

# DATA CLEANING GUIDANCE

## COMMUNITY MENTAL HEALTH SURVEY 2022

## Contacts

The Survey Coordination Centre for Existing Methods for the NHS Patient Survey Programme  
Picker Institute Europe  
Suite 6,  
Fountain House,  
1200 Parkway Court,  
John Smith Drive,  
Oxford,  
OX4 2JY

Tel: 01865 208127  
Fax: 01865 208101  
E-mail: [mentalhealth@surveycoordination.com](mailto:mentalhealth@surveycoordination.com)  
Website: [www.nhssurveys.org](http://www.nhssurveys.org)  
Updates

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 [NHS Surveys website](http://www.nhssurveys.org).

## Questions and comments

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



### For trusts and contractors taking part in the survey:

Contractors submitting final data for the NHS Community Mental Survey **must not** clean their data before submitting it to the Survey Coordination Centre for Existing Methods. Please refer to [Survey Handbook](#) and [Entering and Submitting Final Data](#) instructions for more details.

## Contents

<b>DATA CLEANING – AN OVERVIEW .....</b>	<b>5</b>
Introduction .....	5
Definitions .....	5
Raw / uncleaned data: .....	5
Data cleaning: .....	6
Ask-all questions:.....	6
Filtered questions:.....	6
Routing questions: .....	<b>Error! Bookmark not defined.</b>
Multiple response questions:.....	6
Multiple questionnaire responses:.....	7
Sample data:.....	7
Response data:.....	7
Out-of-range data:.....	7
Outcome: .....	7
Non-specific response:.....	7
Missing responses: .....	8
<b>ENTERING AND CODING DATA PRIOR TO SUBMISSION .....</b>	<b>8</b>
<b>EDITING AND CLEANING DATA AFTER SUBMISSION .....</b>	<b>10</b>
Approach and rationale .....	10
Dealing with filtered questions.....	10
Example 1:.....	12
Example 2:.....	12
Multiple Questionnaire Responses - De-duplication and Inclusion ...	<b>Error! Bookmark not defined.</b>

N:\Service Delivery\All Live SCC Projects\MH 2022 P101678\Instruction manuals\Data Cleaning  
Guidance\P101678\_CMH22\_Data Cleaning  
Guidance\_V3.0\_CLEAN\_HB\_PSS\_CK\_150622\_PROTECT.docx  
©2022 Care Quality Commission. All Rights Reserved.  
Survey Coordination Centre for Existing Methods

Tel: + 44 (0) 1865 208127

[mentalhealth@surveycoordination.com](mailto:mentalhealth@surveycoordination.com)

[www.nhssurveys.org](http://www.nhssurveys.org)

Outcome code priorities .....	<b>Error! Bookmark not defined.</b>
Cleaning special cases .....	13
Cleaning of Q12 and Q13 .....	13
Cleaning of Q18 and Q19 .....	13
Cleaning of Q40, Q41 and Q42 .....	13
Eligibility .....	14
Age / Year of birth .....	14
Q1 / Contact with NHS mental health services .....	15
Demographics .....	15
Out-of-range data .....	166
Usability .....	166
Example .....	177
Missing responses .....	177
Non-specific responses .....	188
<b>APPENDIX A: EXAMPLE OF CLEANING .....</b>	<b>19</b>
<b>APPENDIX B: OUT-OF-RANGE DATA .....</b>	<b>210</b>

# Data cleaning – an overview

## Introduction

Once fieldwork for the Community Mental Health Survey 2022 has been completed, contractors are required to submit final data to the SCCEM in a raw / uncleaned format. The SCCEM will clean the data once all contractors have submitted their files. To ensure that the cleaning process is comparable across all NHS trusts, data for all trusts in the survey are collated and cleaning is carried out on the full collated dataset.

