­­

A green and blue circle

Description automatically generated

Data Cleaning Guidance:

2025 NHS Maternity Survey

July 2025

Data Cleaning Guidance:

2025 NHS Maternity Survey

April 2025

Picker

Picker is a leading international health and social care charity. We carry out research to understand individuals’ needs and their experiences of care. We are here to:

* Influence policy and practice so that health and social care systems are always centred around people’s needs and preferences.
* Inspire the delivery of the highest quality care, developing tools and services which enable all experiences to be better understood.
* Empower those working in health and social care to improve experiences by effectively measuring, and acting upon, people’s feedback.

© Picker 2025

Published by and available from:

Picker Institute Europe

Suite 6, Fountain House,

1200 Parkway Court,

John Smith Drive,

Oxford OX4 2JY

Tel: 01865 208100

Email: [Info@PickerEurope.ac.uk](mailto:Info@PickerEurope.ac.uk)

Website: [picker.org](http://www.picker.org)

Registered Charity in England and Wales: 1081688

Registered Charity in Scotland: SC045048

Company Limited by Registered Guarantee No 3908160

Picker Institute Europe has UKAS accredited certification for ISO20252:2019 (GB08/74322) via SGS and ISO27001:2022 ((certificate number 23715) via Alcumus ISOQAR. We comply with Data Protection Laws including the General Data Protection Regulation, the Data Protection Act 2018 and the Market Research Society's (MRS) Code of Conduct.

Before using this document

Before you use this document, check that you have the latest version, as there might be some small amendments from time to time (the date of the last update is on the front page). In the unlikely event that there are any major changes, we will e-mail all trust contacts and contractors directly to inform them of the change.

This document is available from the 2025 NHS Maternity Survey website: <https://nhssurveys.org/surveys/survey/04-maternity/year/2025/>.

Questions and comments

If you have any questions or concerns regarding this document, please contact the Survey Coordination Centre (SCC) using the details provided at the top of this page.



**For in house trusts and contractors taking part in the survey:**

Contractors and in house trusts submitting final data for the Maternity Survey **must not** clean their data before submitting it to the Survey Coordination Centre (SCC). Please refer to the [Survey Handbook](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Survey%20handbook.docx) and [Entering and Submitting Final Data](http://www.nhssurveys.org/Filestore/Generic_instructions/Generic_Entering_submitting_data_V2.pdf) instructions for more details.

Contents

[1. Data cleaning – an overview 6](#_Toc204867494)

[Introduction 6](#_Toc204867495)

[Definitions 6](#_Toc204867496)

[Raw / uncleaned data: 6](#_Toc204867497)

[Free text comments: 7](#_Toc204867498)

[Data cleaning: 7](#_Toc204867499)

[Ask-all questions: 7](#_Toc204867500)

[Routing questions: 7](#_Toc204867501)

[Filtered questions: 7](#_Toc204867502)

[Multiple response questions: 7](#_Toc204867503)

[Multiple questionnaire responses: 7](#_Toc204867504)

[Sample data: 8](#_Toc204867505)

[Response data: 8](#_Toc204867506)

[Out-of-range data: 8](#_Toc204867507)

[Outcome: 8](#_Toc204867508)

[Outcome code to adjusted response rate mapping: 9](#_Toc204867509)

[Non-specific response: 9](#_Toc204867510)

[Missing responses: 9](#_Toc204867511)

[2. Entering and coding data prior to submission 11](#_Toc204867512)

[Multiple Questionnaire Responses - De-duplication and Inclusion 12](#_Toc204867513)

[Outcome code priorities 12](#_Toc204867514)

[Online partial responses 12](#_Toc204867515)

[3. Editing and cleaning data after submission 14](#_Toc204867516)

[Approach and rationale 14](#_Toc204867517)

[Cleaning multi-code questions – Incompatible answer codes 14](#_Toc204867518)

[Dealing with filtered questions 15](#_Toc204867519)

[Example 1: 16](#_Toc204867520)

[Example 2: 16](#_Toc204867521)

[Eligibility 17](#_Toc204867522)

[Age / Year of birth 17](#_Toc204867523)

[Demographics 17](#_Toc204867524)

[Out-of-range data 18](#_Toc204867525)

[Usability 18](#_Toc204867526)

[Example 19](#_Toc204867527)

[Conditions for usability 19](#_Toc204867528)

[Missing question responses 19](#_Toc204867529)

[Question suppressions 20](#_Toc204867530)

[Non-specific responses 20](#_Toc204867531)

[3. Weighting 21](#_Toc204867532)

[Weights used in National Patient Surveys 21](#_Toc204867533)

[Population weight (pop\_weight) 21](#_Toc204867534)

[Trust weights (qx\_tr\_weights) 21](#_Toc204867535)

[Public service agreement weight (psa\_weight) 22](#_Toc204867536)

[Appendix A: Example of cleaning 23](#_Toc204867537)

[Appendix B: In-range data 25](#_Toc204867538)

1. Data cleaning – an overview

Introduction

Once fieldwork for the 2025 NHS Maternity Survey (MAT25) has been completed, data needs to be submitted to the SCC in a **raw, uncleaned** format (for details of this, see the guidance on the NHS surveys website on [Entering and Submitting Final Data](https://nhssurveys.org/survey-instructions/entering-and-submitting-final-data/). To ensure that the cleaning process is comparable across all NHS trusts, the SCC cleans the full dataset of all trusts.

