

# DATA CLEANING GUIDANCE

## 2022 URGENT AND EMERGENCY CARE SURVEY

## Contacts

The Survey Coordination Centre for Existing Methods  
Picker  
Suite 6, Fountain House  
1200 Parkway Court  
John Smith Drive  
Oxford  
OX4 2JY

Tel: 01865 208127  
Fax: 01865 208101  
E-mail: [emergency@surveycoordination.com](mailto:emergency@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 very 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](#).

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

Contractors and in house trusts submitting final data for the Urgent and Emergency Care Survey **must not** clean their data before submitting it to the Survey Coordination Centre for Existing Methods. Please refer to the [survey handbook](#) and [Entering and Submitting Final Data](#) instructions for more details.

# Contents

<b>DATA CLEANING – AN OVERVIEW .....</b>	<b>4</b>
Introduction .....	4
Definitions .....	4
Raw / uncleaned data.....	4
Free text comments:.....	5
Data cleaning .....	5
Ask-all questions: .....	5
Routing questions.....	5
Filtered questions .....	5
Multiple response questions .....	6
Sample data .....	6
Response data .....	6
Out-of-range data .....	6
Outcome.....	6
Non-specific responses .....	7
Missing responses.....	7
<b>ENTERING AND CODING DATA PRIOR TO SUBMISSION .....</b>	<b>7</b>
Multiple questionnaire responses - De-duplication and inclusion .....	8
Outcome code priorities.....	9
<b>EDITING AND CLEANING DATA AFTER SUBMISSION .....</b>	<b>9</b>
Approach and rationale .....	9
Dealing with filtered questions .....	10
Cleaning Special Cases .....	12
Cleaning of the long-term condition questions .....	12
Dealing with multiple response questions .....	14
Incompatible response options for multiple response questions .....	14
Eligibility .....	14
Age / Year of birth .....	14
Demographics .....	15
Out-of-range data .....	16
Usability.....	17
Example .....	17
Missing responses.....	18
Non-specific responses .....	18
<b>APPENDIX A: EXAMPLE OF CLEANING .....</b>	<b>19</b>
<b>APPENDIX B: OUT-OF-RANGE SAMPLE DATA – TYPE 1 AND TYPE 3.....</b>	<b>21</b>

# Data Cleaning – An overview

## Introduction

At the end of fieldwork for the 2022 Urgent and Emergency Care Survey participating in house trusts and contractors are required to submit data to the SCCEM in a raw / uncleaned format<sup>1</sup>. The SCCEM will clean the data once all in house trusts and 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 by the SCCEM.

This guidance document will aid understanding of how the data submitted by contractors is cleaned and standardised so that it is ready for the reporting undertaken by the SCCEM and CQC for the 2022 Urgent and Emergency Care Survey. By following the guidance in this document, it should be possible to recreate this cleaning process. This document should be used alongside the 2022 Urgent and Emergency Care [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 2022 Urgent and Emergency Care 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](#)). The requirement for raw / uncleaned data does not, however, preclude the checking of data for errors resulting from problems with data entry or similar. 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.

---

<sup>1</sup> For specific information on how data is entered and coded by trusts/contractors before submission to the SCCEM, please refer to the [Entering and Submitting Final Data](#) guidance.

### Free text comments:

These are verbatim comments provided by a patient in response to the open question at the end of both surveys: 'If there is anything else you would like to tell us about your experiences in the A&E department / Urgent Treatment Centre, please do so here'. These responses should be provided in an excel file containing the trust code, patient record number, 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 patient 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. For the 2022 Urgent and Emergency Care Survey **Type 1** questionnaire, the ask-all questions are **Q1, Q5-Q8, Q12-Q19, Q21-Q27, Q30-Q34, Q46-Q49, Q52-Q58**.

For the 2022 Urgent and Emergency Care Survey **Type 3** questionnaire, the ask-all questions are **Q1, Q5-Q9, Q12-Q19, Q21-Q26, Q28-Q32, Q40-Q43, Q46-Q52**.

