

# Nairobi Newborn Study

## Standard Operating Procedure 6 [SOP6]: Data management

Written by Georgina Murphy

Version 1: 09.09.15

### 1. Purpose and scope

Standard Operating Procedures (SOPs) are required to assure quality, to reduce errors, and to standardise activities and tasks throughout the Nairobi Newborn Study.

This SOP outlines the procedures for data handling and survey data checks. Secure procedures for inputting, synchronisation, transferral and storage of data are necessary to ensure that data are not lost and are always kept securely in order to protect the data and the participants from whom they were collected. Reviewing the data after collection is important in ensuring accurate and consistent results.

### 2. Staffing, equipment, and resources

Mobile computers with internet modems and pre-loaded REDCap tools will be required for each data entry officer and nursing interviewer.

### 3. Procedure

#### 3.1 Data collection and entry

- REDCap will be used for direct electronic data collection for the nursing questionnaire and abstraction of data from maternal and newborn admissions registers and newborn medical records.
- These REDCap data collection tools are designed with inbuilt skips to structure the data entry form in accordance to data entered in earlier fields in the form. Validation checks are also inbuilt to prompt the data entry officer if data are entered that conflict with the expected answer type (e.g. integer) or outside the value range. Furthermore, required data entry fields are marked as such and if left blank, the data entry officer is prompted to complete the field.
- Data will be collected on a paper form for the structural assessment. These forms will then be data entered into a REDCap tool at the office. Data will be entered twice and cross-checked. Any inconsistencies will be resolved by the study supervisor by referring to the paper form.
- Before going to a new facility, the first ID number for each data clerk and nurse interviewer will be entered by data management at the office. When entering the first record, the data enterer must select to edit this existing but empty record. Subsequent records for that facility should be entered by selecting 'add record', the ID number will then be autogenerated.

#### 3.2 Number of records entered per facility

- A minimum of 500 records are to be entered from the maternal admissions register within the timeframe [1<sup>st</sup> April 2014- 31<sup>st</sup> March 2015]. For facilities with ≤510 admissions annually, all records will be entered. In facilities with >510 admissions annually, the data entry officers will follow instructions on selecting a random sample of records (see SOP4.0). In all cases, the number of entries should be tracked by the data entry officers themselves and checked by the data management team to ensure that a minimum of 500 records are collected at each facility. In those facilities with >510 admissions annually, data entry should be stopped at 500 records even if admissions for a randomised day or week have not been completely entered.

## Nairobi Newborn Study

### Standard Operating Procedure 6 [SOP6]: Data management

Written by Georgina Murphy

Version 1: 09.09.15

- The number of newborn medical records to be entered will vary from facility to facility. Proportional sampling will be done where the same proportion of annual records will be sampled from each facility. Since the total number of annual records will vary per facility, so too will the number of records to be entered at each facility. Combined, a total of 800 records across all facilities will be entered. The data entry officers will be notified about the number of records to enter at a given facility. As above, it is the responsibility of both the data entry officers and the data management team to check that the required number of records has been entered before moving on to the next facility.

#### 3.3 Data checks for registers and records

##### Daily data checks

- Before synchronising data from REDCap at the end of each day, the data clerks will run a cleaning script. The fields to be checked are outlined in **appendix 1**.
- Where the script picks up on errors in the data entry, these errors will be communicated to data entry officers.
- The officers will check the paper records and ensure that the data they have entered is correct or update it when incorrect.

##### Weekly data checks

- At least weekly, data checks will be performed for data quality assurance purposes. Data included in these checks will be accumulative data for a single facility (i.e. if three weeks' worth of data have been collected in a single facility, then the checks will be run on the dataset that contains all three weeks, not just data collected since the last report). The checks outlined in **appendix 2** will to be performed. The results of these checks will be compiled by the data management team into a report for review by the study research officer.
- It will be the responsibility of the research officer to discuss these reports with the study management team. Where necessary, updating of the data collection tools, re-training of staff, and/or updating of look-up lists, should be considered.

#### 3.4 Data checks for nursing questionnaire

##### Daily data checks

- At the end of every day, a cleaning script will be run on nursing questionnaire data collected on that day by the research officers who conducted the interviews. The checks outlined in **appendix 3** will to be performed.
- When inconsistencies in the data are found, the research officers should consider the query and correct those which they realise to be errors of data entry. In the case where a query needs to be resolved by asking the nurse who was interviewed, this can be done for any section of the questionnaire apart from section two which examines the knowledge of the nurse. Re-asking these questions provides an opportunity for the nurse to change their answer, which is not in line with the aims of the questionnaire.

##### Weekly data checks

- At least weekly, data checks will be performed for data quality assurance purposes. Data included in these checks will be accumulative data for a single facility (i.e. if three weeks' worth of data

## Nairobi Newborn Study

### Standard Operating Procedure 6 [SOP6]: Data management

Written by Georgina Murphy

Version 1: 09.09.15

have been collected in a single facility, then the checks will be run on the dataset that contains all three weeks, not just data collected since the last report). The checks outlined in **appendix 3** will to be performed. The results of these checks will be compiled by the data management team into a report for review by the study research officer.