This document provides a description of the processes that will be used by the SCC to clean and standardise data submitted for the 2025 NHS Maternity Survey. By following the guidance contained in this document it should be possible for all data users to replicate this cleaning process on raw uncleaned data. These instructions focus on the selected answer codes, rather than the free text comments, which are reviewed separately to ensure confidentiality and identify safeguarding concerns. This document should be used alongside the 2025 NHS Maternity Survey [data mapping document](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Data%20mapping%20document.xlsx) which provides further information on specific and non-specific responses.



Please note the only data cleaning to be undertaken on the data file before it is submitted to the SCC is the de-duplication of cases and prioritisation of outcome codes where multiple questionnaires have been returned for a respondent. No further data cleaning should be applied to the raw data before it has been submitted.

Definitions

Definitions of terms commonly used in this document, as they apply to the 2025 NHS Maternity Survey are as follows:

Raw / uncleaned data:

‘Raw’ or ‘uncleaned’ data are data that have been entered verbatim from completed questionnaires without any editing taking place to remove contradictory or inappropriate[[1]](#footnote-1) responses; thus, all response boxes crossed on the questionnaire should be included in the [data entry spreadsheet](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Data%20Entry%20Spreadsheet.xlsx) (see the [Entering and Submitting Final Data document](http://nhssurveys.org/survey-instructions/entering-and-submitting-final-data/)). If two response options are selected on a single-choice question, the response should be left blank (see [Section 2](#Sec2) for more details).

While data are required in a raw/uncleaned format, data should still be checked for errors resulting from problems with data entry and should have duplicates removed. Ensuring high data quality is paramount and errors resulting from data entry problems can and should be corrected, as detailed in the Final [Data Entry Checklist](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Data%20Entry%20Checklist.xlsx).

Free text comments:

These are verbatim comments provided by a service user in response to the ‘Other comments’ section at the end of the survey. These responses should be included within the [data entry spreadsheet](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Data%20Entry%20Spreadsheet.xlsx). A service user may have only answered the free text comments and none of the quantitative questions. We would still want their free text comments to be provided to the SCC even though they have not answered the other questions in the survey, and regardless of outcome code.

Data cleaning:

The SCC uses the term ‘data cleaning’ to refer to all editing processes undertaken upon survey data once the survey has been completed and the data has been entered and collated.

Ask-all questions:

These are items in the questionnaire which are not subject to any filtering, and which should therefore be answered by all respondents. For the 2025 NHS Maternity Survey, the ask-all questions are **A1-C1, C9-C22, E1-F1, G1-H5, H7-H11** and **H13-H14**.

Routing questions:

These are items in the questionnaire which instruct respondents either to continue to the next question or to skip irrelevant questions, depending on their response to the routing question. For the 2025 NHS Maternity Survey, the routing questions in the questionnaire are **C1-C3, C22-C23, F1, F3, H5** and **H11**.

Filtered questions:

These are items in the questionnaire which are not intended to be answered by all respondents. Whether individual respondents are expected to answer filtered questions depends on their responses to preceding routing questions. For the 2025 NHS Maternity Survey, the filtered questions in the questionnaire are **C2-C8, C23, D1-D7, F2-F4, H6** and **H12**.

Multiple response questions:

These are items in the questionnaire where either multiple responses to a single item are permissible, or the question is treated this way for analysis purposes. For the 2025 NHS Maternity Survey, the multiple response questions are **B1, C5, C11, D6, H1, H5, H7** and **H11**.

Multiple questionnaire responses:

For the 2025 NHS Maternity Survey, respondents have the option to complete the survey either online or on paper. This may lead to one respondent completing both the online and paper questionnaire. Below is a description of how multiple questionnaire responses may occur:

All service users receive the first letter inviting them to complete the survey online by including a link and log in details, and a QR code. After 3 days from the first mailing (day 4), those who haven't completed the survey receive an SMS text reminder with a unique link to the online survey. Following this, a first postal reminder letter is sent on day 15 from the first mailing to those who haven’t responded. After 3 days (day 18 from the first mailing) non-respondents receive a second SMS text message reminder. On day 29 from the first mailing, those who haven’t completed the survey receive a second postal reminder letter which also includes a paper questionnaire to complete. This is followed by a final reminder letter on day 43 and a final SMS message on day 46.

Though the reminders are only sent to non-respondents, there is the potential for an individual to complete both the online and paper version. This situation may arise if there is an overlap between when the respondent completes the online survey and when the paper questionnaire arrives in the mail. A respondent having just completed the online survey may assume that their response was not recorded if they receive the paper questionnaire. Then they may fill out and return the paper questionnaire as well.

To address such cases of multiple responses from the same individual, [procedures are in place](#_Multiple_Questionnaire_Responses) during the coding of data prior to submission to identify and remove duplicate entries.