### Routing questions

These are items in the questionnaire which instruct respondents to either continue on to the next question or to skip irrelevant questions depending on their response to the routing question.

For the 2022 Urgent and Emergency Care Survey **Type 1** questionnaire, the routing questions in the questionnaire are **Q1, Q2, Q8, Q19, Q27, Q28, Q34, Q36 and Q49**.

For the 2022 Urgent and Emergency Care Survey **Type 3** questionnaire, the routing questions in the questionnaire are **Q1, Q2, Q9, Q19, Q26, Q32 and Q43**.

### Filtered questions

These are items on the questionnaire that 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 2022 Urgent and Emergency Care Survey **Type 1** questionnaire, the filtered questions in the questionnaire are **Q2-Q4, Q9-Q11, Q20, Q28, Q29, Q35-Q45, Q50 and Q51**.

For the 2022 Urgent and Emergency Care Survey **Type 3** questionnaire, the filtered questions in the questionnaire are **Q2-Q4, Q10, Q11, Q20, Q27, Q33-39, Q44 and Q45**.

## 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 2022 Urgent and Emergency Care Survey **Type 1** questionnaire, the multiple response questions are **Q2, Q3, Q40, Q50, Q52**.

For the 2022 Urgent and Emergency Care Survey **Type 3** questionnaire, the multiple response questions are **Q2, Q3, Q34, Q44, Q46**.

## Sample data

Patient data that is provided from the trust as part of the sampling process. This includes, gender, year of birth, ethnicity, patient's postcode, department type, date of attendance, time of attendance, sub location ICB code (formerly CCG code), NHS site code and mobile phone indicator as it is recorded on the trust's system.

## Response data

Data from the completed questionnaire which is provided from the patient. This includes answers to Q1 through Q58 for the Type 1 questionnaire, and Q1 to Q52 for the Type 3 questionnaire.

## Out-of-range data

This refers to instances where data within a variable have values that are not permissible. For categorical data – most of the variables in this survey – this would mean a value not allowed in the data entry sheet for that specific variable. For example, a value of '3' being entered in for a variable with only two response categories (1 or 2). A full list of valid responses for the 2022 Urgent and Emergency Care Survey can be found in the [data mapping document](#) in the 'all response values' column. Out-of-range sample data is listed in [Appendix B: Out- of-range sample data – Type 1 and Type 3](#).

## Outcome

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

- Outcome 1: Returned completed questionnaire (including accessible versions<sup>2</sup>)
- Outcome 2: Undelivered / moved house
- Outcome 3: Deceased after the start of fieldwork
- Outcome 4: Too ill / opt out
- Outcome 5: Ineligible
- Outcome 6: Unknown
- Outcome 7: Deceased before the start of fieldwork

---

<sup>2</sup> This includes completed easy read, braille and large print questionnaires.

## Non-specific responses

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 did not need any help” or “I did not need an explanation”. 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. A full list of non-specific responses for the 2022 Urgent and Emergency Care Survey can be found in the ‘excluded values’ column of the [data mapping document](#).

## 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, a number of different missing response codes are used to identify how data for a particular respondent has been handled. These codes are as follows:

- 999: Missing response - 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: Inapplicable - this code is used when someone answered a question but should not have. For example, filtered questions.
- 996: Cleaned - this code is used to suppress data at trust level when a question has fewer than 30 responses<sup>3</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, in house trusts and 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 patients 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, where respondents may select more than one response option.

---

<sup>3</sup> Non-specific responses (i.e., “Don’t know / can’t remember”) are excluded from the count.

- For Q47 the overall experience question (Q41 in Type 3 questionnaire), where a respondent has circled two numbers on the scale or has placed a marker in-between two numbers on the scale a code of '98' should be inputted.
- 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 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 Q55 in the Type 1 questionnaire ("What was your year of birth?"), then their year of birth should be entered.
- For the year of birth / age questions, unrealistic responses should still be entered except following the rule above. For example, if a respondent enters '2021' 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 patient'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 patient. This may include applying the outcome code priority order detailed in the next section.
- Free text comments given to the final question in both Type 1 and Type 3 surveys ('Any other comments?') should be submitted in the data entry spreadsheet. All free text comments, regardless of whether a respondent was outcome=1 or not.