This document provides a description of the processes that will be used by the SCCEM to clean and standardise data submitted by contractors as part of the Community Mental Health Survey 2022. By following the guidance contained in this document it should be possible to recreate this cleaning process. This document should be used alongside the [2022 Community Mental Health data mapping document](#) 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 SCCEM 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 Community Mental Health Survey 2022 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](#)). The requirement for raw / uncleaned data does not, however, preclude the checking of data for errors resulting from problems with data entry or de-duping the data file. Ensuring high data quality is paramount and errors resulting from data entry problems can and should be corrected by checking against the appropriate completed questionnaire.

### 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: 'Is there anything particularly good about your care?'; 'Is there anything that could be improved?'; and 'Any other comments?'. These responses should be provided in an excel file containing the SURN, outcome code and free text response. This file should be separate to the quantitative data file and should include free text comments for all respondents, regardless of their outcome code. For example, 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 SCCEM even though they have not answered the other questions in the survey.

### Data cleaning:

The SCCEM 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 Community Mental Health Survey 2022, the ask-all questions are **Q1-Q4, Q7-Q9, Q10, Q15, Q18, Q20, Q23, Q28, Q33-Q39, Q40 and Q43-Q49.**

### Routing questions:

These are items in the questionnaire which instruct respondents either to continue on to the next question or to skip irrelevant questions, depending on their response to the routing question. For the Community Mental Health Survey 2022, the routing questions in the questionnaire are **Q1, Q4, Q10, Q15, Q18, Q20, Q21, Q23, Q26, Q28 and Q40.**<sup>1</sup>

### 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 Community Mental Health Survey 2022, the filtered questions in the questionnaire are **Q5-6, Q11-Q14, Q16-Q17, Q19, Q21-Q22, Q24-Q27, Q29-Q32 and Q41-Q42.**

### 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 Community Mental Health Survey 2022, the multiple response questions are **Q5, Q11, Q37, Q41.**

- A code of '98' is used at Q37 if respondents have circled more than one answer at Q37 and it cannot be determined which number they wanted to circle.
- Q11 is treated as a multiple response question, although it is not presented as such in the questionnaire.

---

<sup>1</sup> Q1 is considered a routing question. However, it is cleaned differently from the other routing questions due to the nature of the question. If option 4 ('I have never seen anyone from NHS mental health services') is selected on Q1 then the respondent has indicated, they are ineligible to participate. Therefore, in addition to removing all subsequent responses their outcome code is recoded to '5' (ineligible).

### Multiple questionnaire responses:

Two paper copies of the questionnaire (an initial mailing and a final reminder) are posted to an individual service user, there is therefore the potential for an individual respondent/service user to return multiple questionnaires.

### Sample data:

Service user data that is provided from the trust as part of the sampling process. This includes postcode, gender, year of birth, ethnicity, date of last contact, Care Programme Approach (CPA) status, Clinical Commissioning Group (CCG) code, Care Cluster, mobile phone indicator, email address indicator and mode of contact.

### Response data:

Data from the completed questionnaire which is provided from the service user. This includes answers to Q1 through to Q49.

### 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 in 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 SCCEM. A full list of such responses for the 2022 Community Mental Health Survey can be found in [Appendix B: Out-of-range data](#).

### 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
- Outcome 2: Undelivered / moved house
- Outcome 3: Deceased after fieldwork started
- Outcome 4: Too ill / opt out
- Outcome 5: Ineligible
- Outcome 6: Unknown
- Outcome 7: Deceased prior to fieldwork
- Outcome 8: Easy Read completed questionnaire
- Outcome 9: Braille completed questionnaire
- Outcome 10: Large Print completed questionnaire

### Non-specific response:

This is a generic term for response options that can be considered as not being applicable to the respondent in terms of directly answering the specific question to which they are linked. 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' or 'I did not want to be involved in making decisions'. Please note: non-specific responses are set to user missing in the final respondent level dataset. 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 classed as 'user missing' data. 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 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.
- 996: this code is used to suppress data at trust level when a question has fewer than 30 responses<sup>1</sup>. These responses would also remain suppressed from the overall base at a national level.

