**CHRONIC KIDNEY DISEASE IN CHILDREN COHORT STUDY (CKiD)**

# Data Archive

The CKiD Study is a multi-center, cohort study of children aged 6 months to 16 years with mild to moderate impaired kidney function. Two clinical coordinating centers (CCCs) (at Children's Hospital of Philadelphia and at Children's Mercy Hospital in Kansas City), a central biochemistry laboratory (at the University of Rochester), and a data coordinating center (at Johns Hopkins Bloomberg School of Public Health) have formed a cooperative agreement to conduct a prospective study of chronic kidney disease in children. The study population currently consist of two cohorts. Recruitment of the first cohort of 586 children occurred from January 2005 through August 2009 at 48 pediatric nephrology programs acrossed the United States and two sites in Canada. Recruitment of the second cohort of 305 children occurred from Februry 2011 to March 2014. Participants complete annual visits and since its inception in 2003, the scientific aims of CKiD have been to determine the risk factors for decline in kidney function and to define how progressive decline in kidney function impacts biomarkers of risk factors for cardiovascular disease; neurocognitive function and behavior; and growth failure and its associated morbidity.

The CKiD data archive contains data collected for Cohort 1 and Cohort 2 at baseline and follow-up up to December 31, 2015. The baseline visit is comprised of two components: V1a and V1b. The first component of the baseline visit (V1a) occurred during the participant's initial visit to the clinical site and included procedures to obtain an iohexol-based GFR measurement. For the purpose of obtaining baseline data on neurocognitive function and growth, the second component of the baseline visit (V1b) occurred within 3 months after the first baseline visit for children between 1 and 3 years of age, and occurred within 6 months for children over the age of 3. The one to six month lag between the two components of visit one was necessary because the procedures needed to measure the GFR did not provide an environment conducive to an unbiased assessment of some of the neurocognitive tests. Markers related to the four specific aims were measured every year (e.g., serum creatinine, Cystatin C, standardized clinical blood pressure obtained with a uniformed centrally-calibrated device, pediatric quality of life, height and weight). GFR was measured annually during the first two years and then every two years thereafter. The cardiovascular component was implemented concurrently with the kidney component after the second year. In turn, the core markers of neurocognitive function and growth was measured in the odd visit years from the third year on.

The archive is organized into the following directories:

- CKiD Documentation
- CKiD Data Integrity Check
- CKiD Forms
- CKiD Study Data

## 1. CKiD Documentation

The CKiD_Documentation directory contains:

- **CKiD Protocol.** Protocol amendments are normally performed annually. The directory contains the 2014 protocol. (filename: CKiD Study Protocol OSMB 07-01-13 to 06-01-14_clean.pdf)
- **CKiD Manual of Procedures (MOP)**. As of 2015, the manual was comprised of 40 PDF files. Changes to sections of the MOP are performed throughout the year and are noted by the date in the footer of each section. The directory contains one compiled document of the MOP as of 2015. (filename: CKiD MOP_02-15-14.pdf)
- **CKiD Publications**. List of all CKiD study publications as of December 5, 2017. (filename: CKiD Publications_072816.doc)

## 2. CKiD Dataset Integrity Check

The CKID_Dataset integrity Checks (DSIC) directory contains a report describing the examiniation of the data by statisticians and data management specialists at KIDMAC. Archived datasets were checked for completeness and comparability to the original study data sets.

## 3. CKiD Forms

All CKiD forms contain a footer with the date the form was implemented. Subsequent changes to forms are indicated in the footer in one of two ways:

1. In the case of substantial revision to a form (i.e., a change in the data collected or the addition or deletion of variables), the change will be indicated by assigning a new date to the form. For example, if F01 version 01/01/05 received substantial changes, the version date in the footer would change to the date of implementation of the new version (e.g., from 01/01/**05** to 01/01/**06**).
2. In the case of minor revision to a form (i.e., no change in data collected – correction of typo or change in wording), the change will be indicated by adding a letter after the date in the footer. For example, if F01 version 01/01/05 received minor changes, an "a" would be added to the end of the version date in the footer. If F01 version 01/01/05a were to receive subsequent minor changes, the "a" in the footer would be changed to a "b" (e.g., from 01/01/**05** to 01/01/**05a** to 01/01/**05b**).