## Multiple questionnaire responses - De-duplication and inclusion

This section outlines how to approach situations when a patient 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 patient in the 2022 Urgent and Emergency 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.	First
In the event that the total number of completed questions is equal on all questionnaires, the earliest questionnaire received should be selected.	Second
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.	Third

## Outcome code priorities

As patients are sent up to two copies of the questionnaire throughout the fieldwork period, there may be duplicate questionnaires returned, or the patient 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 2022 Urgent and Emergency Survey

Outcome Code	Priority
Outcome 1: Returned 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. We do this by removing responses that are known to be erroneous or inappropriate, but 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>4</sup>. Tables 3 and 4 below detail the response values that require cleaning and the appropriate filtered questions to recode as '998': table 3 lists all routing questions included in the 2022 Urgent and Emergency Care Survey Type 1 questionnaire, and table 4 lists all routing questions included in the 2022 Urgent and Emergency Care Survey Type 3 questionnaire.

Table 3. Appropriate cleaning for routing questions in the 2022 Urgent and Emergency Care Survey – Type 1 Questionnaire

Routing question	Response values requiring cleaning	Filtered questions to be recoded as '998'
Q1	1	Q2
Q2	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9	Q3, Q4
Q8	1	Q9, Q10, Q11
Q19	2	Q20
Q27	4	Q28, Q29
Q28	1 or 2 or 3 or 4	Q29
Q34	1	Q35 - 45
Q34	3	Q35
Q36	2	Q37, Q38
Q49	2	See <a href="#">Cleaning Special Cases</a>

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

<sup>4</sup> Code '998' is an arbitrary value chosen because it is out of range for all other questions on the survey.

Table 4. Appropriate cleaning for routing questions in the 2022 Urgent and Emergency Care Survey – Type 3 Questionnaire

Routing question	Response values requiring cleaning	Filtered questions to be recoded as '998'
Q1	1	Q2
Q2	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9	Q3, Q4
Q9	1	Q10, Q11
Q19	2	Q20
Q26	4	Q27
Q32	1 or 2	Q33 – Q39
Q43	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 removing unexpected responses to filtered questions is included in [Appendix A: Example of Cleaning](#).

The recoding of filtered questions into 998 only applies where the response to the routing question is not missing. Indeed, in this case, it is considered that the respondent contradicted their previous answer as seen in example 1 below.

### Example 1:

#### TESTS

Tests could include X-rays, scans, blood tests or urine tests.

27. If you had any tests, did a member of staff explain **why you needed them** in a way you could understand?

- 1  Yes, completely → Go to 28
- 2  Yes, to some extent → Go to 28
- 3  No → Go to 28
- 4  I did not have any tests → Go to 30

28. Before you left A&E, did a member of staff explain the **results of the tests** in a way you could understand?

- 1  Yes, definitely → Go to 30
- 2  Yes, to some extent → Go to 30
- 3  No → Go to 30
- 4  Not sure / can't remember → Go to 30
- 5  I was given the results after I left A&E → Go to 29

In the example above, the response to Q28 would be recoded to '998' because according to their answer from Q27 (the routing question), respondents were supposed to skip Q28.

However, where the response to a routing question is missing, responses to filtered questions are not removed. It is considered that the respondent might have been unsure or missed the routing question and therefore, their responses to filtered questions are still relevant (example 2 below).

### Example 2:

#### TESTS

Tests could include X-rays, scans, blood tests or urine tests.