Sample data:

Service user data that is provided by the trust as part of the sampling process. This includes trust code, patient record number, postcode, mobile phone number indicator, mother’s year of birth, mother’s gender, mother’s ethnic group, time, day, month and year of delivery, number of babies born, actual delivery place, site code for place of birth and whether any babies born received neonatal care, including what type of care was received if they did.

Response data:

Data from the completed questionnaire which is provided from the service user. This includes answers to A1 through to H14.

Out-of-range data:

This refers to instances where data within a variable has a value that is not permissible. For categorical data – most of the variables in this survey – this would mean a value not allowed in the data, for example, a value of ‘3’ being entered for a variable with only two response categories (1 or 2). Out-of-range responses entered into the dataset should not be automatically (e.g., algorithmically) removed prior to submitting the data to the SCC. A full list of valid responses for the 2025 NHS Maternity Survey can be found in the [data mapping document](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Data%20mapping%20document.xlsx). In range **sample** data is listed in [Appendix B](#_Appendix_B:_In-range).

Outcome:

An outcome code is given to each service user to indicate the end result of their participation in the survey. This is used when calculating the adjusted response rate for the survey and is therefore vital to ensure all service users are coded appropriately. The coding for outcome is as follows:

* Outcome 1: Returned completed questionnaire (including accessible versions[[2]](#footnote-2))
* Outcome 2: Undelivered / moved house
* Outcome 3: Mother / baby deceased after fieldwork starting
* Outcome 4: Too ill / opted out / returned blank questionnaire
* Outcome 5: Ineligible
* Outcome 6: Unknown
* Outcome 7: Mother / baby died prior to fieldwork starting

Outcome code to adjusted response rate mapping:

Outcome codes are used in calculating the adjusted response rate. When an outcome code is marked as ‘Excluded’, this means it is not included in the base count. The Adjusted Response Rate (ARR) is calculated as:

where:

* Outcome 1: Returned usable
* Outcome 4: Did not respond - Opted out
* Outcome 5: Did not respond - Ineligible
* Outcome 6: Did not respond during fieldwork period

The following outcomes codes are not included in calculating the adjusted response rate:

* Outcome 2: Excluded – Returned undelivered
* Outcome 3: Excluded – Deceased after fieldwork starting
* Outcome 7: Excluded – Deceased prior to fieldwork starting

Non-specific response:

This refers to response options that essentially indicate the question is not directly applicable or relevant to the respondent. Most commonly, non-specific responses include responses such as “Don’t know / can’t remember”, and those indicating that the question is not applicable to the respondent such as “I did not want / need support” or “I did not contact a midwifery team”. Please note, non-specific responses are set to missing in the final respondent level dataset[[3]](#footnote-3). This does not delete the data in any way but alters how that data is used in analysis.

Missing responses:

This term is used to describe data which are not stored as a valid response for a question or variable in a dataset. There can be a number of different types of missing data, with the most common being missing responses for cases where outcome is 1.

Within the data cleaning process, several different missing response codes are used to identify how data for a particular respondent has been handled.

These codes are as follows:

* 999 ‘missing’: this code is used for out-of-range question responses but is primarily used when someone should have answered a question but did not. For example, ask-all questions or filtered questions where the respondent meets the filter criteria, or where the respondent has correctly skipped the filtered question.
* 998 ‘inapplicable’: this code is used when someone answered a question but should not have. For example, filtered questions where the respondent does not meet the filter criteria.
* 997 ‘incompatible’: this code is used when someone provided two incompatible responses to a multi-code question. It is also used if an out-of-range response has been provided for the year of birth question.

**Please note, contractors should submit raw ‘uncleaned’ data to the SCC, as per the ‘**[**Entering and coding data prior to submission**](#Entering)**’ section below.**

2. Entering and coding data prior to submission

For the 2025 survey, in house trusts and contractors are required to submit raw (‘uncleaned’) data to the SCC. For clarification, raw data is created as follows:

* All responses should be entered into the dataset, regardless of whether or not the respondent was meant to respond to the question (e.g., where service users answer questions that they have been directed to skip past, these responses should still be entered).
* Where a respondent has selected more than one response category on a question, the response should be left blank for that person in the data. The exceptions to this are for the ‘multiple response’ questions (e.g., B1), where respondents may select more than one response option. These should be coded ‘0’ response not selected and ‘1’ response selected[[4]](#footnote-4).
* Where a respondent has crossed out a response, this should not be entered in the data (the response should be left blank). Where a respondent has crossed out a response and instead selected a second response option, the second choice should be entered into the data.
* Where a respondent has given their response inconsistently with the formatting of the questionnaire but where their intended response is nonetheless unambiguous upon inspection of the completed questionnaire, then the respondent’s intended response should be entered. For example, where a respondent has written their date of birth underneath the boxes at H4 (“In what year were you born?”), then their year of birth should be entered.
* For the year of birth question, unrealistic responses should still be entered except following the rule above. For example, if a respondent enters ‘2025’ in the year of birth box, this should still be entered unless the respondent has unambiguously indicated their actual year of birth to the side.
* Once the data has been entered, no responses should be removed or changed in any way except where responses are known to have been entered incorrectly or where inspection of the questionnaire indicates that the service user’s intended response has not been captured. This includes ‘out-of-range’ responses, which must not be removed from the dataset. Responses in the dataset should only be changed before submission to the SCC where they are found to have been entered inconsistently with the respondent’s intended response.
* The data file should be de-duped. In practice, this means removing multiple questionnaire submissions, so the file only contains one record per service user. This may include applying the outcome code priority order detailed in the next section.
* Free text comments given in the final ‘Other comments’ section of the survey should be submitted in the [data entry spreadsheet](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Data%20Entry%20Spreadsheet.xlsx). Free text comments should be cleaned of any symbols and/or non-readable characters that may result from the survey software before submission.