The CKiD_Forms directory contains baseline forms in PDF format. Multiple versions of some data forms are provided and as previously noted changes to the forms are noted in the footer.

The CKiD_Forms directory contains baseline forms in PDF format. Multiple versions of some data forms are provided and as previously noted changes to the forms are noted in the footer.

| **Name** | **Descritpion** |
|---|---|
| ADVR | Adverse Event Form |
| F01 | Symptoms List |
| F02 | Smoking, Alcohol, Drug Use and Physical Activity |
| F12 | Smoking, Alcohol, Drug Use and Physical Activity Follow-up |
| F13 | General History Follow-up (includes repeating segment on siblings, F13S1) |
| F13a | Abbreviated General History Follow-up (includes repeating segment on siblings, F13S1) |
| F14 | Medical History Follow-up |
| F15 | Nutritional Assessment and Steroid Use (includes repeating segment on supplements, F15S1) |
| F16 | Sun Exposure |
| F17 | Overall Physical Activity |
| F19 | Hand Grip Test |
| GH | General History (includes repeating segment on siblings, GHS1) |
| HLC01 | Health Literacy Assessment Coding Sheet |
| L01 | Specimen Collection Form for Visit 1a |
| L02 | Specimen Collection Form for Visit 1b |
| L03 | Local Laboratory – Renal Panel Results |
| L04 | Local Laboratory – CBC Results |
| L05 | Central Laboratory – Renal Panel Tests |
| L06 | Local Laboratory – Urine Assay Results |
| L07 | Central Laboratory – Iohexol Concentration Results |
| L08 | Central Laboratory – Intact Parathyroid Hormone (iPTH)<br>and Wide Range C-Reactive Protein (wrCRP) |
| L09 | Central Laboratory – Lipid Profile |
| L11 | Central Laboratory – Cystatin C Results |
| L12 | Central Laboratory – Iron Tests |
| L13 | Central Laboratory – Vitamin D |
| L21 | Specimen Collection Form for Visit 2 |
| L31 | Specimen Collection Form for Visit 3 |
| L41 | Specimen Collection Form for Visit 4 |
| L51 | Specimen Collection Form for Visit 5 *(form discontinued)* |
| MEDS | Medication and Supplement Inventory (MEDSUM_FULL) |
| MH | Medical History (includes repeating segment on nutritional supplements, MHS1) |
| NRC03A | Cognitive/Development Data Coding Sheet: 12 to < 30 months |
| NRC03B | Cognitive/Development Data Coding Sheet: 30 months to < 4 years |
| NRC03C | Cognitive/Development Data Coding Sheet: 4 to < 6 years |
| NRC03D | Cognitive/Development Data Coding Sheet: 6 to < 17 years and older |
| NRC03E | Cognitive/Development Data Coding Sheet: 17 years and older |
| NRC04A | Behavioral Data Coding Sheet: 12 to < 24 months |
| NRC04B | Behavioral Data Coding Sheet: 2 to < 6 years |
| NRC04C | Behavioral Data Coding Sheet: 6 to < 8 years |
| NRC04D | Behavioral Data Coding Sheet: 8 to < 12 years |

| | |
|---|---|
| NRC04E | Behavioral Data Coding Sheet: 12 to < 18years and older |
| NRC04F | Behavioral Data Coding Sheet: 18 to < 21 years and older |
| NRC04G | Behavioral Data Coding Sheet: 21 years and older |
| PE | Physical Exam |
| PFU01 | Phone/In-Person Follow-up Interview Form (includes repeated segment, PFU01S1) |
| PFU02 | Follow-up Site Questionnaire |
| WFU01 | Web-based Follow-up Survey |

Prior to 2016, the following forms were not included: F19, HCL01 and WFU01.

