Monitoring plan for the SafeBoosC-III follow up study

Version	Author(s)	Date	Changes	Approved by
1.0	Marie Isabel S Rasmussen	22.03.23	Initial version	Gorm Greisen, Mathias L Hansen, Markus Olsen and Janus Jakobsen
2.0	Marie Isabel S Rasmussen	10.08.24	Change of site % above average of lost to follow up and informal assessments (F00-F02) Addition of predefined list of "other neurodevelopmental test"' (F15) Addition of date check (F17-20) Change of frequency of data quality monitoring report	Gorm Greisen

The purpose of the central monitoring plan is to maximize data quality. This will be done by identifing centres with 1) missing data, and 2) centres with predefined quality deficiencies or noteworthy data deviations as identified by visual inspection of data distributions.

As monitoring of missing data will be done once a month, while the additional monitoring will be done every third month, the central monitoring plan is divided into two parts; 1) missing data monitoring, and 2) monitoring of quality deficiencies and noteworthy data deviations.

Part 1 - Missing data

For each centre we will report missing data, for both clinical data as well as parental questionnaires.

1) Clinical data (OpenClinica)

We collect clinical data on participants between 18-30 months of corrected age. The eCRF is hosted on the platform OpenClinica. To complete a data form, the investigator must perform the data entry "blinded 2-year follow up", all relevant fields are filled out and press "submit".

When the participant surpasses 24 months of corrected age + 1 month and has not had a completed data entry, it will count as missing data in the system. Even though the data is not per definition missing until after 30 months of corrected age, most participants will have a routine follow up around 24 months of corrected age, and therefore this threshold is set to remind the investigator to fill out the data, when available.

2) Parental questionnaires (REDCap)

This is hosted on the platform REDCap. The parental questionnaires are standardized to be filled out between 23,3-27,5 months of corrected age. To complete a questionnaire, a language must be chosen, the

participants study ID must be reported, the questionnaire must be answered, and the submit button must be pressed.

Course of action regarding missing data

Once a month, data extraction on missing data entries will be generated from OpenClinica and REDCao. Data completion will be reported for each centre and reported in the monthly newsletter which will be circulated to all investigators and uploaded to 'safeboosc.eu'. The identified study IDs for participants with missing data will not be included in the report. However, the study manager will send these to the relevant primary investigators. Missing data monitoring will start after 100 babies have been randomised.

Part 2 – quality deficiencies

We will aim at identifying centres with noteworthy data deviations. Any noteworthy data deviations will be defined as 1) outliers 2) systematic deviations in data due to suspected misunderstandings; or 3) Suspected fabricated data

1) Outliers
   For the quantitative measures, we will analyse plots for outliers. Since there is input check ranges on almost all quantitative measures in OpenClinica, we expect few random errors.

2) Suspected misunderstandings/deviations for the following datapoints:

   - F00: "Is the participant lost to follow up". If a site has more than the average percentage of lost to follow up, the investigator will be contacted to confirm this.
   - F02: "Are there relevant health care records, such as formal assessments of vision, hearing, psychomotor, neurodevelopment, from corrected age 18-30 months available?" if a site has more than the average percentage of lost to follow up, the investigator will be contacted to confirm this.
   - F07: "Has the child been diagnosed with cerebral palsy". If a Global Motor Function Classification System score has been reported = 1, the investigator will be asked to confirm this.
   - F14b: "Cognitive composite score". If a score has been reported that does not follow the table of 5 (85-90-95-100-105 etc), it is most likely not a composite score, but the raw score which has been report. In this case the investigator will be contacted and asked to correct this. This only applies for Bayley-III scores, as Bayley-IV scores follow another scoring regimen.

- F15: “other neurodevelopmental tests”. The scores will be screened for any notable deviations. Not all tests will be known, but for most common neurodevelopmental tests, this will be possible. If anything stands out, investigators will be contacted and asked to confirm.
- F15: “other neurodevelopmental tests”. If a test which is not one of the predefined neurodevelopmental test, the investigator will be asked to delete this and do an informal assessment if there is no Bayley and no other neurodevelopment test done.
- F17-F20: if the inputted date the child has been seen is before 18 months of corrected age, the date should be double checked and if the date is before 18 months of corrected age, the form should be changed to an informal assessment in F02.

3) Suspected fabricated data

   Suspected fabricated data will be defined as an unexpected distribution or variance in data of each centre, that may represent fabricated data. This will be relevant for the binary data. An unexpectedly large or small prevalence of a binary data point can also be a sign of data fabrication, but it is also likely that this could be due a misunderstanding (see suspected misunderstandings above).

Course of action regarding quality deficiencies and noteworthy data deviations

Every month will generate a report on outliers for clinical data. We will only include data from centres with more than ten included participants, and for every variable a minimum of five participants must be added to be presented. This is to ensure some degree of anonymisation and due to the fact that data assessment is difficult in very small samples. Centre IDs will be blinded under a new name in the report. Two months (11th of August 2024) before completion of the study and at the end of the study (20th of October 2024), a data report focusing on quality deficiencies and noteworthy data deviations, will be generated after extraction of data from the eCRF. The data report will be assessed and results from the monitoring will be logged in a ‘central monitoring log’ by the monitoring group.