## Entering and coding data prior to submission

For the 2022 survey, contractors are required to submit raw ('uncleaned') data to the SCCEM. 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, this question should be set to 'missing' for that person in the data (i.e., left blank, or coded as a full stop (.)). The exceptions to this are for the 'multiple response' questions (e.g., Q41), where respondents may select more than one response option and for Q37 where a code of '98' is used for respondents who have circled more than one answer as it cannot be determined which number they wanted to circle.
- Where a respondent has crossed out a response, this should not be entered in the data (the response should be left blank or coded as a full stop (.)). Where a respondent has

---

<sup>1</sup> Non-specific responses (i.e., "Don't know / can't remember") are excluded from the count.



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. For example, if a respondent enters ‘2022’ 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 automatically removed from the dataset. Responses in the dataset should only be changed before submission to the SCCEM 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 to the final three questions in the survey (‘Is there anything particularly good about your care?’; ‘is there anything that could be improved?’; and ‘any other comments?’) should be submitted as a separate file to the quantitative questions. This file should include all free text comments, regardless of whether a respondent was outcome=1 or not. The file should be in excel and include the SURN, outcome code and free text responses for that case.

## 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 Community Mental Health Survey 2022*

Scenario	Priority
The total number of completed questions should be calculated, and the questionnaire with the highest number of completed questions should be selected.	<b>First</b>
In the event that the total number of completed questions is equal on all questionnaires, the earliest questionnaire received should be selected.	<b>Second</b>

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, the first questionnaire that was scanned/data entered should be selected.	<b>Third</b>
---	--------------

## Outcome code priorities

As service users are sent up to two copies of the questionnaire throughout the fieldwork period, there may be duplicate questionnaires returned, or the service user may fall into multiple unproductive outcomes. The following priority list to remove duplicate outcome codes, should be used:

Table 2. Selecting an outcome code if multiple questionnaires or outcome codes in the Community Mental Health Survey 2022

Outcome Code	Priority
Outcome 1: Returned completed questionnaire	1 <sup>st</sup> =
Outcome 8: Easy Read completed questionnaire	1 <sup>st</sup> =
Outcome 9: Braille completed questionnaire	1 <sup>st</sup> =
Outcome 10: Large Print completed questionnaire	1 <sup>st</sup> =
Outcome 7: Deceased prior to fieldwork	2 <sup>nd</sup>
Outcome 3: Deceased during fieldwork	3 <sup>rd</sup>
Outcome 5: Ineligible	4 <sup>th</sup>
Outcome 4: Opted out	5 <sup>th</sup>
Outcome 2: Undelivered / moved house	6 <sup>th</sup>
Outcome 6: Unknown	7 <sup>th</sup>

# Editing and cleaning data after submission

## Approach and rationale

The aim of the SCCEM 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.

## 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.<sup>1</sup> See table 3 for a list of all routing questions included in the Community Mental Health Survey 2022, the response values that require cleaning, and the appropriate filtered questions to recode as '998'.

Table 3. Appropriate cleaning for routing questions in the Community Mental Health Survey 2022

Routing question	Response values requiring cleaning	Filtered questions to be recoded
Q1	4	See <a href="#">Eligibility</a> .
Q4	2 or 3	Q5 – Q6
Q10	2 or 3	Q11 – Q14
Q15	3 or 4	Q16 – Q17
Q18	2 or 3	Q19
Q20	2 or 3	Q21 – Q22
Q21	4, 5 or 6	Q22
Q23	2	Q24 – Q27
Q26	2 or 3	Q27
Q28	2, 3, 4 or 5	Q29 – Q32
Q40	2	See <a href="#">Cleaning special cases</a> .

Please note that these instructions should be followed in the order shown above.

A worked example of the cleaning process for recoding non-applicable responses to filtered questions is included in

<sup>1</sup> Code '998' is an arbitrary value chosen because it is out-of-range for all other questions on the survey.

