Safeguarding the Brains of our smallest Children

SafeBoosC-IIIv

Central data monitoring plan

Version	Author(s)	Date	Changes
1.0	Mathias Lühr Hansen Markus Harboe Olsen Johanne Juul Petersen Gorm Greisen Janus Christian Jakobsen	24-02-25	Initial version
1.1	Mathias Lühr Hansen Markus Harboe Olsen Johanne Juul Petersen Caroline Kamp Janus Christian Jakobsen	13-08-25	Addition of section on sites participating in substudies

Contents

Purpose

Central data monitoring for step one of the SafeBoosC-IIIv trial (90 days outcome assessment) will be conducted as outlined in this central data monitoring plan. The purpose of central data monitoring is to optimize data quality by identifying:

1) Participants with missing data
2) Sites with pre-defined quality deficiencies or noteworthy data deviations

Safety monitoring is not included in the central data monitoring plan; this task is delegated to the Data Monitoring and Safety Committee.

The central data monitoring group will include Caroline Kamp, Johanne Juul Petersen, Janus Christian Jakobsen, Markus Harboe Olsen, two board-certified neonatologists, and Mathias Lühr Hansen.

Blinding

The data manager will provide two separate data sets. Data will be provided without any information about the group assignment to assess quality deficiencies and noteworthy data deviations. A second dataset with scrambled participants' IDs will be provided to assess quality deficiencies and the variable 'Change of medical management due to cerebral hypoxia' (see section on 'Noteworthy data deviations'). This dataset will have information about intervention (coded as A or B). All participants from sites with fewer than five participants included will be removed from the second file.

Missing data

The first part of the central data monitoring plan will focus on identifying participants with missing data entries.

Form completeness

Each participant enrolled in the trial has two data entry forms that should be completed at a specific due date. Participants randomised to the cerebral oximetry group have a third data entry form on Serious Adverse Reactions (SARs) (**Table 1**). Since this is a pragmatic trial with limited resources and time allocated to the principal investigators, it is unrealistic to expect data entry forms to be completed at the exact due date. Therefore, a data entry form will not be flagged as 'missing data' until 10 days from the due date, defined as 'overdue date' (Table 1).

Table 1. Due and overdue time points for completion of data entry forms.

Form	Due	Overdue
SARs	28 days (from birth)	38 days (from birth)
End of monitoring	28 days (from birth)	38 days (from birth)
90 days follow-up	90 days (from randomisation)	100 days (from randomisation)

Flagging of data entry forms

The 'enrolment list' report in the electronic case report form (eCRF) indicates the completion status of each data entry form for each participant. Once a month, the 'enrolment list' will be exported from the eCRF and used to assess the completion of the data entry forms for each enrolled participant. Only participants that have reached 28 days from birth will be included in the data completion report since participants that have not reached this time point do not have 'missing data' yet, as per the definition of missing data below. Depending on the completion status, the data entry forms will be flagged as follows:

Waiting for data entry

A form will be marked as 'waiting for data entry' if the overdue date has not surpassed (e.g., the end of monitoring form for a participant 30 days after birth, will be flagged as 'waiting for data entry' if the form has not been completed, since the due date of 28 days from birth has surpassed – but not the overdue date of 38 days from birth).

Missing data

A form will be flagged as 'missing data' if the overdue date has surpassed (e.g., the end of monitoring form for a participant 39 days after birth will be flagged as 'missing data' if the form has not been completed, since the overdue date of 38 days from birth has passed).

Completed data

A form will be marked as 'completed data' if the data entry form has been completed, regardless of the previous status of the form (e.g., a form that was previously flagged as 'missing data' will, after the form has been completed, be flagged as 'completed data' like those forms that was completed in due time).

For infants that have reached 28 days from birth and thereby are included in the data completion report, but have not reach 90 days from randomisation, the 90 days follow-up form will also be marked as 'completed data', since neither the due nor the overdue time point has been reached for this form.

Sites participating in substudies

For sites participating in substudies, we will monitor completion of substudy data entry forms based on the principles described above.