27. If you had any tests, did a member of staff explain **why you needed them** in a way you could understand?
- 1  Yes, completely → Go to 28
  - 2  Yes, to some extent → Go to 28
  - 3  No → Go to 28
  - 4  I did not have any tests → Go to 30
28. Before you left A&E, did a member of staff explain the **results of the tests** in a way you could understand?
- 1  Yes, definitely → Go to 30
  - 2  Yes, to some extent → Go to 30
  - 3  No → Go to 30
  - 4  Not sure / can't remember → Go to 30
  - 5  I was given the results after I left A&E → Go to 29

In the example above (example 2), Q27 would be coded as missing ('999') and the response to Q28 would remain as code 1.

## Cleaning Special Cases

### Cleaning of the long-term condition questions

The long-term condition questions refer to **Q49, Q50 and Q51** in the **Type 1** questionnaire and **Q43, Q44 and Q45** in the **Type 3** questionnaire. The cleaning rules are the same for both questionnaires though in the following instructions, **Type 1** question numbers are used.

**Q49** 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.

**Q50** 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.

**Q51** Do any of these reduce your ability to care out day-to-day activities?

Although **Q49** is a routing question and **Q50** and **Q51** 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 **Q49** but has answered **Q50** by selecting one or more long term condition, their response to **Q49** is cleaned by setting it to missing (code ‘999’). This is because their response to **Q50** indicates that they do in fact consider they have a long-standing condition. Instead of changing their response to **Q49** to an answer the patient did not select (i.e. “Yes”), we set it to missing so that it is no longer inconsistent with their response to **Q50**.

However, when a respondent has answered “No” to **Q49** and has not selected any long-term conditions in **Q50** but has answered **Q51** then their response to **Q51** is cleaned by setting it to not-applicable (code ‘998’). As they have indicated that they do not have a long-term condition and have also skipped **Q50**, this would suggest that **Q51** was not applicable to them. See table 5 for a summary of how **Q49**, **Q50** and **Q51** in the Type 1 questionnaire are cleaned, and table 6 for a summary of how **Q43**, **Q44** and **Q45** in the Type 3 questionnaire are cleaned.

Table 5. Cleaning for Q49, Q50 and Q51 in the Type 1 questionnaire

Q49 response	Q50 response	Q51 response	Cleaning
Ticked option 2 (no)	Ticked one or more options	No response	Q49 is set to missing – ‘999’ and as there was no response to Q51 this would also be set to missing – ‘999’.
Ticked option 2 (no)	Ticked one or more options	Ticked any option	Q49 is set to missing – ‘999’.
Ticked option 2 (no)	No response	Ticked any option	Q51 is set to not-applicable – ‘998’.

Table 6. Cleaning for Q43, Q44 and Q45 in the Type 3 questionnaire

Q43 response	Q44 response	Q45 response	Cleaning
Ticked option 2 (no)	Ticked one or more options	No response	Q43 is set to missing – ‘999’ and as there was no response to Q45 this would also be set to missing – ‘999’.
Ticked option 2 (no)	Ticked one or more options	Ticked any option	Q43 is set to missing – ‘999’.
Ticked option 2 (no)	No response	Ticked any option	Q45 is set to not-applicable – ‘998’.

## Dealing with multiple response questions

For most questions, each column corresponds to one survey question. There are five exceptions to this rule: Q2, Q3, Q40, Q50 and Q52 in the Type 1 questionnaire and Q2, Q3, Q34, Q44 and Q46 in the Type 3 questionnaire where each response option is treated as a separate question.

For multiple response questions each response option that is ticked are coded as '1' in the data file and all other responses are coded as '0'. This is to ensure these questions can be appropriately counted for when assessing usability.

### Example

3. Before going to this A&E department, where did you go to, or contact, for help with your condition? **(Cross ALL that apply)**

- 1  999 emergency service
- 2  NHS 111 telephone service
- 3  NHS 111 online service
- 4  A different A&E department
- 5  Pharmacist
- 6  GP practice
- 7  GP out-of-hours service
- 8  Urgent Treatment Centre/ Urgent Care Centre / Minor Injuries Unit / Walk-in Centre
- 9  Somewhere else

The above example would be coded as the following:

Column headings	Q3_1	Q3_2	Q3_3	Q3_4	Q3_5	Q3_6	Q3_7	Q3_8	Q3_9
Coding for this example	1	0	1	0	0	1	0	0	0

## 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 16. When this occurs, respondents will *not* be considered ineligible for the survey and therefore remain as outcome 1 (see table 5). This is because of the difficulty of inferring the source of errors when year of birth from sample and response data are

mismatched. We cannot be certain whether this mismatch occurs due to an error in the sample file or an error in the patient's completion of the questionnaire form.