Multiple Questionnaire Responses - De-duplication and Inclusion

This section outlines how to approach situations when a service user returns multiple questionnaires. The below table details how to approach different scenarios where this may occur.

Table 1. Selecting a questionnaire if multiple questionnaires are returned by a service user in the 2025 NHS Maternity Survey

|  |  |
| --- | --- |
| **Scenario** | **Priority** |
| The total number of completed questions should be calculated, and the questionnaire with the highest number of completed questions should be selected. | **1st** |
| In the event that the total number of completed questions is equal on all questionnaires, the data used are selected according to a priority order, and the earliest questionnaire received (either online or paper) should be selected. | **2nd** |
| In the rare event that the total number of completed questions is equal on all questionnaires, and the questionnaires were received at the same time, priority will be given to the response completed online. | **3rd** |

Outcome code priorities

There may be scenarios when a service user falls into multiple unproductive outcomes, for example a service user returns a completed questionnaire, but they are later flagged as deceased during fieldwork. In such cases, the following priority list should be used to determine which outcome should be kept and which outcome(s) should be removed.

Table 2. Selecting an outcome code if multiple questionnaires or outcome codes in the 2025 NHS Maternity Survey

|  |  |
| --- | --- |
| **Outcome Code** | **Priority** |
| Outcome 1: Returned completed questionnaire | **1st** |
| Outcome 7: Mother / baby died prior to fieldwork starting | **2nd** |
| Outcome 3: Mother / baby deceased after fieldwork starting | **3rd** |
| Outcome 5: Ineligible | **4th** |
| Outcome 4: Too ill / Opted out / returned blank questionnaire | **5th** |
| Outcome 2: Undelivered / moved house | **6th** |
| Outcome 6: Unknown | **7th** |

Online partial responses

For the online component of the survey, where a respondent has not clicked the submit button, responses are to be accepted and recorded as a complete based on the following:

* If a respondent has left a free text comment, the respondent is included in the data file;
* If a respondent has NOT left a free text comment and has selected a response option for H14 (ethnicity question), the respondent is included in the data file[[5]](#footnote-5);
* Any respondents who didn't reach the ethnicity question are NOT included in the data file.

This approach ensures that all responses where the respondent had progressed to this specified point will be captured and included in the dataset.

3. Editing and cleaning data after submission

Approach and rationale

The aim of the SCC in cleaning the collated final data is to ensure an optimal balance between data quality and completeness. Thus, we seek to remove responses that are known to be erroneous or inappropriate, but to do so in a relatively permissive way to enable as many responses as possible to contribute to the overall survey results.

Cleaning multi-code questions – Incompatible answer codes

Where participants have answered two incompatible codes in a multi-code question, these should be removed, as it is not possible for both those answers to be correct. For example, at B1 “Were you offered a choice about where to have your baby?” participants cannot select both "Yes – a choice of hospitals" and "No - I had no choices due to medical reasons".

Table 3. List of multi-code questions that include mutually exclusive response options and how to clean them

|  |  |
| --- | --- |
| **Condition for multi-code questions** | **Recode** |
| B1 any option of 1 to 4 = 1 AND B1 any option 5 to 7 = 1 | B1 = 997 |
| B1 any option of 5 to 6 = 1 AND B1 option 7 = 1 | B1 = 997 |
| C5 any option of 1 to 3 = 1 AND C5 any option 4 to 5 = 1 | C5 = 997 |
| C5 option 4 = 1 AND C5 option 5 = 1 | C5 = 997 |
| C11 any option of 1 to 4 = 1 AND C11 option 5 = 1 | C11 = 997 |
| D6 option 1 = 1 AND D6 any option 2 to 5 = 1 | D6 = 997 |
| D6 any option of 2 to 4 = 1 AND D6 option 5 = 1 | D6 = 997 |
| H1 any option of 1 to 6 = 1 AND H1 any of option 7 to 9 = 1 | H1 = 997 |
| H1 option 7 = 1 AND H1 option 8 to 9 = 1 | H1 = 997 |
| H5 any option of 1 to 18 = 1 AND H5 any option 19 to 20 = 1 | H5 = 997 |
| H5 option 19 = 1 AND H5 option 20 = 1 | H5 = 997 |
| H7 any option of 1 to 2 = 1 AND H7 any option 3 to 4 = 1 | H7 = 997 |
| H7 option 3 = 1 AND H7 option 4 = 1 | H7 = 997 |
| H11 any option of 1 to 5 = 1 and H11 option 6 = 1 | H11 = 997 |

Where participants have selected more than one answer code at a single code question, these responses should be recoded as 997.