For the substudy on renal outcomes, the due time point will be 90 days from randomisation and the overdue time point will be 100 days from randomisation.

For the substudy on hypocapnia, the due time point will be 28 days after birth and the overdue time point will be 38 days after birth.

Quality deficiencies

The second part of the central data monitoring will focus on identifying quality deficiencies and noteworthy data deviations.

We will aim at identifying sites with prespecified quality deficiencies as defined below:

1. Lack of cerebral oximetry monitoring during the first episode of invasive mechanical ventilation, despite being randomised to the cerebral oximetry group (E09 = No)
2. Discontinuation of cerebral oximetry monitoring for more than 12 continuous hours during the intervention period if the participant is randomised to cerebral oximetry (E10 = Yes)
3. Late initiation of cerebral oximetry monitoring in hours, defined as >6 hours from initiation of invasive mechanical ventilation if the participant is randomised to cerebral oximetry (E11 >6 hours)
4. Termination of cerebral oximetry monitoring due to other reasons than the four stopping criteria depicted in the protocol if the participant is randomised to cerebral oximetry (E15 = Other)
5. Deviation from per-protocol intervention (E09 = No OR E10 = yes OR E11>6 OR E15 = Other)
6. Visible cerebral oximetry monitoring despite being randomised to the usual care group (E17a = Yes)
7. Invalid days from birth to cerebral oximetry (E12 ≠ days from S01 to R05 (+1 day)
8. Continued consent/assent after randomisation by the parents (E18 = No OR E19 = No OR E20 = No)
9. Eligible infants not randomised, i.e., screening failures (S05 = No OR S06 = Consent not sought OR S06 = Opt-out AND S06b = No OR S06c = Yes OR S06 = prior consent AND S06a = no OR S09 = No)

Assessment

For visualisation of each quality deficiency (except for the assessment of screening failures, see below), we will create a stacked bar chart showing the proportion of participants in the different categories for each of

the variables per site. To allow further exploration of the registered quality deficiencies, the trial manager may request the IDs of participants registered with the specific quality deficiencies.

Assessment of screening failures

A screening failure will be defined as a participant fulfilling the following eligibility criteria: Postnatal age <28 days (S01<28 days from actual date); GA >28 weeks (S03 = Yes); expected to receive mechanical ventilation >24 hours (S04 = Yes); no suspicion of or confirmed brain injury disorder (S07 = No); no suspicion or diagnosis of congenital heart malformation likely to require surgery (S08 = No)

And where the infant was still not randomised due to lack of adherence to the additional eligibility criteria or due to other (specified reason) as defined below:

- No possibility of initiating cerebral oximetry monitoring <6 hours from initiation of mechanical ventilation (S05 = No)
- Consent not sought (S06 = Consent not sought)
- Infants screened with opt-out and no parental information provided or parental opt-out before randomisation (S06 = Opt-out AND S06b = No OR S06c = Yes)
- Infants screened with prior consent and no parental consent (S06 = Prior consent AND S06a = No)
- Not randomised despite adhering to all eligibility criteria (S09 = 0, possibility to specify)

For visualisation, we will

1) Create a flow chart that includes the number of screened participants, the number of screening failures, including reasons, and the number of randomised participants.
2) Create a stacked bar chart showing the proportion of screening failures per site.

Noteworthy data deviations

We will aim at identifying sites with noteworthy data deviations, defined as:

1) Outliers due to suspected random errors in data entries or insufficient validation ranges in the eCRF (suspected outliers)
2) Suspected systematic errors in data entries due to misunderstandings (suspected misunderstandings)
3) Potentially fabricated data (suspected fabricated data).

Suspected outliers

Suspected outliers of the eCRF are defined as outlying observations for continuous data points in the monitoring data report, identified by visual inspection. If such a data entry was identified in the data monitoring report, it would be marked as a suspected outlier.

Suspected misunderstandings

Suspected misunderstandings are defined as any unexpected differences in the distribution of data among sites, that may represent differences in coding practice or misinterpretation of the eCRF, or the overall study design. An example of this could be a local misunderstanding of the definition of a trial outcome, e.g., diagnosing the outcome by different criteria than described in the study protocol. In such a situation, we would expect an abnormally high or low frequency of the event for binary data, or a shift/abnormally large variability in the data distribution for a continuous outcome, in the site where the misunderstanding occurred.