In addition, the CKiD study collected pediatric quality of life data; however, these forms are copyrighted and therefore are not released due to copyright laws.

## 4. CKiD Study Data

The CKiD_Data directory contains 52 SAS data files. Each dataset has corresponding codebooks (<Filename>.cdb) which are text files and should be opened with an editor like notepad. Also, five (5) of the data files contain repeating segment data: GHS1, F13S1, MHS1, F15S1 and PFU01S1. Repeating segment data are groups of similar data captured within the context of a form. For example, date of birth for each sibling, which is considered repeating segment data, is captured within the general history (GH, F13, F13a (followup)) form and the data is stored in GHS1/F13S1. Similarly, information on the types and amounts of nutritional supplements taken is captured within the medical history (MH), nutritional assessment (F15) and Phone/In-Person Follow-up Interview (PFU01) forms, and stored in MHS1/F15S1 and PFU01S1.

In addition to datasets for each of the study and laboratory forms, the directory includes 12 summary files: GFRCALIBRATEDSUMMARY (formerly referred to as GFRSUMMARY), KIDHIST, LABMARKERS, GROWTH, CARDIO, ABPM, ECHO, CIMT, MEDSUM_SHORT, NEUROSUMMARY, SOCDEM and VERT_DATEBASE. The **GFRCALIBRATEDSUMMARY** data file contains the glomerular filtration rate measurements based on calibrated iohexol concentrations. In addition to the calibrated iohexol-based GFR we also provide a bedside estimated GFR using height and serum creatinine. The variables in **KIDHIST** include primary chronic kidney disease (CKD) diagnosis, CKD diagnosis group, date of CKD onset, dates associated with birth, baseline and clinical endpoints of the study (i.e., date of transplant, dialysis, death) and last date free of renal replacement therapy (RRT). For participants who did not experience renal replacement therapy prior to the administrative censoring date (July 31, 2014), a random value between 0.01 and 0.26 was added to the variable duration of time between baseline visit and last date free of renal replacement therapy (*LDATRTFREE* in KIDHIST). The same is done for the variable *LDATALIVE* in KIDHIST, in which participants alive as of July 31, 2014 have a random number between 0.01 and 0.26 added to their time from baseline visit and last date alive. The rationale for doing this is to prohibit the user from determining personal identifiers, namely back-calculating the exact date of study entry (i.e., baseline date) and potentially the exact date of birth. The **LABMARKERS** data file contains variables for laboratory markers such as basic metabolic panel, complete blood count, urine analysis, intact parathyroid, c-reactive protein, lipid panel and iron results as well as calculated proteinuria, acidosis, hypoalbuminemia, abnormal calcium and phosphate (based on KDOQI thresholds), calcium-phosphate product, elevated CRP, anemia and hemoglobin z-scores and

percentiles based on age, sex and race per CDC guidelines. The **GROWTH** data file contains height, weight and body mass index (BMI) average, percentile and z-score variables, and birth history variables. The **CARDIO** data file contains the casual/clinical blood pressure percentile, z-score and index variables. The **ABPM** data file contains ambulatory blood monitoring (ABP) variables. The **ECHO** data file contains echocardiogram measurement data. The **CIMT** data file contains variables for the measurement of caritod intimal medial thickness (cIMT) in a subset of the study population. The **MEDSUM_SHORT** data file contains one record per medication per participant-visit and summarizes whether or not the study participant has been prescribed during the past 30 days any medication that falls into one of several major medication classes including antihypertensives, ESAs, growth hormones, immunosuppressives, anticholinergics, and antidepressants. More detailed information for the medications including dosing amounts, schedules and adherence are provided in the MEDSUM_FULL data file. **NEUROSUMMARY** data file contains key variables from the neurocognitive and behavioral battery. **SOCDEM** data file contains sociodemographic variables such as race, ethinicity, income and maternal education. **VERT_DATEBASE** data file contains all the study visits, and status of each visits (i.e., regular, irregular, transitional or disenrolled).