Dealing with filtered questions

Some of the questions included in the survey are only relevant to a subset of respondents, and in these cases filter instructions are included in the questionnaire to route respondents past questions that are not applicable to them.

It is necessary to clean the data to recode responses where filter instructions have been incorrectly followed. In such cases, participants’ responses to questions that were not relevant to them are recoded in the dataset. Responses are only recoded where respondents have answered filtered questions despite selecting an earlier response on a routing question instructing them to skip these questions.

In such cases, participants’ responses to questions that were not relevant to them are recoded to ‘998’ to indicate a non-applicable response.[[6]](#footnote-6) See Table 4 for a list of all routing questions included in the 2025 NHS Maternity Survey. If any of the answer codes shown are selected at the routing question, then the answers at filtered questions should be recoded as 998.

Table 4. Appropriate cleaning for filtered questions in the 2025 NHS Maternity Survey

|  |  |
| --- | --- |
| **Condition for routing questions** | **Recoding for filtered questions (if answered)** |
| C1 options 1 or 2 | C2 = 998 |
| C2 options 2 or 3 | C3, C4, C5, C6, C7 and C8 = 998 |
| C3 options 2 or 3 | C4 and C5 = 998 |
| C22 option 2 | C23 = 998 |
| C23 option 2 | D1, D2, D3, D4, D5, D6 and D7 = 998 |
| F1 options 2 or 3 | F2, F3 and F4 = 998 |
| F3 options 4 or 5 | F4 = 998 |
| H5 options 19 or 20 = 1 | H6 = 998 |
| H11 option 6 = 1 | H12 = 998 |

A worked example of the cleaning process for recoding non-applicable responses to filtered questions is included in [Appendix A: example of cleaning](#_Appendix_A:_Example).

The recoding of filtered questions into ‘998’ only applies where respondents have answered filtered questions despite ticking an earlier response on a routing question instructing them to skip these questions.

Example 1:A screenshot of a questionnaire

Description automatically generated

In the example above (Example 1), the response to D1 would be recoded to ‘998’ because according to their answer from C23 (the routing question), they were supposed to skip to E1.

Responses to filtered questions are not removed where the response to the routing question is missing (Example 2).

Example 2:

A questionnaire with blue text

AI-generated content may be incorrect.

In the example above, H11 would be coded as missing (‘999’) and the response to H12 would remain as code 1.

Eligibility

Age / Year of birth

There may be instances where the sample and response data are mismatched, and the response data indicates that the respondent is under the age of 16. When this occurs, respondents will be considered eligible for the survey if their sample data is not missing and therefore remain as outcome 1. This approach aims to avoid removing legitimate responses because of an overly conservative approach to assessing eligibility; in other words, where the respondent’s age is uncertain (because sample and response information contradict each other and in different instances either of these may be accurate or inaccurate) the benefit of the doubt is given in any assessment of eligibility, so respondents are not excluded based on potentially inaccurate data. We cannot be certain whether the mismatch occurs due to an error in the sample file or an error in the service user’s completion of the questionnaire. It is also possible that there has been an error in data entry.

In instances where the sample data is missing, the response data is the only proof of age available. If the response data indicates the respondent is under the age of 16, the respondent will be considered ineligible (outcome 5). See Table 5.

Table 5. Eligibility and outcome codes of service users based on sample and response data of age

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Original outcome code** | **Sample data** | **Response data** | **Eligibility** | **Final outcome code** |
| 1 | YoB ≤ 2009 | H4 > 2009 | Eligible | 1 |
| 1 | YoB ≤ 2009 | H4 ≤ 2009 | Eligible | 1 |
| 1 | YoB ≤ 2009 | H4 = missing | Eligible | 1 |
| 1 | YoB = missing | H4 ≤ 2009 | Eligible | 1 |
| 1 | YoB = missing | H4 > 2009 | Ineligible | 5 |
| 1 | YoB = missing | H4 = missing | Ineligible | 5 |

Demographics

In a small number of cases, sample data and response data does not correspond for age. For example, the sample data may identify an individual as being born in 1980 only for the service user to report being born in 1985.

Where service user responses to demographic questions in the questionnaire are present, it is assumed these are more likely to be accurate than the sample data (since it is assumed that respondents are best placed to know their own age). However, because questions about demographics tend to produce relatively high item non-response rates, it is not appropriate to rely on response data alone. In cases of mismatch where it is obvious that the respondent data contains a clear and unambiguous error (e.g. when the respondent enters the current year instead of their year of birth), then sample data is used.

For demographic analysis on groups of cases, it is therefore necessary to use some combination of the information supplied in the sample data and response data. To do this, we first copy all valid responses to survey demographic questions into a new variable. Where response data is missing, we then copy in the relevant sample data[[7]](#footnote-7) (note that for a very small number of respondents demographic information may be missing in both the sample and response data; in such cases data must necessarily be left missing in the new variable)[[8]](#footnote-8).