Suspected fabricated data

Suspected fabricated data is defined as an unexpected distribution or variance in data at each site, that may represent fabricated data. For continuous data, we would expect to see a different shape or distribution of data when data visualised graphically, since natural variance is difficult to fabricate. An example of this could be an unexpected narrow or wide distribution of continuous variables at a particular site. An unexpectedly large or small prevalence of a binary data point can also be a sign of data fabrication or due to a misunderstanding (see suspected misunderstandings below).

Assessment

Assessment of continuous variables

We will create box plots showing the median, interquartile range, upper and lower range, and outliers for each of the following variables per site.

- Randomisation age in days and hours (time from S01 to R05)
- Gestational age at randomisation in weeks and days (E01 numeric)
- Hospital-free days based on health care records (F08 numeric)
- Parental-reported additional hospital days (F09 numeric)
- Total number of hospital-free days (F10 numeric)
- Days with invasive mechanical ventilation (F21 numeric)

Assessment of binary and categorical variables

We will create a stacked bar chart showing the proportion of participants in the different categories for each of the following variables per site.

- Consent method used for randomised infants (R05 timestamp AND S06 prior informed consent, deferred consent, opt-out, or consent not sought)
- Reason for mechanical ventilation (E08 = surgery to allow for optimal pain relief, surgery for physiological/anatomical reasons, primary pulmonary problem, other)
- Birth weight <-2/>+2 SD (E01 gestational age and E02 birthweight to determine)
- Change of medical management due to cerebral hypoxia (E14 = yes)
    - Intervention categories (E14a = volume expansion, vasopressor/inotrope, PDA treatment, RBC transfusion, FiO2 adjustment, ventilator settings, other)
- Reason for termination of cerebral oximetry monitoring (E15 = stabilised cardio-pulmonary function, extubated, >28 days from birth, deceased, other)
- Death (F01 = Yes)
- Reintubation and mechanical ventilation beyond first episode (F04 = Yes)
- Bronchopulmonary dysplasia (F11 = Yes)
- Brain injury (F12 = Yes)
- Seizure (F13 = Yes)
- Vasoactive drugs or hydrocortisone due to hemodynamic insufficiency (F14 or F15 = Yes)
- ECMO (F16 = Yes)
- NEC or SIP (F17 = Yes)
- Renal replacement therapy (F18 = Yes)
- Late-onset sepsis (F19 = Yes)
- Invasive mechanical-ventilation related infection (F20 = Yes)

Data completion report and corrective actions

Missing data report

Based on the monthly data extraction from the eCRF on completion of data entry forms, a site-specific completion rate for each form (SARs, end of monitoring, 90 days follow-up) will be calculated. Data form completion will be calculated by dividing the number of forms flagged as ‘completed’ by the number of forms expected to be completed (forms that have surpassed the overdue date).

Example:
In site DK01, 10 participants have surpassed the overdue date (>100 days from randomisation) for the 90 days follow-up form, but the form has only been completed for 7 participants.

*90 days follow-up form completion for site DK01: 7/10*100 = 70%*

The percentage of incomplete forms for each site will be reported in a monthly data completion report. The report will be assessed and approved by the central data monitoring group and uploaded to 'www.safeboosc.eu'.

Site-specific reports

A site-specific missing data reported will automatically be generated as well. This report will include the following information:

- Completion rate for each data entry form for all sites in the trial combined
- Completion rate for each form in the specific site
- Overview of completion status for each form per participant in the specific site

To minimise missing data and loss to follow-up for the 90-day primary outcome, the site-specific report will also include a list of participant IDs for the specific site, on participants where the time to primary outcome assessment is less than 30 days, including the specific due date for primary outcome assessment (review of clinical records and contact to parents).

The trial manager will, each month, forward the site-specific report to the relevant principal investigator. As there will be zero tolerance for missing data entries, the investigators will be requested to fill out the missing data entries as soon as possible.

