrflag) alerting data users to these participants.

3.7 Snack observation

3.7.1 Correction or removal of impossible values

The time when participants entered and left the *Food Stop* station, and 'time since last eating or drinking' (recorded in the *Young Bloods* station before *Food Stop)* were sometimes entered unreliably by assessors due to AM/PM discrepancy, duration/ clock time confusion, and transcription errors. Following a review of extreme values, some data were corrected or removed due to impossible values.

Food Stop station start and finish times were cross-checked with *Check-In* time, *Young Bloods* finish time, *Young Bloods* and *Food Stop* comments, and data time stamps for other stations before and after *Food Stop*. Data for 87 study children and 70 attending parents were corrected by changing from AM to PM time. Negative or long (>35 minutes) durations at the *Food Stop* station were corrected for 20 participants by shifting times earlier or later by one to five hours, and for one participant by reversing the start and finish times, while data for 12 participants were removed as it was not clear what the correct times were.

Participants who spent 0 to 4 minutes at the *Food Stop* station and didn't consume any food were considered not to have participated in *Food Stop* and their times were recoded as missing (4 study children and 14 attending parents). Times were not removed from the dataset if any food items were consumed and no other time errors were apparent.

The time that the participant last consumed food or drink was cross-checked with *Check-In* time, *Young Bloods* finish time, *Young Bloods* and *Food Stop* comments, data time stamps for other stations before and after *Food Stop*, and average sleep and wake times from accelerometry data (where available). Data corrections included:

- Where the time was recorded in the free-text comment field instead of the appropriate data entry field, the data were transferred to the appropriate field (4 study children and 5 attending parents).
- Where a comment stated that the time entered was not the last time the participant consumed food or drink prior to their assessment but no correct time was recorded, the data were removed (4 study children and 3 attending parents).
- Where the time was entered as an AM time but was more likely to be a PM time (i.e. the CheckPoint team had not used the 24 hours clock), data were updated to the PM time (21 study children and 36 attending parents).
- Where the participant last ate between 02:00 and 04:00 AM, the average sleep times, wake times and the start of assessment times were cross-checked to determine the plausibility that the time stated was the *time* that the participant last ate, or whether it was more likely to be the *duration (hours)* since the participant last ate. Data were corrected where the cross-check revealed inconsistencies with other data collected that suggested the CheckPoint team had incorrectly interpreted the question as duration since last ate (4 study children and 3 attending parents). Data were removed from the dataset where both the time since last eaten and the duration since last eaten values were impossible since both would occur during the assessment time (10 study children and 3 attending parents).

Corrected values for 'time of last food or drink' and calculated durations for fasting before and eating at *Food Stop* station are included in the dataset. Additionally, a flag variable (fcatebfs, faatebfs) alerts data users to participants known to have consumed food or drink during their CheckPoint assessment visit or completing stations in a modified order. Further details regarding data cleaning are provided in the Diet and Food Choices data management SOP.

3.7.2 Changes in snacks due to being discontinued or altered by manufacturer

Two food items provided in *Food Stop* changed during the data collection period due to the manufacturer discontinuing or altering the item.

Part-way through data collection, the peaches in syrup (150g pot) were discontinued by the manufacturer. This item was replaced with fruit salad in syrup of the same weight and brand. Nutritional differences between the two items can be found in the Food Stop data management SOP. Because the majority of participants received the peaches (not the replacement fruit salad), all the related variables (e.g. item grams, kilojoules, nutrients; total grams, kilojoules, nutrients) used weight and nutritional information from the peaches.

Shortly after data collection commenced, the miniature 11g chocolate bars were discontinued by the manufacturer and replaced with 13g chocolate bars. We therefore used 13g, rather than 11g chocolate bars for the majority of data collection. Nutritional differences between these two items can be found in the SOP. Because the majority of participants received the 13g chocolate (not the 11g chocolate), all the related variables (e.g. item grams, kilojoules, nutrients; total grams, kilojoules, nutrients) used weight and nutritional information from the 13g item.