Out-of-range data

A common error when completing year of birth questions is for respondents to accidentally write in the current year. In this case, the response to **H4** would be considered as an out-of-range response and would therefore be set to missing. For the 2025 NHS Maternity Survey, out-of-range responses for **H4** are recoded as ‘997’. The out-of-range responses for **H4** are defined as **H4 ≤ 1953 or H4 ≥ 2010**.This must only be done after the Age / Year of Birth eligibility cleaning as described in the earlier section titled ‘[Eligibility](#_Cleaning_special_cases)’

Out-of-range data must also be set for invalid responses to all other questions in the survey. The out-of-range responses will depend on the number of response options given for each question. For instance, all questions with three response options (e.g., A1, B15, C12) with response data of **≤ 0 or ≥ 4** would be set to missing and coded as ‘999’.

A list of in-range responses for the 2025 NHS Maternity Survey are listed in [Appendix B](#AppB).

Usability

Sometimes questionnaires (paper or online) are returned with only a very small number of questions completed. As in previous years and across the NHS Patient Survey Programme, for the 2025 NHS Maternity Survey, questionnaires containing responses to fewer than five questions are considered ‘unusable’ – we will set all responses (except free text) pertaining to such cases as system missing ‘left blank’ and recode the outcome to 6. Contractors and in-house trusts should submit these at outcome 1 and the SCC will re-code accordingly. This should only affect a very limited number of cases and so should not have a significant impact on response rates. The number of responses per questionnaire (including responses to the demographic questions) will be counted after all other cleaning has been conducted.

Additional clarification regarding what constitutes as five responses in determining if a questionnaire is usable (or not):

* Verbatim comments in the ‘Other comments’ free-text response box are not counted towards the five responses.
* Multiple choice questions are counted as one response, even where multiple response options are selected. For instance, H1 would be counted as one response in the below scenario.

Example

A screenshot of a questionnaire

AI-generated content may be incorrect.

Conditions for usability

**It is possible that a questionnaire could be considered usable because there are five or more responses, despite having an outcome code of 2, 3, 4 or 6.** In this case the outcome would be recoded to 1 to indicate a complete usable questionnaire.

Missing question responses

It is useful to be able to see the number of missing responses for each question. Responses are considered to be missing when a respondent is expected to answer a question, but no response is present.

For ask-all questions, responses are expected from all respondents – thus any instance of missing data constitutes a missing response.

For filtered questions, only respondents who have answered a previous routing question instructing them to go on to that filtered question or set of filtered questions are expected to give answers. Where respondents have missed a routing question, they are not expected to answer subsequent filtered questions; thus, only where respondents were explicitly instructed to answer filtered questions should such blank cells be coded as missing responses.

The SCC codes missing responses in the data as ‘999’[[9]](#footnote-9). For results to be consistent with those produced by the SCC, missing responses should be reported but excluded from the base number of respondents used to calculate percentages.

Question suppressions

The SCC will suppress results at both national and trust level for questions that have fewer than 30 respondents[[10]](#footnote-10). The suppression of 30 is achieved by weights at national level or base size at trust level. Note: non-specific responses are excluded from this count.

Non-specific responses

As well as excluding missing responses from results, non-specific responses are removed from the base numbers for percentages. The rationale for this is to facilitate easy comparison between institutions by presenting only results from those respondents who felt able to give an evaluative response to questions.

A list of non-specific responses is detailed in the [data mapping document](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Data%20mapping%20document.xlsx).

As shown in Table 6, using hypothetical data, non-specific response option 4 has been excluded from the base number when calculating percentages for question B2. This is because those selecting answer option 4 said they did not know or could not remember, so were not able to provide an evaluative response to the question. Therefore, any percentages used based on 2025 Maternity data would use the percentages in the column on the far right of Table 6, excluding the non-specific response options.

Table 6. Example of how percentages are calculated excluding non-specific response options with hypothetical data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **B2: Did you get enough information from either a midwife or doctor to help you decide where to have your baby?** | | | | |
| **Response options** | **Original**  **base**  **numbers** | **Percentage including non-specific response options** | **Base numbers for percentages** | **Percentage excluding non-specific response options** |
| 1. Yes, definitely | 6,000 | 58.5% | 6,000 | 60.0% |
| 1. Yes, to some extent | 2,000 | 19.5% | 2,000 | 20.0% |
| 1. No | 2,000 | 19.5% | 2,000 | 20.0% |
| 1. Don’t know / can’t remember | 250 | 2.4% | - | - |
| **Total base** | 10,250 | - | 10,000 | - |

3. Weighting

Weights used in National Patient Surveys

For some analysis purposes, case weights are applied to the data. The weighted result comprises the case-level data multiplied by the case-level weight, aggregated to the reporting level and then divided by the sum of case weights at the reporting level.

