

Data Cleaning Guidance:

2025 Community Mental Health Survey

October 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

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 email all trust contacts and contractors directly to inform them of the change.

This document is available from the 2025 NHS Community Mental Health Survey website: <https://nhssurveys.org/surveys/survey/05-community-mental-health/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 Community Mental Health 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/05-community-mental-health/03-instructions-guidance/2025/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](#_Toc208919327)

 [Introduction………………………………………………………………………………………………...6](#_Toc208919328)

 [Definitions………………………………………………………………………………………………….6](#_Toc208919329)

 [Raw / uncleaned data.……………… …………………………………………………….……..…....6](#_Toc208919330)

 [Free text comments.…………………………………………………………………………….…..….6](#_Toc208919331)

 [Data cleaning.…………………………………………………………………………………….….….7](#_Toc208919332)

 [Ask-all questions.…………………………………………………………………………………….....7](#_Toc208919333)

 [Routing questions.……………………………………………………………………………….……...7](#_Toc208919334)

 [Filtered questions.…………………………………………………………………………………….…7](#_Toc208919335)

 [Multiple response questions.……………………………………………………………………….…..7](#_Toc208919336)

 [Multiple questionnaire responses.………………………………………………………………….….7](#_Toc208919337)

 [Sample data.………………………………………………………………………………………….….8](#_Toc208919338)

 [Response data.……………………………………………………………………………………….…8](#_Toc208919339)

 [Out-of-range data.………………………………………………………………………………….…...8](#_Toc208919340)

 [Outcome.…………………………………………………………………………………………….…..8](#_Toc208919341)

 [Outcome code to adjusted response rate mapping.………………………………………………...9](#_Toc208919342)

 [Non-specific response.…………………………………………………………………………………9](#_Toc208919343)

 [Missing responses.……………………………………………………………………………………..9](#_Toc208919344)

[2. Entering and coding data prior to submission……………………………………………………….11](#_Toc208919345)

 [Multiple questionnaire responses – de-duplication and inclusion…………………………………12](#_Toc208919346)

 [Outcome code priorities………………………………………………………………………………..12](#_Toc208919347)

 [Online partial responses……………………………………………………………………………...12](#_Toc208919348)

[3. Editing and cleaning data after submission………………………………………………………….14](#_Toc208919349)

 [Approach and rationale………………………………………………………………………………...14](#_Toc208919350)

 [Cleaning multi-code questions – incompatible answer codes……………………………………..14](#_Toc208919351)

 Cleaning single code questions – incompatible answer codes……………………………...…….14

 [Dealing with filtered questions………………………………………………………………………...14](#_Toc208919352)

 [Example 1.……………………………………………………………………………………………..16](#_Toc208919353)

 [Example 2.……………………………………………………………………………………………..17](#_Toc208919354)

 [Recoding correctly skipped questions………………………………………………………………..18](#_Toc208919355)

 [Example 3.……………………………………………………………………………………………..18](#_Toc208919356)

 [Cleaning of Q17……………………………………………………………………………….………..19](#_Toc208919357)

 [Eligibility………………………………………………………………………………………………….19](#_Toc208919358)

 [Age / Year of birth ……………………………………………………………………….…………….19](#_Toc208919359)

 [Q1 / Contact with NHS mental health services…………………………………………….……....20](#_Toc208919360)

 [Demographics…………………………………………………………………………………….…..…20](#_Toc208919361)

 [Out-of-range data………………………………………………………………….……………………21](#_Toc208919362)

 [Usability………………………………………………………………………………………………….21](#_Toc208919363)

 [Example 4.………………………………………………………………………………………….….22](#_Toc208919364)

 [Conditions for usability………………………………………………………………………………..22](#_Toc208919365)

 [Missing question responses…………………………………………………………………….……..23](#_Toc208919366)

 [Question suppressions………………………………………………………………………….….…..23](#_Toc208919367)

 [Non-specific responses……………………………………………………………………….………..23](#_Toc208919368)

[4. Weighting………………………………………………………………………………………….……..25](#_Toc208919369)

