# BaBi Network

# Guidance Note:

# Advice on the Use of Identifiers and Data Management for BaBi Participant Records

Author: Alex Newsham

(Research IT Development Manager, Born in Bradford)

Table of Contents

[Identifiers 3](#_Toc113008725)

[Recruitment Identifier 3](#_Toc113008726)

[Suggested Format for the Recruitment Identifier 3](#_Toc113008727)

[Participant Identifier 4](#_Toc113008728)

[Suggested Format for the Participant Identifier 4](#_Toc113008729)

[Pregnancy Identifier 4](#_Toc113008730)

[Screening Identifier 4](#_Toc113008731)

[Suggested Format for the Screening Identifier 4](#_Toc113008732)

[A Word about Check digits 4](#_Toc113008733)

[Data Items to Support Data Linkage, Contact & Cohort Management 5](#_Toc113008734)

[Non-study personal identifiers and demographics 5](#_Toc113008735)

[Additional Data and Data Sources 6](#_Toc113008736)

[Concluding Remarks 6](#_Toc113008737)

# Identifiers

For BaBi we do not wish to be too prescriptive in advising you how to store or manage your data but we do recommend using certain types of identifier types and formats.

In the following examples, the use of a prepended (i.e. prefixed) letter with identifiers (e.g. R\*, PN\* and S\*) is optional, but may help to distinguish identifiers by type, especially when used in combination or conjunction with other identifiers (it also helps prevent against the removal of leading zero’s in some proprietary data formats, such as Excel; you could, of course, use a starting integer of >0 if you wish to avoid the use of prepended letters in ID formation).

We would recommend keeping the identifier as simple as possible, so that the identifier remains atomic and therefore does not encode any individual characteristics about the person or the BaBi centre within the identifier itself.

If needed at a later stage in order to distinguish identifiers from different BaBi centres, then BaBi centre-specific encoded information can be prepended to the identifiers at the time of sharing the data.

Here are our recommendations of the types and format of identifiers to use.

## Recruitment Identifier

The **Recruitment** **Identifier** uniquely identifies each instance of participant consent to the study. This is allocated each time a participant is recruited, with a participant being recruited anew for each pregnancy registered as a BaBi-recruited pregnancy at your centre. This means in effect, that over the lifetime of BaBi, a woman may acquire multiple BaBi recruitment IDs, i.e. for her first and subsequent pregnancies that are recruited to BaBi. A child is also allocated a Recruitment ID when their details are obtained following notification of their birth.

### Suggested Format for the Recruitment Identifier

R + **00000001** + a Luhn check Digit, with each successive recruitment identifier allocated incrementing by one digit.

For example, this should produce a set of incrementing identifiers with the check digit on the end. Note that the second digit in from the right is the incrementing digit and the last digit the Luhn check digit:

* R100000017
* R100000025
* R100000033
* R100000041
* R100000058

Typically, it is the Recruitment identifier that is provided as the Registration Number in the accruals lists that are submitted to the EDGE system.

At Born in Bradford, it is our practice to database identifiers of a fixed length with a Microsoft SQL Server data type of Char(). In the case of this example of a Recruitment Identifier, this identifier is stored as Char data type of ten characters, e.g. Char(10).

## Participant Identifier

The **Participant Identifier** that uniquely identifies a participant in the study and is, therefore, assigned only once to a participant, for the lifetime of the study.

For BaBi recruited mothers, this is assigned when they are first recruited to BaBi. In the case of the child(ren), born of the mother’s BaBi-recruited pregnancy, each child receives their BaBi participant identifier once we receive details of their birth, providing the mother has not withdrawn from the study and the child is a live birth.

### Suggested Format for the Participant Identifier

PN +00000001 + a Luhn check digit,with each successive recruitment identifier allocated incrementing by one digit.

As with the Recruitment Identifier, a set of incrementing Participant Identifiers should be produced with a check digit on the end, and we would recommend storing them as a fixed character data type (e.g. SQL Server data type Char(11), given that the identifier is a fixed length of 11 characters).

## Pregnancy Identifier

The **Pregnancy Identifier** is shared by a mother and her children born of a particular pregnancy. It serves to link a mother and her children and to uniquely identify a pregnancy. A pregnancy identifier may be routinely allocated by and available from your local maternity system, so obtaining the record from the maternity system would support linkage between a mother and her babies. In the absence of a Pregnancy Identifier, the BaBi data team can advise on how best to go about creating one.