Population weight (pop\_weight)

The purpose of this weight is to adjust for differential non-response within trusts among demographic groups (for Maternity, by age group only) to ensure that the data is representative in terms of the demographic make-up. A single population weight is computed for each respondent.

The pop\_weight is calculated for each case as follows:

1. Use sample data for age to create three age groups (i.e. 16-26, 27-32, 33+).
2. Calculate the proportion of *total cases* within each trust that were sampled in January and February and fit into each of the three age groups (the population proportion).
3. Weight the respondent data by the weight and calculate the *weighted* proportion of respondent cases within each trust who fit into each of the three age groups (the weighted respondent proportion).
4. Create a pop\_weight for each case by dividing the population proportion (step 2) by the respondent proportion (step 3).
5. To contain the impact of extreme outliers we truncate weights to a maximum of 5, meaning any weights larger than 5 will be replaced by the value of 5, thus reducing the impact of any very large weights. Note that the outliers are checked and truncated *before* the weight is added to ensure the truncating focusses only on those outliers resulting from differential non-response.

Missing weights are set to one to avoid data loss.

The population data are taken from the selected cases in February (the core sample month), with data from small trusts included from January – as per previous years. This provides a representative snapshot of the overall population that is not affected by the boosted sample and is comparable with the population used to weight the data in previous years.

The pop\_weight is applied to categorical data in national tables analysis (with the exception of demographic questions), when it is multiplied by the trust weight (tr\_weight) for the question.

Trust weights (qx\_tr\_weights)

This weight has been used since the beginning of the survey programme. Its purpose is to adjust for the differing numbers of respondents between trusts for any one question so that each trust has equal influence on the national table figures.

One qx\_tr\_weight is calculated per trust per question (with the exception of demographic questions). For a given question and trust, the qx\_tr\_weight is the average number of specific responses across all trusts divided by the number of specific responses in that trust.

The weight is applied to categorical data in national tables analysis when it is multiplied by the pop\_weight.

Public service agreement weight (psa\_weight)

This originates with the Department of Health public service agreement for which the national statistics indicators were developed. It has also been applied in trust benchmarking from the time this was undertaken by CQC and subsequently transferred to Picker. Its purpose is to standardise the scores between trusts to create a more level ‘playing field’ when comparing trusts.

One psa\_weight is calculated per respondent equal to the proportion in that person’s weighting group in the national achieved sample divided by the proportion of that person’s weighting group in the trust’s achieved sample.

Weights are based on age group (age3\_group) and parity for this survey and are capped to a maximum of 5.

The weight is applied to scored data in trust benchmarking. Trust weights are not applied for benchmarking but are implicit because trust scores are averaged to produce the national mean scores. Psa\_weight is also used to weight the NHS National statistic output.

Trusts with 30 responses (including both specific and non-specific responses) or less are suppressed. These responses are not weighted.

Appendix A: Example of cleaning

Figure 1 shows hypothetical raw / uncleaned data for eight service users, five of whom have responded to the survey. It can be seen from this data that some of the respondents have followed instructions from routing questions incorrectly:

Respondents ‘003’ and ‘005’ have answered the question about in- triage (F4) even though they said they did not attend triage face-to-face or didn’t know / couldn’t remember if they had been through face-to-face triage in F3, meaning they should have skipped F4 and gone straight to G1. Respondents ‘002’, ‘007’ and ‘008’ have followed the routing correctly.

Figure 1. Example of raw / uncleaned data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Record** | **Outcome** | **F3** | **F4** | **G1** |
| Patient Record Number | Outcome of sending questionnaire (N) | Thinking about the last time you attended triage face-to-face, did the midwife or doctor you spoke to listen to you? | Thinking about the last time you attended triage in person, how did you feel about the length of time you waited before you were seen by a midwife? | Thinking about your postnatal care, were you involved in decisions about your care? |
| E001XXX | 6 |  |  |  |
| E002XXX | 1 | 1 | 1 | 2 |
| E003XXX | 1 | 4 | 3 | 4 |
| E004XXX | 4 |  |  |  |
| E005XXX | 1 | 5 | 2 | 3 |
| E006XXX | 6 |  |  |  |
| E007XXX | 1 | 1 | 3 | 3 |
| E008XXX | 1 | 4 |  | 3 |

Following the cleaning instructions above, the SCC will remove these inapplicable responses. Firstly, the filter instructions specify that:

|  |  |  |
| --- | --- | --- |
| **Routing question** | **Response values requiring cleaning** | **Filtered questions to be recoded** |
| **F3** | 4 or 5 | **F4** |

In accordance with this, if the respondent has answered **F3 = 4 or 5** (i.e., did not go through triage face-to-face or didn’t know / couldn’t remember if they had been listened to during face-to-face triage), then responses for **F4** must be recoded as ‘998’ (not applicable). If the respondent has correctly answered **F3 = 4 or 5** (i.e., did not go through triage face-to-face or didn’t know / couldn’t remember if they had been listened to during face-to-face triage), and then correctly skipped **F4**, then **F4** should be set to ‘999’.

Figure 2 below shows how the data would look after cleaning is done by the SCC to remove responses to filtered questions that should have been skipped.