Another scenario could be that the respondent has indicated they are under the age of sixteen in the response data, but year of birth is missing from the sample data. As this is the only available data for their year of birth, these respondents would be considered ineligible for the survey and would therefore be recoded to outcome 5 (see table 7 and table 8).

Table 7. Eligibility and outcome codes of patients based on sample and response data of age – Type 1 questionnaire

Original outcome code	Sample data	Response data	Eligibility	Final outcome code
1	YoB ≤ 2006	Q55 > 2006	Eligible	1
1	YoB ≤ 2006	Q55 ≤ 2006	Eligible	1
1	YoB ≤ 2006	Q55 = missing	Eligible	1
1	YoB ≤ 2006	Q55 = out of range (≤ 1903 ≥ 2007)	Eligible	1
1	YoB = missing	Q55 > 2006	Ineligible	5
1	YoB = missing	Q55 ≤ 2006	Eligible	1
1	YoB = missing	Q55 = missing	Ineligible	5

Table 8. Eligibility and outcome codes of patients based on sample and response data of age – Type 3 questionnaire

Original outcome code	Sample data	Response data	Eligibility	Final outcome code
1	YoB ≤ 2006	Q47 > 2006	Eligible	1
1	YoB ≤ 2006	Q47 ≤ 2006	Eligible	1
1	YoB ≤ 2006	Q47 = missing	Eligible	1
1	YoB ≤ 2006	Q47 = out of range (≤ 1903 ≥ 2007)	Eligible	1
1	YoB = missing	Q47 > 2006	Ineligible	5
1	YoB = missing	Q47 ≤ 2006	Eligible	1
1	YoB = missing	Q47 = missing	Ineligible	5

## 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 patient 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 patient to report being born in 1985.

Where responses to demographic questions are present, it is assumed these are more likely to be accurate than 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 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>5</sup>.

## Out-of-range data

Out-of-range data are responses that have been entered into the questionnaire or the data entry spreadsheet that do not meet the set range of valid responses for that question. Out-of-range data is highlighted in the data entry spreadsheet via conditional formatting.

For example, a common error when completing the year of birth question is for respondents to accidentally write in the current year. In this case, the response to **Q55** in the Type 1 questionnaire (**Q49** in the Type 3 questionnaire) would be considered as an out-of-range response and would therefore be set to missing. For the 2022 Urgent and Emergency Care Survey, out of range responses for **Q55** in Type 1 (and **Q49** in Type 3) are defined as **Q55 ≤ 1903 or ≥ 2007 (this is the same for Q49 in the Type 3 questionnaire)**. 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 3 response options (i.e. Q9, Q11, Q13, Q15, Q17 Q18, Q23, Q25, Q29, Q32, Q34, Q43, Q44, Q46, Q51, Q54 in the Type 1 questionnaire and Q6, Q10, Q13, Q15, Q17, Q23, Q24, Q30, Q32, Q37, Q38, Q40, Q45, Q48 in the Type 3 questionnaire) with response data of **≤ 0 or ≥ 4** would be set to missing. A full list of valid responses for the 2022 Urgent and Emergency Care Survey is listed in the 'all response values' column of the [data mapping document](#). Out-of-range sample data is listed in [Appendix B: Out- of-range sample data – Type 1 and Type 3](#).

---

<sup>5</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.



## Usability

Sometimes questionnaires are returned with only a very small number of questions completed. As in previous years, questionnaires containing fewer than five responses are considered 'unusable' – we will set all responses pertaining to such cases as missing '999' and recode 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 will be counted after all cleaning has been conducted.