 [Weights used in National Patient Surveys……………………………………………………..…….25](#_Toc208919370)

 [Population weight (pop\_weight)……………………………………………………………….…….25](#_Toc208919371)

 [Trust weights (tr\_weights)……………………………………………………………………….…...25](#_Toc208919372)

 [Public service agreement weight (psa\_weight)…………………………………………………....26](#_Toc208919373)

[Assessment Service Group (ASG) trust\_asg\_psa\_weight……………………………………….…...26](#_Toc208919374)

[Appendix A: Example of cleaning………………………………………………………….…………….27](#_Toc208919375)

[Appendix B: In-range data………………………………………………………………………….…….29](#_Toc208919376)

# 1. Data cleaning – an overview

## Introduction

Once fieldwork for the 2025 Community Mental Health Survey (CMH25) 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](http://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 Community Mental Health 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 [CMH25 data mapping document](https://nhssurveys.org/surveys/survey/05-community-mental-health/year/2025/) 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 Community Mental Health 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 responses; thus, all response boxes crossed on the questionnaire should be included in the data entry spreadsheet (see the [Entering and Submitting Final Data document](http://nhssurveys.org/survey-instructions/entering-and-submitting-final-data/)).

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/surveys/survey/05-community-mental-health/year/2025/).

### Free-text comments

These are verbatim comments provided by a service user in response to the three open questions at the end of the survey: “Was there anything particularly good about your care?”; “Was there anything that could be improved?”; and “Any other comments?”. These responses should be included within the [data entry spreadsheet](https://nhssurveys.org/wp-content/surveys/05-community-mental-health/03-instructions-guidance/2025/Date%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. In these instances, records should be coded as ‘Outcome 1’. If the questionnaires contain responses to fewer than five questions, they will be considered ‘unusable’ and recoded as ‘Outcome 6’ during data cleaning, as per the [Usability](#_Usability) section.

### 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 (except those who ticked response option 4 at Q1). For the CMH25 survey, the ask all questions are **Q1, Q2, Q8-Q20, Q23, Q26, Q31-Q35, Q38-Q42, Q44-Q50.**

### Routing questions

These are items in the questionnaire that instruct respondents either to continue to the next question or to skip irrelevant questions, depending on their response to the routing question. For the CMH25 Survey, the routing questions are **Q1, Q2, Q6, Online-Q8, Q20, Q23, Q26, Q27, Q35, Q42.**

### 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, or on whether questions are targeted at a specific group (e.g. respondents aged 16-25 years responding to Online Q8 in the online survey). For the CMH25 Survey, the filtered questions are **Q3-Q7, Online-Q8-Online-Q12, Q21, Q22, Q24, Q25, Q27-Q30, Q36, Q37, Q43.**

### 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 CMH25 Survey, the multiple response questions are **Q27, Q36** and **Q42.**

### Multiple questionnaire responses

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

All eligible service users receive the first letter inviting them to complete the survey online by including a link and login details, and QR code. After 5 working days from the first mailing (day 7), 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, which includes a paper questionnaire to complete, is sent on day 14 from the first mailing to those who haven’t responded. After 5 working days (day 21 from the first mailing) non-respondents receive a second SMS text message reminder. On day 28 from the first mailing, those who haven’t completed the survey receive a second postal reminder letter which includes another paper questionnaire.

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 provided by the trust as part of the sampling process. This includes: trust code; service user record number; postcode; mobile phone number indicator; year of birth; gender; ethnic category; day, month and year of last contact; sub-ICB code; mental health inpatient indicator; service or team type; assessment service group; mode of contact.

### Response data

Data from the completed questionnaire which was provided from the respondent(s). This includes answers to Q1 through Q50.

### 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 Community Mental Health Survey can be found in the [data mapping document](https://nhssurveys.org/surveys/survey/05-community-mental-health/year/2025/). In range **sample data** is listed in [Appendix B](#AppB).