Screening Identifier

The **Screening Identifier** is applied to all mothers approached, or, considered as eligible for recruitment to BaBi, which includes both those who are and who are not successfully recruited. This will help to identify all those who have been considered and /or approached for participation in BaBi, regardless of whether or not they were subsequently recruited.

### Suggested Format for the Screening Identifier

S +00000001 + a Luhn check digit,with each successive recruitment identifier allocated incrementing by one digit.

Like the Recruitment Identifier, we would recommend storing this as a fixed character data type. If using Microsoft SQL Server, the Char(10) data type would support the suggested format.

## A Word about Check digits

We have advised incorporating check digits in the IDs you generate for BaBi data storage. Check digits are routinely used for error detection on IDs. Familiar examples of IDs that contain check digits are NHS Numbers and credit card numbers. They are produced using algorithms, the Luhn algorithm being one such instance, and we use a GetLuhn() function to apply the Luhn algorithm and attach a check digit when generating our IDs. This then helps us to spot errors in data capture.

If you would like to learn more, here’s a wiki page URL with further information: <https://en.wikipedia.org/wiki/Luhn_algorithm>.

We can provide a working example of a Luhn check digit function, should you need one, and can advise on deployment.

# Data Items to Support Data Linkage, Contact & Cohort Management

## Non-study personal identifiers and demographics

We also collect some basic data from the hospital record, which is stored as part of a participant’s BaBi record, to support identification & profile management, as well as future data linkage and contact.

We have permission to obtain the following core data items for our BaBi participants from our hospital maternity system:

* NHS Number - supports health record linkage
* Hospital Number/PAS number – supports hospital data record linkage
* Date of Birth
* Time of Birth (child only)
* Gender
* Ethnicity
* Contact details (Address with postcode; telephone numbers; email address)
* Booking Date
* Date of Assessment (the date of the appointment at which the mother consented)
* Expected Date of Delivery
* Language spoken
* Native tongue
* Reading ability

Most of these variables will already be familiar to you and therefore probably require no further explanation. However, the spoken and reading language variables probably merit further description, especially given that these are more than likely to vary from one Trust system to another, depending on the maternity and/or PAS system used at your NHS Trust, which will impact on the type and scope of language-related variables you are able to derive.

The benefit of obtaining these variables, or their equivalent, is to be found in their ability to support communications planning and aid preparation for bridging language barriers. They should help in providing an indication of an individual’s spoken language ability (i.e. fluency) in English; their preferred/main spoken language if that language is not English; help you to identify whether an interpreter is required; and provide some idea of the individual’s reading ability in English, or, an indication of their preferred/main reading language if that language is not English.

At Bradford, we previously sourced these variables from our Medway Maternity system and, now source them from our Cerner Maternity system, which superseded Medway as the Trust’s maternity system as of March 2022. As you will note from the following summary of these variables, there is considerable variation between what the Medway and the Cerner systems provide in terms of scope and utility. From Medway Maternity we sourced several variables to facilitate knowledge of our eligible BaBi population’s language needs. The specific variables we sourced were:

* [Spoken language ability] – this indicates fluency in spoken English
* [Spoken language ability Native tongue] – this indicates an individual’s preferred main language where their main language is not English
* [Spoken language ability Interpreter needed] – this indicates whether or not an interpreter is required
* [Reading ability] – this indicates the individual’s reading ability in English
* [Reading ability Native tongue] – this indicates the individuals main reading language where that language is not English

From Cerner Maternity, we now source the following variables, which best approximate the variables we used to receive from Medway Maternity:

* [Spoken\_Language\_Ability\_Understands\_English] – indicates with simple ‘Yes/No’ categories as to whether an individual understands spoken English
* [Spoken\_Language\_Ability\_Preferred\_Language] - this indicates an individual’s preferred main language where that language is not English (this variable is currently appearing blank in the data we are receiving from Cerner)
* [Spoken\_Language\_Ability\_Interpreter\_Required] - this indicates whether or not an interpreter is required
* [Difficulties\_Reading\_Writing\_English] - indicates with simple ‘Yes/No’ categories whether or not an individual can read English

## Additional Data and Data Sources

In addition to this core data, we also have permission to obtain further information about the pregnancy and birth, such as gravida, parity, gestation & gestational length, outcome of birth, birth order, and birth weight.

We would also advise setting up a routine submission to the NHS Tracing Service to maintain over time the completeness and accuracy of the data you hold on a participant.

# Concluding Remarks

We hope this advice and the set of recommendations helps to provide a robust foundation for structuring and managing your Centre’s BaBi data.

However, if you require further advice, or clarity, please get in touch with the BaBi data team.