Missing data monitoring will start after 100 participants have been randomised. An example on the overall missing data report, and the site-specific missing data report, including the participant-specific missing data overview available in the Appendix (Figure A1, A2 and A3).

Data quality report

Every third month, a data quality report on potential quality deficiencies and noteworthy data deviations will be generated automatically after data extraction from the eCRF. We will only include data from sites with more than five included participants and for every variable, a minimum of five participants must be added to be presented in the report. This is done to ensure some degree of anonymisation and since data assessment is difficult in very small samples. Site IDs will be blinded in the report, by assigning the sites with a new report-specific name based on a seed generated by the absolute time of the computer. Quality deficiencies or noteworthy data deviations flagged during previous central data monitoring meetings and registered in the assessment log, will be flagged in the data quality report.

The central monitoring group will assess the data quality report during a central monitoring meeting.

Quality deficiencies

If a site is found to have a high number or proportion of the prespecified quality deficiencies, the trial manager will contact the relevant principal investigator for further assessment. The central monitoring group may also decide that the principal investigator should be contacted for further assessment, if only few (or a single) deficiencies have been registered. Contact despite only a few (or a single) quality deficiencies being registered could be 1) clarify if the data entry flagged as a quality deficiency is correct, and 2) discuss how such quality deficiencies can be minimised in the future.

Noteworthy data deviations

Every variable will be assessed by the central monitoring group via the following steps:

1. Is there a site with noteworthy data as per the previous definitions? [Yes] / [No]
2. If yes, which suspicion has been raised? [Describe]
3. Will any course of action be taken? [Yes] / [No]
4. If yes, the trial manager will contact the principal investigator and take the relevant course of action
5. Results of the course of action will be noted

Monitoring of quality deficiencies and noteworthy data deviations will start after the first 100 participants have been randomised.

Assessment log

Quality deficiencies and noteworthy data deviations, including the course of action taken (e.g., contact to principal investigator) and the summary of the course of action will be registered in the assessment log (template available in Appendix, Table A1 and A2).

Final data quality report

Once the data quality monitoring report and the planned actions registered in the assessment log have been approved by the central monitoring group, a conversion key will be sent to the trial manager for re-identification of sites, in order to contact relevant principal investigators. Once the trial manager has been in touch with all relevant investigators, the report and assessment log will be updated accordingly. A short version of the final report will be uploaded to safeboosc.eu.

Appendix

Figure A1. Example of data completion report

Site	Randomisations	End of monitoring (28 days from birth) % completed	Serious Adverse Reactions (28 days from birth) % completed	90 days follow-up (90 days from randomisation) % completed
CH01	7	100%	100%	100%
DK02	8	80%	100%	n/a
ES01	4	75%	33%	0%
ES02	15	87%	63%	83%
ES04	44	100%	100%	100%
PL05	1	0%	0%	n/a
CZ01	32	82%	56%	75%

Figure A2. Example of site-specific missing data report

Sites

Forms

	End of monitoring	SAEs	90 days follow-up
All (n=730)	78%	73%	66%
CE (N=53)	67%	54%	34%

Figure A3. Example of a participant-specific missing data overview from a site-specific report.

Participants

Forms

In total: 53 participants are included, of them

- 10 (19%) have completed data *(green fields)*
- 14 (26%) are waiting for data entry *(yellow fields)*
- 29 (55%) have missing data *(red fields)*

Participant ID	Missing forms	End of monitoring	SAEs	90 days follow-up
CE0001	1			
CE0002	0			
CE0003	0		N/A	
CE0004	2		N/A	
CE0005	0			
CE0006	0			

Table A1. Quality deficiencies on assessment log, template

Quality deficiency	Blinded site ID	Comment	Will any course of action be taken? (yes/no)	Unblinded site ID	Summary of course of action

Table A2. Noteworthy data deviations on assessment log, template.

Variable	Blinded site ID	Suspicion raised (suspected outlier, misunderstanding, or fabrication)	Will any course of action be taken? (yes/no)	Unblinded site ID	Summary of course of action