### 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 it 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[[1]](#footnote-2))
* Outcome 2: Undelivered / moved house
* Outcome 3: Service user died after fieldwork starting
* Outcome 4: Too ill / opted out / returned blank questionnaire
* Outcome 5: Ineligible
* Outcome 6: Unknown
* Outcome 7: Service user 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:

$$ARR=\frac{Outcome 1}{Outcome 1+Outcome 4+Outcome 5+Outcome 6}$$

* 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, these are responses such as “Don’t know / can’t remember”. Likewise, responses that indicate the question is not applicable to the respondent are considered ‘non-specific’ – for example, responses such as “I do not need support for this”. Please note that non-specific responses are set to user missing in the final respondent-level dataset[[2]](#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: this code is used for out-of-range question responses and 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.
* 998: 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: 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.
* 995: this code is used when someone correctly skipped a question, by following the routing indicated in the questionnaire.
* 98: This code is used for respondents who have selected more than one answer for the scale of 0 to 10 overall experience question (Q38), and it cannot be determined which answer they intended to select.

**Please note, contractors should submit raw ‘uncleaned’ data to the SCC, as per the ‘**[**Entering and coding data prior to submission**](#_2._Entering_and)**’ 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. Q36), where respondents may select more than one response option. These should be coded ‘0’ for response not selected and ‘1’ for response selected[[3]](#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 Q44 (“What was your year of birth?”), 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.
* 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 three questions in the survey (“Was there anything particularly good about your care?”; “Was there anything that could be improved?”; and “Any other comments?”) should be submitted in the [data entry spreadsheet](https://nhssurveys.org/wp-content/surveys/05-community-mental-health/03-instructions-guidance/2025/Date%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 CMH25 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 outcome codes, 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 CMH25 Survey

|  |  |
| --- | --- |
| **Outcome Code** | **Priority** |
| Outcome 1: Returned completed questionnaire | **1st** |
| Outcome 7: Service user died prior to fieldwork starting  | **2nd** |
| Outcome 3: Service user 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 Community Mental Health Survey, partial responses will not be accepted for the online component of the survey. Service users completing the survey online will be required to press the ‘submit’ button at the end of the survey for their data to be included. This approach ensures that full consent has been given by service users in order for their responses to be included in the dataset.

Service users do not need to complete the recontact question on the online survey to have a completed response.

# 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 answers to be correct. For example, at Q27 “In the last 12 months, have you contacted NHS mental health crisis support?” participants cannot select both “Yes, I contacted NHS 111 mental health option” and “No, I have not contacted NHS crisis care".

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

|  |  |
| --- | --- |
| **Condition for multi-code questions** | **Recode** |
| Q27 any option of 1 to 3 = 1 AND Q27 any option 4 to 6 = 1 | Q27 = 997 |
| Q27 any option of 4 to 5 = 1 AND Q27 option 6 = 1 | Q27 = 997 |
| Q42 any option of 1 to 16 = 1 AND Q42 any option 17 to 18 = 1 | Q42 = 997 |
| Q42 option 17 = 1 AND Q42 option 18 = 1 | Q42 = 997 |

## Cleaning single code questions – incompatible answer codes

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.[[4]](#footnote-5) See table 4 for a list of all routing questions included in the CMH25. 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 CMH25

|  |  |
| --- | --- |
| **Condition for routing questions** | **Recoding for filtered questions (if answered)** |
| Q1 option 4  | See [Eligibility](#_Q1_/_Contact). |
| Q2 options 3, 4, 5, 6, or 7 | Q3 – Q7 = 998 |
| Q6 options 2 or 3 | Q7 = 998 |
| Online Q8 options 3 or 4 | Online Q9 – Online Q12 = 998 |
| Q20 option 5 | Q21 – Q22 = 998 |
| Q23 options 2, 3, 4, or 5 | Q24 and Q25 = 998 |
| Q26 options 2 or 3 | Q27 – Q30 = 998 |
| Q27 options 4, 5 or 6 | Q28 – Q30 = 998 |
| Q35 options 2 or 3 | Q36 and Q37 = 998 |
| Q42 options 17 or 18 | Q43 = 998 |

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

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



In the example above (Example 1), the response to Q28 would be recoded to ‘998’ because according to their answer from Q27 (the routing question), they were supposed to skip Q28.