The variable 'snack observation issues for sensitivity analyses' (fcsenst, fasenst) allows data users to identify which fruit and chocolate items were offered to participants.

3.8 Hospital admissions – data from two sources combined to overcome incorrect branching logic

The Parent Questionnaire included questions about the study child's overnight hospital admissions (excluding birth). These were separated into admissions 'in the past 12 months' and 'not in the last 12 months'. Unfortunately, there was an error in the branching logic for hospitalisation questions relating to hospitalisations 'not in the last 12 months'. A total of 641 attending parents reported hospitalisations more than 12 months ago, but additional questions about these admissions were only presented if the attending parent also reported more than one hospitalisation within the last 12 months (64 attending parents).

In preparation for the CheckPoint visit, attending parents were asked to complete a Pre-Visit Checklist. This checklist reminded attending parents of what to bring and asked the attending parent to recall the study child's history of hospital admissions. It was to be used as a prompt when completing the Parent Questionnaire. Therefore, hospitalisation data from the Pre-Visit Checklist has been transcribed and combined with the Parent Questionnaire responses for a more complete dataset.

There were slight differences in the wording of the questions in the Pre-Visit Checklist and Parent Questionnaire. The Pre-Visit Checklist included a single question about the total number of admissions during the study child's lifetime, while the Parent Questionnaire asked separate questions for the number of admissions during last 12 months and prior to the last 12 months.

The Pre-Visit Checklist contained a free-text box for the child's age at each admission, and the response format varied between whole years, years and months, and age ranges. The Parent Questionnaire specified age in years and months, and only asked about child age for admissions prior to the last 12 months.

Regarding the reason for each admission, the Pre-Visit Checklist allowed parents to tick one or more tick-boxes (including "illness" but not including "other") and describe the main reason for the

admission in a free-text box. The Parent Questionnaire allowed a single selection from a drop-down list, not including "illness" but including "other", and further categorical selections if the reason was related to infections.

Data were consolidated by combining data from both sources, then removing obvious duplicate admissions, but keeping the age for admissions within the last 12 months if available, and free-text fields on the Pre-Visit Checklist were coded by a qualified medical professional to the infection-related admission type available on the Parent Questionnaire. This process recovered hospitalisation data for 540 study children. However, some data remains missing because parents who didn't return a Pre-Visit Checklist had a shortened Pre-Visit interview in *Check-In* that didn't include these questions. Flag variables (fchsrce1 to fchsrce10) alert data users to the data source for each hospital admission, and discrepancies between different data sources (fchospflag). Further details are described in the Health, Welfare and Community Services data management SOP.

3.9 Community participation – activities data reshaped and rearranged to categorise activities

The Parent Questionnaire prompted parents to describe their child's participation in community activities over the last 12 months (including community groups or clubs, team sport, individual sport, art, music or performance lessons, classes to learn new skills, religious services or classes, and other activities). They were then asked ‘How many activities?’ within each category of activities their child had participated in. Branching logic presented the corresponding number of free-text fields for the attending parent to report “What was this activity?” Attending parents were also asked additional questions about hours spent per week on the activity and the cost of the activity.

Although a single free-text field was provided for each activity within each category, in some instances more than one activity was reported (e.g. ‘tennis, swimming and running’ entered into the 'individual sport’ category). In some cases, activities were described in unexpected categories (e.g. 'football' entered in the 'art' category).