- It will be the responsibility of the research officer to discuss these reports with the study management team. Where necessary, updating of the data collection tools, re-training of staff, and/or updating of look-up lists, should be considered.

#### 3.5 Per facility checks and quality assurance

- 10% of all newborn medical records at each facility will be re-entered. At the end of data collection at a facility, a random sample of 10% of records will be selected by the data management team. The data entry officers will locate these records. The research officer/supervisor will re-enter these records. The data management team will then cross-check the double entry for consistency. Where inconsistencies have arisen, the records will be reviewed again. Where agreement is unclear, final decision on the correct data will be made by the study coordinator.
- As previously mentioned, each structural assessment will be doubled entered by two different research officers. These will be cross-checked for consistency. Any inconsistencies will be checked against the paper version. Where agreement is unclear, final decision on the correct data will be made by the study coordinator.
- Before completion of data collection at a facility, it is the responsibility of the study coordinator to check that each nursing questionnaire has been appropriately consented form.
- For each data entry tool, it is the responsibility of the research officer to extract any comments inserted into the comments bubbles on the local machines/laptops, discuss these comments with the data entry officers, and report the relevant comments to the study coordinator.

#### 3.6 Data cleaning

- Data cleaning will consider the following actions:
  - Recoding of -1 to 'missing'
  - Separating out lists with colons : into separate variables
  - Checking free-text field to see if and how they can be categorised
  - Restructuring of variables e.g. merging the two variable 'weight in g' and 'weight in kg' into one single variable with consistent units.
- No changes should be made directly to the original data set.
- Where changes are made, they should be done so on a copy of the data set. A detailed log of all changes made to those data, compared to a previous version of (or the original) dataset, must be kept.

#### 3.7 Data synchronisation

- All data entry laptops are installed with a synchronisation module, which includes each of the data entry tools and is configured to the main KWTRP server. This synchronisation module enhances the security of data transfer and works through a token system that can identify the source of data.
- Each data collection tool is clearly labelled on the synchronisation interface, it's thus important for the data entry person to ensure that data from all tools in use are synchronised.

# Nairobi Newborn Study

## Standard Operating Procedure 6 [SOP6]: Data management

Written by Georgina Murphy

Version 1: 09.09.15

- Data should be synchronised after verification and correction of any inconsistencies from the data checking process.
- Synchronisation is done at the end of every day during which data were entered.
- The data clerk is notified on the status of the synchronisation as either **successful** or **failed**. If failed, the clerk is be able to try again and again until it's successful. In case there is a problem, the clerk should conduct a simple diagnosis by checking the modem signal colour strength and its connectivity. For further clarification, the clerk should contact the data managers for assistance.

### 3.8 Data storage

- Data will be securely stored and backed-up on KWTRP and Oxford servers. On a daily basis, data clerks are required to download, compress, and encrypt a copy of the data from each data collection tool. A copy of this encrypted data is transferred to the KWTRP office. On a daily basis, data on the main server at KWTRP are backed up. This is a major automated backup for all data stored within the REDCap database. A replica of the data from the REDCap database is pushed to a different server for storage. This process is automated and monitored by the data management team.
- Data will not be stored or transported on portable drives (e.g. USB) or unprotected computers.
- All paper-based documents (structural assessments, permission forms, consent forms, etc) will be scanned and securely stored in digital format. Note: consent forms must not be stored in the same physical or electronic location as data as the identity of the nurse's data might be compromised.
- All paper-based documents must be stored in a locked cabinet.

### 4. Related procedures and documents

SOP4.1: Data entry from maternal admission register

SOP4.2: Data entry from newborn admission register

SOP4.3: Data entry from newborn medical record

SOP5.1: Nursing questionnaire instructions

### 5. SOP-user signature log

By signing in the table below, I confirm I have read and am familiar with the SOP for data management.

Name	Signature	Date

# Nairobi Newborn Study

## Standard Operating Procedure 6 [SOP6]: Data management

Written by Georgina Murphy

Version 1: 09.09.15

### Appendix 1: Daily cleaning script checks for register and record data

[These checks are to be reported directly back to data entry officers at the end of each day]

#### All records

1. Check if variables marked as required in REDCap are missing. Allowances will be made for branching logic. I.e. if variable x is mandatory but only appears when y=1, then when y=1 x must not be missing but when y=0 then x can be missing.
2. Check all soft validations on ranges (being sure to ignore -1 from the ranges) and data type
3. Check that today's date (date\_today) is actually the date of data entry.
4. Date\_adm should not be later than date\_today and date\_adm should be within study time frame 01July2014 to 30June2015 unless it is 1915-01-01.
5. Date\_discharge should not be earlier than date\_adm unless date\_discharge is 1915-01-01 to indicate missing date\_discharge.
6. Loc and other\_loc\_1 must not both contain information
7. Sub\_loc and other\_sub\_loc\_1 must not both contain information
8. Check that the ID number is in the correct structure [hospital ID][data clerk ID][00001] and that there are no duplicates in ID number
9. Check that all forms for each instrument have been selected as completed and saved.