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

### Example 2



In the example above (Example 2), Q35 would be coded as missing (‘999’) and the response to Q36 would remain as code 4.

## Recoding correctly skipped questions

Where participants correctly followed the filtered questions instructions, and skipped the question(s) that are incompatible with the routing, this is recoded as ‘995[[5]](#footnote-6)’. This code indicates that the response is consistent with the intended survey logic.

### Example 3



In the example above (Example 3), the response to Q7 would be recoded as ‘995’ because the respondent correctly skipped answering to this question.

## Cleaning of Q17

Question 17 (“In the last 12 months, have you had a care review meeting with your NHS mental health team to discuss how your care is working?”) requires additional cleaning due to the nature of the question, to ensure that the results of the analysis are as fair as possible when assessing trust performance.

As Q17 specifies a time period of ‘the last 12 months’, if Q2(“Overall, how long have you been in contact with NHS mental health services?”) has been given a response option of 1 (‘Less than 1 year’ of contact with mental health services) then Q17 should be coded as 998. This is because it is not fair to penalise a mental health trust for not having reviewed a person’s care, if the person has not been in contact with these services long enough to have reasonably expected them to have had a meeting to discuss how their care is working.

Table 5. Additional cleaning for Q17

|  |  |
| --- | --- |
| **Condition for Q17** | **Recode** |
| Q2 option 1 = 1 | Q17 = 998 |

## 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 6.

Table 6. 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 | Q44 > 2009 | Eligible | 1 |
| 1 | YoB ≤ 2009 | Q44 ≤ 2009 | Eligible  | 1 |
| 1 | YoB ≤ 2009 | Q44 = missing | Eligible | 1 |
| 1 | YoB = missing | Q44 ≤ 2009 | Eligible | 1 |
| 1 | YoB = missing | Q44 > 2009 | Ineligible | 5 |
| 1 | YoB = missing | Q44 = missing | Ineligible | 5 |

### Q1 / Contact with NHS mental health services

A respondent who has marked response option 4 for Q1[[6]](#footnote-7) (stating they ‘have never seen anyone from NHS mental health services’) is recoded from outcome 1 to outcome 5. This is because the respondent is not eligible to take part in the survey if they have never seen anyone from NHS mental health services. For all ineligible respondents (outcome 5), any responses to Q1 through Q50 are set to system missing ‘999’.

## Demographics

In a small number of cases, sample data and response data does not correspond for age and gender of service users. For example, the sample may identify a service user as male only for them to report being female, or 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, gender, and other characteristics). 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 for age and gender. 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-8) (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 as system missing in the new variable) [[8]](#footnote-9).

## 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 Q44 would be considered as an out-of-range response. For the CMH25, out-of-range responses for Q44 are recoded as ‘997’. The out-of-range responses for Q44 are defined as Q44 ≤ 1919 or Q44 > 2009**.** This must only be done after eligibility has been set as described in the earlier section titled [‘Eligibility’](#_Cleaning_special_cases).

Out-of-range data must be set for invalid responses to all 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., Q6, Q13, Q16, Q17, Q26, Q34, Q35, Q39, Q40, Q43, Q46) with response data of **≤ 0 or ≥ 4** would be set to missing and coded as ‘999’.

A list of in-range responses for the sample data in CMH25 are listed in [Appendix B](#AppB).

## Usability

Sometimes paper and online questionnaires are returned with only a very small number of questions completed. As in previous years and across the NHS Patient Survey Programme, for the CMH25, paper and online questionnaires containing responses to fewer than five questions are considered ‘unusable’ – we will set all possible responses pertaining to such cases as system missing ‘left blank’ and recode the outcome to 6. 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 (the final three questions in the survey “Was there anything particularly good about your care?”; “Was there anything that could be improved?”; and “Any other comments?”) are not counted towards the five responses.
* Multiple choice questions are counted as one response, even where multiple response options are selected. For instance, Q42 would be counted as one response in the below scenario (example 4).
* Online recontact question is not counted towards the five responses.