Figure 2. Example of cleaned data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Record** | **Outcome** | **F3** | **F4** | **G1** |
| Patient Record Number | Outcome of sending questionnaire (N) | Thinking about the last time you attended triage face-to-face, did the midwife or doctor you spoke to listen to you? | Thinking about the last time you attended triage in person, how did you feel about the length of time you waited before you were seen by a midwife? | Thinking about your postnatal care, were you involved in decisions about your care? |
| E001XXX | 6 |  |  |  |
| E002XXX | 1 | 1 | 1 | 2 |
| E003XXX | 1 | 4 | 998 | 4 |
| E004XXX | 4 |  |  |  |
| E005XXX | 1 | 5 | 998 | 3 |
| E006XXX | 6 |  |  |  |
| E007XXX | 1 | 1 | 3 | 3 |
| E008XXX | 1 | 4 | 999 | 3 |

Appendix B: In-range data

The [data mapping document](https://nhssurveys.org/wp-content/surveys/04-maternity/03-instructions-guidance/2025/MAT25_Data%20mapping%20document.xlsx) published for 2025 NHS Maternity Survey indicates the in-range values for each survey question. The below table (Table 7) covers the in-range data for sample information, or any information completed during fieldwork.

Table 7. In-range sample data

|  |  |
| --- | --- |
| **Sample Variable** | **In-range data** |
| Trust code | Three-character code of the organisation e.g. RTH. |
| Patient Record Number (PRN) | The unique serial number allocated to each mother by the trust. This should include the survey code (E), followed by a four-digit number (e.g. 0001, 0002, …), followed by the trust code (e.g. RTH). E.g. E0001RTH. |
| Postcode | The mother’s full postcode |
| Mobile number indicator | 0,1 |
| Mother’s Year of Birth | ≥ 1954 ≤ 2009 |
| Mother’s Gender | 1, 2, 9 |
| Mother’s Ethnic Category | A, B, C, D, E, F, G, H, J, K, L, M, N, P, R, S, Z |
| Time of Delivery | 24-hour clock without seconds, e.g. 14:18 |
| Day of Delivery | If Month of Delivery = 2: ≥1 or ≤ 28;  If Month of Delivery = 1: ≥1 or ≤ 31 |
| Month of Delivery | 1 or 2 |
| Year of Delivery | 2025 |
| Number of Babies Born at Delivery | ≥ 1 ≤ 5 |
| Actual Delivery Place | 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 |
| Place of Birth: NHS Site Code | Five-character NHS trust site code |
| Baby 1 Received Neonatal Care | ≥ 1 or ≤ 8 |
| Baby 2 Received Neonatal Care | ≥ 1 or ≤ 8 or left blank |
| Baby 3 Received Neonatal Care | ≥ 1 or ≤ 8 or left blank |
| Baby 4 Received Neonatal Care | ≥ 1 or ≤ 8 or left blank |
| Baby 5 Received Neonatal Care | ≥ 1 or ≤ 8 or left blank |

A green and blue circle

Description automatically generated

Picker Institute Europe  
Suite 6, Fountain House,

1200 Parkway Court,

John Smith Drive,

Oxford OX4 2JY

Tel: +44 (0) 1865 208100

info@pickereurope.ac.uk

picker.org

Charity registered in England and Wales: 1081688

Charity registered in Scotland: SC045048

Company limited by guarantee registered in England and Wales: 3908160

1. ‘Inappropriate’ responses refer to where a respondent has incorrectly answered a filtered or routed question. [↑](#footnote-ref-1)
2. This includes completed easy read, braille and large print questionnaires. The cut-off date is the end of fieldwork (15th July 2025). Submissions received on the cut-off date are to be included. Additionally, the date (day, month, year) of the questionnaire being received should be attributed to the related record prior to submitting the final dataset to the SCC. All outcome 1s should include the date of the questionnaire being received. [↑](#footnote-ref-2)
3. Please note that this rule doesn’t apply to multiple response questions, where all responses are included in the final respondent level dataset. [↑](#footnote-ref-3)
4. Please note that code 997 was used in previous iterations where more than 1 response option was selected to a single-response question but is no longer used for this purpose. In this case, the response should be left blank during data entry and would subsequently be coded 999. [↑](#footnote-ref-4)
5. The ethnicity question (H14) precedes the free text comments question. [↑](#footnote-ref-5)
6. Code ‘998’ is an arbitrary value chosen because it is out-of-range for all other questions on the survey. [↑](#footnote-ref-6)
7. The sample data have already been checked for eligibility upon submission. [↑](#footnote-ref-7)
8. The exception to this is when response rates are calculated. Because response rates vary between demographic groups, using response and sample data to calculate response rates would create a systematic source of bias in that we are only able to amend information for the *respondents*. Therefore, only the sample data should be used to calculate response rates by demographic groups. [↑](#footnote-ref-8)
9. This is an arbitrary value chosen because it is ‘out-of-range’ for all other questions on the survey. [↑](#footnote-ref-9)
10. This does not include the demographic items included in the ‘About you’ section of the questionnaire. [↑](#footnote-ref-10)