To assist data users, these data were reshaped and rearranged into variables ready to use in analyses (e.g. fcactc01to fcactc09, see Figure 1 below). This reshaping means that variables do not exactly reflect the way in which questions were asked, but rather how the data collected presented itself. Each activity was assigned a category and sub-category, to allow simple description of the data and remove reliance on string variables (included in our supplementary data file). Where activities were reported in unexpected categories ('football' in the 'art' category), data were moved to a more appropriate category ('team sport'); or moved to the 'other' category. Where multiple activities were described in the same field, a flag variable was created (fcmultir01 to fcmultir14), data from follow-up questions such as cost were removed (because it is not possible to separate out the costs of each activity), and each activity was moved to a separate variable to accurately represent the number of activities the study child participated in. For each activity, the original category in which the activity was reported is provided (fcqtype01 to fcqtype14, from the above example, ‘football’ would have an associated variable showing it was recorded under ‘art’).

Original Data Shape

ID	Club1	Club2	Club3	Team1
148	Chess	Football	Scouts	
297	Puzzle club			Water polo

New Data Shape

ID	Activity1	Code1	Activity2	Code2	Activity3	Code3
148	Chess	53-Board Games	Football	21- Football	Scouts	11- Scouts/Guides
297	Puzzle club	53- Board Games	Water polo	26- Team water sports		

Figure 1. Examples of reshaping the Community Participant data.

3.10 Biospecimens – sample tracking

Pre-barcoded sample collection tubes and envelopes were used to collect the biological samples. At the point of collection, the CheckPoint team member scanned the sample collection vessel (tube or envelope) barcode into the relevant sample field in the participant's data entry page (i.e. linking the participant to the sample collection vessel). In the laboratory, CheckPoint team members kept a parallel set of records where they scanned the sample collection vessel barcode and the barcodes of the aliquot tubes that the sample was transferred into (i.e. linking the sample collection vessel and aliquot storage tubes). Later, the CheckPoint team also linked the participant ID to the aliquot tubes.

Of the approximately 140,000 aliquots and envelopes, 98.5% are linked to a participant. Errors in the linkage process at either sample collection or in the laboratory could occur due to barcode scanner errors, manual barcode input errors when barcode scanners were not available, barcode input to incorrect fields, and database errors (including periodic connectivity, syncing and autocorrect issues). Linkage errors were investigated by comparing participant consent for sample collection (including collection notes), date and time of sample collection and processing, and the order in which aliquots were placed in storage boxes. Data has been corrected where it is clear which participant provided the sample.

References

Bach, M. (1996). The Freiburg Visual Acuity test: Automatic measurement of visual acuity. *Optometry and Vision Science, 73*(1), 49-53.

Bach, M. (2006). The Freiburg Visual Acuity Test-Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology, 245*(7), 965-971. doi:10.1007/s00417-006-0474-4

Carskadon, M. A., & Acebo, C. (1993). A self-administered rating scale for pubertal development. *Journal of Adolescent Health, 14*(3), 190-195.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377-381. doi:10.1016/j.jbi.2008.08.010

Homans, N. C., Metselaar, R. M., Dingemanse, J. G., van der Schroeff, M. P., Brocaar, M. P., Wieringa, M. H., . . . Goedegebure, A. (2017). Prevalence of age-related hearing loss, including sex differences, in older adults in a large cohort study. *Laryngoscope, 127*(3), 725-730. doi:10.1002/lary.26150

Jones, G. T., Watson, K. D., Silman, A. J., Symmons, D. P., & Macfarlane, G. J. (2003). Predictors of low back pain in British schoolchildren: A population-based prospective cohort study. *Pediatrics, 111*(4 Pt 1), 822-828.

Niskar, A. S., Kieszak, S. M., Holmes, A., Esteban, E., Rubin, C., & Brody, D. J. (1998). Prevalence of hearing loss among children 6 to 19 years of age: The Third National Health and Nutrition Examination Survey. *JAMA, 279*(14), 1071-1075.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence, 17*(2), 117-133. doi:10.1007/BF01537962

Semel, E., Wiig, E., & Secord, W. (2006). *Clinical Evaluation of Language Fundamentals - Fourth edition, Australian Standardised Edition (CELF-4 Australian)*. Marrickville, Australia: Harcourt Assessment.