###

### Example 4



### 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.

For multiple response questions, if a respondent has not selected at least one of the response options, resulting in answers being coded as ‘0’ or left blank, they should be coded as missing responses.

The SCC codes missing responses in the data as ‘999’[[9]](#footnote-10). For results to be consistent with those produced by the SCC, missing responses should be presented but should not be included in the base number of respondents for percentages.

## Question suppressions

The SCC will suppress results at both national and trust level for questions that have fewer than 30 respondents[[10]](#footnote-11). 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. Question suppressions are applied directly in the reporting outputs, and the data is not deleted from the SPSS dataset.

## Non-specific responses

As well as excluding missing responses from results, the SCC also removes non-specific responses from 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/surveys/survey/05-community-mental-health/year/2025/).

As shown in table 7, using hypothetical data, non-specific response option 4 has been excluded from the base number when calculating percentages for question Q9. 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, in reporting data would use the percentages in the column on the far right of table 7, excluding the non-specific response options.

Table 7. Example of how percentages are calculated excluding non-specific response options with hypothetical data[[11]](#footnote-12)

|  |
| --- |
| **Q9: Did you feel your NHS mental health team listened to what you had to say?** |
| **Response options** | **Original** **base** **numbers** | **Percentage including non-specific response options** | **Base numbers for percentages** | **Percentage excluding non-specific response options** |
| 1. Yes, always
 | 6,000 | 59.5% | 6,000 | 60.0% |
| 1. Yes, sometimes
 | 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 | - |

# Weighting

## Weights used in NHS Patient Survey Programme

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 the Community Mental Health Survey, by age and gender) 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 from age (i.e. 16-35; 36-50; 51-65; 66+) and gender to create eight groups (i.e. male 16-35; male 36-50 etc.)
2. Calculate the proportion of *total cases* within each trust that were sampled in April – May and fit into each of the eight groups (the population proportion)
3. Calculate the proportion of *respondent cases* (outcome=1) within each trust who fit into each of the eight groups (the 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 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 (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 tr\_weight is calculated per trust per question (with the exception of demographic questions). For a given question and trust, the 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, where 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 capped to a maximum of 5 and are based on age group and gender for this survey.

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.

### Assessment Service Group (ASG) trust\_asg\_psa\_weight

A trust\_asg\_psa\_weight will also be calculated for each assessment service group (CAMHS, Adults, Older People). This weight will be calculated as per trust level psa\_weight detailed above. Three separate trust\_asg\_psa\_weights will be produced, on for each assessment service group.

These three weights will standardise the scores for each assessment service group between trusts to create a more level ‘playing field’ when comparing trusts within the assessment service group. The weight is applied to scored data in trust benchmarking.

Weights are capped to a maximum of 5 and are based on age group and gender for this survey.

# 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:

Respondent ‘MH25RYG0002’ has reported that they have not received any therapies in the last 12 months (Q23=2) and respondent ‘MH25RYG0005’ has reported that they cannot remember if they have received any NHS therapies (Q23=5), but they have both responded to filtered questions Q24 and Q25.

Figure 1. Example of raw / uncleaned data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Record** | **Outcome** | **Q23** | **Q24** | **Q25** |
| Service User Record Number | Outcome of sending questionnaire (N)  | In the last 12 months, have you received any therapies for your mental health needs? | How do you feel about the length of time you waited between your assessment with the NHS mental health team and your first therapy appointment? | Thinking about the last time you received therapy, did you have enough privacy to talk comfortably? |
| MH25RYG0001 | 6 |  |  |  |
| MH25RYG0002 | 1 | 2 | 1 | 2 |
| MH25RYG0003 | 1 | 1 | 2 | 4 |
| MH25RYG0004 | 4 |  |  |  |
| MH25RYG0005 | 1 | 5 | 3 | 3 |
| MH25RYG0006 | 6 |  |  |  |
| MH25RYG0007 | 1 | 1 | 2 | 2 |
| MH25RYG0008 | 1 | 1 | 3 | 2 |

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

|  |  |
| --- | --- |
| **Condition for routing question** | **Recoding for filtered questions (if answered)** |
| Q23 options 2, 3, 4, or 5 | Q24 and Q25 = 998  |

In accordance with this, where the respondent has answered **Q23 = 2, 3, 4 or 5** (i.e., had not received any therapies) all responses for **Q24** and **Q25** must be recoded as ‘998’.