Additional clarification on what constitutes as five responses to determine if a questionnaire is usable:

- Demographic questions in the 'About You' section should be included within the count
- Verbatim comments in other comments are not counted towards the five responses
- Multiple response questions are counted once. For instance, Q50 from the Type 1 questionnaire would be counted as one response in the below scenario:

### Example

50. 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 (which 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>6</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.

The SCCEM will suppress results at both national and trust level for questions that have fewer than 30 respondents 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. For a full listing of 'non-specific' responses in the 2022 Urgent and Emergency Care Survey, please see the 'excluded values' column in the [data mapping document](#).

---

<sup>6</sup> This is an arbitrary value chosen because it is 'out-of-range' for all other questions on the survey.

# Appendix A: Example of cleaning

Figure 1 shows hypothetical raw / uncleaned data for eight patients, 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 '0005' and '0007' have reported that before they left A&E they were not prescribed any new medications (Q36=2), but have both responded to subsequent filtered questions which are only applicable for patients that were prescribed new medications before leaving A&E (respondent '0005' has answered the first filter question (Q37) before skipping the remaining filter questions, whilst respondent '0007' has answered Q37 and Q38).

Record	Outcome	Q36	Q37	Q38
Patient Record Number	Outcome of Sending questionnaire (N)	Before you left A&E, were you prescribed any new medications?	Did a member of staff explain the purpose of the medications you were to take at home in a way you could understand?	Did a member of staff tell you about medication side effects to watch for?
UEC22...00001	6			
UEC22...00002	1	2	.	.
UEC22...00003	1	1	2	3
UEC22...00004	4			
UEC22...00005	1	2	1	.
UEC22...00006	6			
UEC22...00007	1	2	3	2
UEC22...00008	1	1	1	2

Figure 1. Example of raw/'uncleaned' data.

Following the cleaning instructions, the filter instructions specify that:

Routing question	Response values requiring cleaning	Filtered questions to be recoded
Q36	2	Q37 – Q38

In accordance with this, all responses for **Q37** and **Q38** must be set to missing (not applicable code 998) in cases where the respondent has ticked **Q36 = 2** (i.e. "no they were not prescribed any new medication before leaving A&E").

Figure 2 below shows how the data would look after cleaning is complete by the SCCEM to remove responses to filtered questions that should have been skipped – cells where responses have been set to missing are shaded.

<b>Record</b>	<b>Outcome</b>	<b>Q36</b>	<b>Q37</b>	<b>Q38</b>
Patient Record Number	Outcome of Sending questionnaire (N)	Before you left A&E, were you prescribed any new medications?	Did a member of staff explain the purpose of the medications you were to take at home in a way you could understand?	Did a member of staff tell you about medication side effects to watch for?
UEC22...00001	6			
UEC22...00002	1	2	.	.
UEC22...00003	1	1	2	3
UEC22...00004	4			
UEC22...00005	1	2	998	.
UEC22...00006	6			
UEC22...00007	1	2	998	998
UEC22...00008	1	1	1	2

Figure 2. Example of cleaned data.

## Appendix B: Out-of-range sample data – Type 1 and Type 3

Variable	Out-of-range data
Birth	≤ 1903 ≥ 2007
Gender	≤ 0 3-8 ≥ 10
Ethnicity	Anything except A-H, J-N, P, R, S or Z
DateOfAttendance	≤ 0 ≥ 31
MonthOfAttendance (Type 1)	≤ 8 ≥ 10
MonthOfAttendance (Type 3)	≤ 7 ≥ 10
YearOfAttendance	≤ 2021 ≥ 2023
TimeOfAttendance	Anything not between the hours of 00:00 and 23:59
DayQRec	≤ 0 ≥ 32
MonthQRec	≤ 0 4-9 ≥ 13
YearQRec	≤ 2021 ≥ 2024

For a full list of valid responses, please refer to the [data mapping document](#).