## Appendix A: Example of cleaning.

Responses are only recoded where respondents have answered filtered questions despite ticking an earlier response on a routing question instructing them to skip these questions.

### Example 1:

**23. In the last 12 months, have you been receiving any **medicines** for your mental health needs?**

1  Yes → Go to 24  
2  No → Go to 28

**24. Has the **purpose** of your medicines ever been discussed with you?**

1  Yes, definitely  
2  Yes, to some extent  
3  No  
4  Don't know / can't remember

In the example above (example 1), the response to Q24 would be recoded to '998' because according to their answer from Q23 (the routing question), they were supposed to skip Q24. Responses to filtered questions are not removed where the response to the routing question is missing.

### Example 2:

**15. Have you and someone from **NHS mental health services** decided what care you will receive? (This may be called a care plan).**

1  Yes, definitely → Go to 16  
2  Yes, to some extent → Go to 16  
3  No → Go to 18  
4  Don't know / can't remember → Go to 18

**16. Were you involved as much as you wanted to be in deciding what care you will receive?**

1  Yes, definitely  
2  Yes, to some extent  
3  No, but I wanted to be  
4  No, but I did not want to be  
5  Don't know / can't remember

In the example above (example 2), the response to Q16 would remain as code 1 and Q15 would be coded as missing (999).

## Cleaning special cases

Some questions require 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.

### Cleaning of Q12 and Q13

**Q11** Is the **main** person in charge of organising your care and services...

**Q12** How well does this person organise the care and services you need?

**Q13** Do you know how to contact this person if you have a concern about your care?

If **Q11** has been given a response option of 1 ('A GP' is the main person in charge of organising their care and services) then **Q12** and **Q13** are recoded as '998' for not applicable. This is because it is not fair to assess the mental health trust if the main person in charge of organising the service user's care and services is not someone from the mental health trust.

### Cleaning of Q18 and Q19

**Q2** Overall, how long have you been in contact with NHS mental health services?

**Q18** In the last 12 months, have you had a care review meeting with someone from NHS mental health services to discuss how your care is working?

**Q19** Did you feel that decisions were made **together** by you and the person you saw during this discussion? (This includes contact in person, via video call and telephone).

As Q18 specifies a time period of 'the last 12 months', if **Q2** has been given a response option of 1 ('Less than 1 year' of contact with mental health services) then **Q18** and **Q19** are recoded as '998' as they are not applicable. 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 for long enough to have reasonably expected them to have had a meeting to discuss how their care is working.

### Cleaning of Q40, Q41 and Q42

**Q40** Do you have any physical or mental health conditions, disabilities or illnesses that have lasted or are expected to last for 12 months or more? Include problems related to old age.

**Q41** Do you have any of the following? Select **ALL** conditions you have that have lasted or are expected to last for 12 months or more.

**Q42** Do any of these reduce your ability to carry out day-to-day activities?

Although Q40 is a routing question and Q41 and Q42 are the corresponding filtered questions, these three questions are cleaned differently due to the nature of the questions.

When a respondent has answered 'No' to **Q40** but has answered **Q41** by selecting one or more long term condition(s), their response to **Q40** is cleaned. This is because their response to **Q41** indicates that they do in fact have a long-standing condition and therefore their response to **Q40** is likely to be incorrect.

However, when a respondent has answered 'No' to **Q40** and has not selected any long-term conditions at **Q41** but has answered **Q42** then their response to **Q42** is cleaned. As they have indicated that they do not have a long-term condition and have also skipped Q41, this would suggest that their response to Q42 is incorrect. See table 4 for a summary of how Q40, Q41 and Q42 are cleaned.