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** | **Q23** | **Q24** | **Q25** |
| Service User Record Number | Outcome of sending questionnaire (N)  | In the last 12 months, have you received any therapies for your mental health needs? | How do you feel about the length of time you waited between your assessment with the NHS mental health team and your first therapy appointment? | Thinking about the last time you received therapy, did you have enough privacy to talk comfortably? |
| MH25RYG0001 | 6 |  |  |  |
| MH25RYG0002 | 1 | 2 | 998 | 998 |
| MH25RYG0003 | 1 | 1 | 2 | 4 |
| MH25RYG0004 | 4 |  |  |  |
| MH25RYG0005 | 1 | 5 | 998 | 998 |
| MH25RYG0006 | 6 |  |  |  |
| MH25RYG0007 | 1 | 1 | 2 | 2 |
| MH25RYG0008 | 1 | 1 | 3 | 2 |

# Appendix B: In-range data

The [data mapping document](https://nhssurveys.org/surveys/survey/05-community-mental-health/year/2024/) published for 2025 Community Mental Health Survey indicates the in-range values for each survey question. Table 8 below covers the in-range data for sample information, or any information completed during fieldwork.

Table 8. In-range sample data

|  |  |
| --- | --- |
| **Sample Variable** | **In-range data** |
| Year of Birth | ≥ 1919 or ≤ 2009 |
| Gender | 0, 1, 2, 8 or 9  |
| Ethnic category | Anything except I, O, Q, T-Y |
| Day of last contact | ≥ 1≤ 31  |
| Month of last contact | ≥ 4 ≤ 8 |
| Year of last contact | 2025 |
| Sub-ICB code | See full list of Sub ICB Locations by clicking [here](https://digital.nhs.uk/binaries/content/assets/website-assets/services/ods/integrated-care-boards/ods-change-summary-icb-22-23---yr1-renaming). |
| Mental Health Inpatient Indicator | ≥ 0≤ 1 |
| Service or Team Type | Please see full list of service and teams by clicking [here](https://www.datadictionary.nhs.uk/attributes/service_or_team_type_for_mental_health.html). |
| Assessment Service Group | ≥ 1≤ 3 |
| Mobile phone indicator | ≥ 0≤ 1 |
| Mode of Contact | ≥ 1≤ 5 |
| Day Questionnaire Received | ≥1≤ 31 |
| Month Questionnaire Received  | ≥ 8≤12 |
| Year Questionnaire Received | 2025 |
| Outcome Code | ≥1≤ 7 |

1. This includes completed easy read, braille and large print questionnaires. [↑](#footnote-ref-2)
2. 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)
3. If a respondent does not select at least one response option, then the question should be left blank during data entry. [↑](#footnote-ref-4)
4. Code ‘998’ is an arbitrary value chosen because it is out-of-range for all other questions on the survey. [↑](#footnote-ref-5)
5. Code ‘995’ is an arbitrary value chosen because it is out-of-range for all other questions on the survey. [↑](#footnote-ref-6)
6. Q1 is a routing question which instructs respondents to go to the next applicable question. If a respondent ticks response option 4, the next applicable question is Q41 because Q2 through Q40 are in regard to their mental health care however they have indicated that they have never seen anyone from NHS mental health services. Due to these service users also being recoded as ineligible (outcome 5), all response data is set to system missing including Q41 through Q50. [↑](#footnote-ref-7)
7. The sample data have already been checked for eligibility upon submission. [↑](#footnote-ref-8)
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-9)
9. This is an arbitrary value chosen because it is ‘out-of-range’ for all other questions on the survey. [↑](#footnote-ref-10)
10. This does not include the demographic items included in the ‘About You’ section of the questionnaire. [↑](#footnote-ref-11)
11. Table 7 presents CMH25 example data. [↑](#footnote-ref-12)