#### Maternal admission register

1. Ref\_out must =no (0) if there is subsequent data available on multi\_birth, outcome\_1, or sex\_1

#### Newborn admission register

1. Date\_of\_birth should not be later than date\_today and not later than date\_adm.
2. One of birth\_weight\_kg or birth\_weight\_g must be answered. Both should not be answered unless they are -1.
3. One of ad\_weight\_kg or ad\_weight\_g must be answered. Both should not be answered unless they are -1.
4. If dx1\_primary=1 then at least one of dx1\_adm OR dx1\_adm\_other must be entered
5. If dx2\_other=1 then at least one of dx2\_other\_1 OR dx2\_other\_3 must be entered
6. If dx1\_primary=0 then at least one of dx1\_other\_1 OR dx1\_other\_4 must be entered
7. Only one of either dx1\_adm OR dx1\_adm\_other should have data entered
8. If dx1\_dsc\_primary=1 then at least one of dx1\_dsc OR dx1\_dsc\_other must be entered
9. If dx2\_dsc\_other=1 then at least one of dx2\_dsc\_other\_1 OR dx2\_dsc\_other\_3 must be entered
10. If dx1\_dsc\_primary=0 then at least one of dx1\_dsc\_other\_1 OR dx1\_dsc\_other\_4 must be entered
11. Only one of either dx1\_dsc OR dx1\_dsc\_other should have data entered

#### Newborn medical record

1. Date\_of\_birth should not be later than date\_today and not later than date\_adm.
2. One of birth\_weight\_kg or birth\_weight\_g must be answered. Both should not be answered unless they are -1.

## Nairobi Newborn Study

### Standard Operating Procedure 6 [SOP6]: Data management

Written by Georgina Murphy

Version 1: 09.09.15

3. One of ad\_weight\_kg or ad\_weight\_g must be answered. Both should not be answered unless they are -1.
4. Temp should be written to one decimal place
5. If dx1\_primary=1 then at least one of dx1\_adm OR dx1\_adm\_other must be entered
6. If dx2\_other=1 then at least one of dx2\_other\_1 OR dx2\_other\_3 must be entered
7. If dx1\_primary=0 then at least one of dx1\_other\_1 OR dx1\_other\_4 must be entered
8. Only one of either dx1\_adm OR dx1\_adm\_other should have data entered
9. If adm\_rx=1 then at least one of adm\_rx1 OR adm\_rx4 should have data entered
10. If dx1\_dsc\_primary=1 then at least one of dx1\_dsc OR dx1\_dsc\_other must be entered
11. If dx2\_dsc\_other=1 then at least one of dx2\_dsc\_other\_1 OR dx2\_dsc\_other\_3 must be entered
12. If dx1\_dsc\_primary=0 then at least one of dx1\_dsc\_other\_1 OR dx1\_dsc\_other\_4 must be entered
13. Only one of either dx1\_dsc OR dx1\_dsc\_other should have data entered

## Nairobi Newborn Study

### Standard Operating Procedure 6 [SOP6]: Data management

Written by Georgina Murphy

Version 1: 09.09.15

#### Appendix 2: Weekly/periodic data checks for register and record data

[These checks will be compiled into a report for regular review by the study research officer]

- Summarise the rate of data entry per data clerk. I.e. how many records were entered each day by each data clerk.
- Check that the branching logic is working properly. I.e. variables that should not appear given a previous answer, remain empty.
- Report any incidences where data errors found during the daily checks were not corrected i.e. the errors checked for during daily checks appear in this synchronised dataset.
- Report the frequency with which -1 is recorded for each (and all!) variables. I.e. quantify the missingness for each variable. Report this for the total dataset and also stratified by data clerk.
- Tabulate the free text fields for all relevant variables (this will include the answers that don't appear on the look up lists)
- Report the proportion of values which lie outside the validation range (outliers) for each variable for which a range has been set in REDCap. Report this for the total dataset and also stratified by data clerk.

## Nairobi Newborn Study

### Standard Operating Procedure 6 [SOP6]: Data management

Written by Georgina Murphy

Version 1: 09.09.15

#### Appendix 3: Nursing questionnaire checks

##### Daily cleaning scripts

1. Check if variables marked as required in REDCap are missing. Allowances will be made for branching logic. I.e. if variable x is mandatory but only appears when y=1, then when y=1 x must not be missing, but when y=0 then x should be missing.
2. Check that today's date (date\_today) is actually the date of data entry.
3. Check that the ID number is in the correct structure [hospital ID][data clerk ID][00001] and that there are no duplicates in ID number
4. Check that all forms for each instrument have been selected as completed and saved.

##### Weekly/periodic reports

- Tabulate use of 'don't know= -1' and 'none of the above=0' for the five instruments entitled 'Knowledge:...'
- Tabulate the free text fields for all relevant variables
- Report any incidences where data errors found during the daily checks were not corrected i.e. the errors checked for during daily checks appear in this synchronised dataset.
- Check that the branching logic is working properly. I.e. variables that should not appear given a previous answer, remain empty.