Table 4. Cleaning for Q40, Q41 and Q42 in the Community Mental Health Survey 2022

Q40 response	Q41 response	Q42 response	Cleaning
Ticked option 2 (no)	Ticked one or more options	No response	Q40 is set to missing (999) and as there is no response to Q42, this would also be set to missing (999)
Ticked option 2 (no)	Ticked one or more options	Ticked any option	Q40 is set to missing (999)
Ticked option 2 (no)	No response	Ticked any option or no response	Q41 is set to not applicable (998)  Q42 is set to not applicable (998)

## Eligibility

### Age / Year of birth

There may be instances where the sample and response data is mismatched and the response data indicates that the respondent is under the age of 18. When this occurs, respondents will *not* be considered ineligible for the survey if their sample data is not missing and therefore remain as outcome 1. This is 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. 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 18, the respondent will be considered ineligible (outcome 5). See table 4.

Table 4. 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 ≤ 2004	Q44 > 2004	Eligible	1
1	YoB ≤ 2004	Q44 ≤ 2004	Eligible	1
1	YoB ≤ 2004	Q44 = missing	Eligible	1
1	YoB ≤ 2004	Q44 = out-of-range <sup>1</sup>	Eligible	1
1	YoB = missing	Q44 > 2004	Ineligible	5
1	YoB = missing	Q44 = missing	Ineligible	5
1	YoB = missing	Q44 = out-of-range	Ineligible	5

### Q1 / Contact with NHS mental health services

A respondent who has marked response option 4 for Q1<sup>2</sup> (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 Q49 are set to system missing.

## Demographics

In a small number of cases, sample data and response data does not correspond for age and gender. 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 gender and 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.

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,

---

<sup>1</sup> Out-of-range responses for Q44 are defined as Q44 < 1900 or Q44 > 2005 (see 'Out-of-range data' for more details)

<sup>2</sup> 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 Q40 because Q2 through Q39 is 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 missing including Q40 through Q49.

we then copy in the relevant sample data (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)<sup>1</sup>.

## 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 and would therefore be set to missing (999). For the Community Mental Health Survey 2022, out-of-range responses for **Q44** are defined as **Q44 ≤ 1900 or Q44 ≥ 2005**. This must only be done after eligibility has been set as described in the earlier section titled 'Eligibility'.

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<sup>2</sup> (i.e., Q4, Q10, Q13, Q18, Q20, Q26, Q27, Q38, Q39 and Q42) with response data of **≤ 0 or ≥ 4** would be set to missing. A list of out-of-range responses for the Community Mental Health Survey 2022 that are not detailed in other documents are listed in [Appendix B: Out-of-range data](#).

## Usability

Sometimes questionnaires are returned with only a very small number of questions completed. For the Community Mental Health Survey 2022, questionnaires containing fewer than five responses are considered 'unusable' – we will set all responses pertaining to such cases as system missing 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 cleaning has been conducted.

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

- Verbatim comments in other comments are not counted towards the five responses.
- Multiple choice questions are counted once. For instance, Q41 would be counted as one response in the below scenario:

---

<sup>1</sup> 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.

<sup>2</sup> With the exception of Q5 which is a multiple response question and Q11 which is treated as a multiple response question.



## Example

**41. Do you have any of the following?**

Select **ALL** conditions you have that have lasted or are expected to last for 12 months or more.

- 1  Autism or autism spectrum condition
- 2  Breathing problem, such as asthma
- 3  Blindness or partial sight
- 4  Cancer in the last 5 years
- 5  Dementia or Alzheimer's disease
- 6  Deafness or hearing loss
- 7  Diabetes
- 8  Heart problem, such as angina
- 9  Joint problem, such as arthritis
- 10  Kidney or liver disease
- 11  Learning disability
- 12  Mental health condition
- 13  Neurological condition
- 14  Stroke (that affects your day-to-day life)
- 15  Another long-term condition

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, 6 or 7. In this case the outcome would be recoded to 1 to indicate a complete usable questionnaire.

## Missing responses

It is useful to be able to see the number of missed 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 SCCEM codes missing responses in the data as '999'<sup>1</sup>. For results to be consistent with those produced by the SCCEM, missing responses should be presented but should not be included in the base number of respondents for percentages.