To enhance the existing quality control measure, a supplementary measure was performed. For each variable in each summaryfile, the first and third quartiles (25th and 75th percentiles), Q1 and Q3 were calculated. Any values that were larger than Q3/Q1 or smaller than 4×Q1 – 3×Q3 were flagged as "far-out values". These limits were based on the methodology of Tukey fences: 3 Tukey fences in the log scale for high values and in the raw scale for low values. The implementation of this quality control methodology yielded subsequent checks for values in the LABMARKERS and ECHO summary files. Flagged values were double-checked to determine whether the measurement entered was a feasible value based on the distribution of the data and/or the participant's previous history for that measurement. The percentages of flagged values were 0.1% in the LABMARKERS and <0.1% in the ECHO summary files. There were no values flagged in the other summary files. After futher investigation of the values in the LABMARKERS and ECHO summary files, approximately 20% of the values were corrected. Uncorrected values were recorded as "-90" to indicate that original values was a far-out value.

Prior to 2017, the following summary files were: LABMARKERS, ABPM and CIMT.

| No. of Records | Dataset |
|---|---|
| 1367 | abpm.sas7bdat |
| 102 | advr.sas7bdat |
| 5128 | cardio.sas7bdat |
| 410 | cimt.sas7bdat |
| 1540 | echo.sas7bdat |
| 3866 | f01.sas7bdat |
| 443 | f02.sas7bdat |
| 1597 | f12.sas7bdat |
| 1934 | f13.sas7bdat |
| 1058 | f13a.sas7bdat |
| 2177 | f13s1.sas7bdat |
| 2985 | f14.sas7bdat |
| 3709 | f15.sas7bdat |
| 177 | f15s1.sas7bdat |
| 1665 | f16.sas7bdat |
| 1248 | f17.sas7bdat |
| 538 | f19.sas7bdat |
| 5045 | gfrcalibratedsummary.sas7bdat |
| 894 | gh.sas7bdat |

| | |
|---|---|
| 1082 | ghs1.sas7bdat |
| 5119 | growth.sas7bdat |
| 608 | hlc01.sas7bdat |
| 891 | kidhist.sas7bdat |
| 891 | l01.sas7bdat |
| 827 | l02.sas7bdat |
| 4185 | l03.sas7bdat |
| 4286 | l04.sas7bdat |
| 4293 | l05.sas7bdat |
| 3077 | l06.sas7bdat |
| 2842 | l07.sas7bdat |
| 2301 | l08.sas7bdat |
| 1916 | l09.sas7bdat |
| 4114 | l11.sas7bdat |
| 2545 | l12.sas7bdat |
| 124 | l13.sas7bdat |
| 807 | l21.sas7bdat |
| 1304 | l31.sas7bdat |
| 1111 | l41.sas7bdat |
| 179 | l51.sas7bdat |
| 5122 | labmarkers.sas7bdat |
| 17828 | medsum_full.sas7bdat |
| 4288 | medsum_short.sas7bdat |
| 891 | mh.sas7bdat |
| 14 | mhs1.sas7bdat |
| 4103 | neurosummary.sas7bdat |
| 29 | nrc03a.sas7bdat |
| 54 | nrc03b.sas7bdat |
| 122 | nrc03c.sas7bdat |
| 1761 | nrc03d.sas7bdat |
| 274 | nrc03e.sas7bdat |
| 10 | nrc04a.sas7bdat |
| 195 | nrc04b.sas7bdat |
| 169 | nrc04c.sas7bdat |
| 571 | nrc04d.sas7bdat |
| 1123 | nrc04e.sas7bdat |
| 158 | nrc04f.sas7bdat |
| 51 | nrc04g.sas7bdat |
| 5121 | pe.sas7bdat |
| 494 | pfu01.sas7bdat |
| 824 | pfu01s1.sas7bdat |
| 528 | pfu02.sas7bdat |
| 5119 | socdem.sas7bdat |
| 6269 | vert_datebase.sas7bdat |
| 45 | wfu01.sas7bdat |