---

<sup>1</sup> This is an arbitrary value chosen because it is 'out-of-range' for all other questions on the survey.

The SCCEM will suppress results at both national and trust level for questions that have fewer than 30 respondents<sup>12</sup> and code as '996'. Note: non-specific responses are excluded from this count.

## Non-specific responses

As well as excluding missing responses from results, the SCCEM 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.

---

<sup>1</sup> This does not include the demographic items included in the 'about you' section of the questionnaire.

<sup>2</sup> Non-specific responses (i.e., "Don't know / can't remember") are excluded from the count.

# 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 '002' has reported that they have not received any NHS talking therapies in the last 12 months (Q28=2) and respondent '005' has reported that they cannot remember if they have received any NHS talking therapies (Q28=5), but they have both responded to filtered questions Q29, Q30, Q31 and Q32.

Record	Outcome	Q28	Q29	Q30	Q31	Q32
Service User Record Number	Outcome of sending questionnaire (N)	In the last 12 months, have you received any NHS talking therapies for your mental health needs that do not involve medicines?	Were these NHS talking therapies explained to you in a way you could understand?	Were you involved as much as you wanted to be in deciding what NHS talking therapies to use?	Do you feel your NHS talking therapies have helped your mental health?	Overall, how did you feel about the length of time you waited before receiving talking NHS therapies?
001	6					
002	1	2	1	2	2	2
003	1	1	2	4	1	1
004	4					
005	1	5	3	3	1	3
006	6					
007	1	1	2	2	2	3
008	1	1	3	2		

Figure 1. Example of raw / uncleaned data

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

Routing question	Response values requiring cleaning	Filtered questions to be recoded
Q28	2, 3, 4 or 5	Q29 – Q32

In accordance with this, all responses for **Q29, Q30, Q31** and **Q32** must be set to not applicable in cases where the respondent has answered **Q28 = 2, 3, 4 or 5** (i.e., had not received NHS talking therapies).

Figure 2 below shows how the data would look after cleaning is done by the SCCEM to remove responses to filtered questions that should have been skipped – cells where responses have been set to not applicable are shaded.

Record	Outcome	Q28	Q29	Q30	Q31	Q32
Service User Record Number	Outcome of sending questionnaire (N)	In the last 12 months, have you received any NHS talking therapies for your mental health needs that do not involve medicines?	Were these NHS talking therapies explained to you in a way you could understand?	Were you involved as much as you wanted to be in deciding what NHS talking therapies to use?	Do you feel your NHS talking therapies have helped your mental health?	Overall, how did you feel about the length of time you waited before receiving NHS talking therapies?
001	6					
002	1	2	998	998	998	998
003	1	1	2	4	1	1
004	4					
005	1	5	998	998	998	998
006	6					
007	1	1	2	2	2	3
008	1	1	3	2	999	999

Figure 2. Example of cleaned data

## Appendix B: Out-of-range data

The [data mapping document](#) published for 2022 Community Mental Health Survey indicates the out-of-range values for each survey question. The below table (table 5) covers the out-of-range data for sample information, or any information completed during fieldwork.

Table 5. Out-of-range sample data

Sample Variable	Out-of-range data
Birth	≤ 1900 or ≥ 2005
Gender	≤ 0 3-8 ≥ 10
Ethnic category	I, O, Q, T-W
Day of last contact	≤ 0 ≥ 32
Month of last contact	≤ 0 5-8 ≥ 13
Year of last contact	≤ 2020 ≥ 2023
Registered on CPA	≤ 0 ≥ 4
Care Cluster Code	9 ≥ 22-98 ≥ 100
Mobile phone indicator	2-8 ≥ 9
Email address indicator	2-8 ≥ 9
Mode of Contact	≤ 0 ≥ 6
Day Questionnaire Received	≤ 0 ≥ 32
Month Questionnaire Received	≤ 1 ≥ 7
Year Questionnaire Received	≤ 2021 > 2022
Outcome Code	≤ 0 ≥ 11