

# DATA CLEANING GUIDANCE

# 2023 ADULT INPATIENT SURVEY

Last updated: March 2024

Contacts

The Survey Coordination Centre 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

E-mail: [inpatient@surveycoordination.com](mailto:inpatient@surveycoordination.com)

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

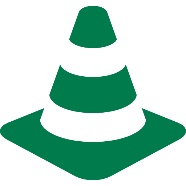
Updates

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](https://nhssurveys.org/surveys/survey/02-adults-inpatients/).

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 trusts and contractors taking part in the survey:**

Contractors submitting final data for the Adult Inpatient Survey **must not** clean their data before submitting it to the Survey Coordination Centre. Please refer to [Survey Handbook](https://nhssurveys.org/wp-content/surveys/02-adults-inpatients/03-instructions-guidance/2023/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

[Data cleaning – an overview 5](#_Toc160704262)

[Introduction 5](#_Toc160704263)

[Definitions 5](#_Toc160704264)

[Raw / uncleaned data: 5](#_Toc160704265)

[Free text comments: 5](#_Toc160704266)

[Data cleaning: 5](#_Toc160704267)

[Ask-all questions: 6](#_Toc160704268)

[Routing questions: 6](#_Toc160704269)

[Filtered questions: 6](#_Toc160704270)

[Multiple response questions: 6](#_Toc160704271)

[Multiple questionnaire responses: 6](#_Toc160704272)

[Sample data: 6](#_Toc160704273)

[Response data: 7](#_Toc160704274)

[Out-of-range data: 7](#_Toc160704275)

[Outcome: 7](#_Toc160704276)

[Non-specific response: 7](#_Toc160704277)

[Missing responses: 7](#_Toc160704278)

[Entering and coding data prior to submission 8](#_Toc160704279)

[Multiple Questionnaire Responses - De-duplication and Inclusion 9](#_Toc160704280)

[Outcome code priorities 9](#_Toc160704281)

[Editing and cleaning data after submission 10](#_Toc160704282)

[Approach and rationale 10](#_Toc160704283)

[Dealing with filtered questions 10](#_Toc160704284)

[Example 1: 11](#_Toc160704285)

[Example 2: 11](#_Toc160704286)

[Cleaning multi-code questions 12](#_Toc160704287)

[Eligibility 12](#_Toc160704288)

[Age / Year of birth 12](#_Toc160704289)

[Demographics 13](#_Toc160704290)

[Out-of-range data 14](#_Toc160704291)

[Usability 14](#_Toc160704292)

[Example 15](#_Toc160704293)

[Missing responses 15](#_Toc160704294)

[Non-specific responses 16](#_Toc160704295)

[Weighting 17](#_Toc160704296)

[National data weights 17](#_Toc160704297)

[Trust data weights 17](#_Toc160704298)

[Other data weights 17](#_Toc160704299)

[Appendix A: Example of cleaning 18](#_Toc160704300)

[Appendix B: In-range data 20](#_Toc160704301)

# Data cleaning – an overview

## Introduction

Once fieldwork for the 2023 Adult Inpatient Survey has been completed, data needs to be submitted to the SCC in a raw / uncleaned format. 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 2023 Adult Inpatient 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. This document should be used alongside the 2023 Adult Inpatient [data mapping document](https://nhssurveys.org/surveys/survey/02-adults-inpatients/) 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 2023 Adult Inpatient 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/)). 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 patient in response to the three open questions at the end of the survey: “Was there anything particularly good about your hospital care?”; “Was there anything that could be improved?”; and “Any other comments?”. These responses should be included within the data entry spreadsheet. 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 SCC even though they have not answered the other questions in the survey.

## Data cleaning:

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

## Ask-all questions:

These are items in the questionnaire which are not subject to any filtering, and which should therefore be answered by all respondents. For the 2023 Adult Inpatient Survey, the ask-all questions are **Q1, Q5-Q7, Q9-Q12, Q16-Q32, Q35-Q39, Q41-Q45, Q47-Q52 and Q54-Q61.**

## Routing questions:

These are items in the questionnaire which instruct respondents either to continue to the next question or to skip irrelevant questions, depending on their response to the routing question. For the 2023 Adult Inpatient Survey, the routing questions in the questionnaire are **Q1, Q7, Q12, Q32, Q39, Q45 and Q52.**

## 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 2023 Adult Inpatient Survey, the filtered questions in the questionnaire are **Q2-Q4, Q8, Q13-Q15, Q33, Q34, Q40, Q46 and Q53.**

## 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 2023 Adult Inpatient Survey, the multiple response questions are **Q6, Q16, Q41, Q52, Q54.**

* A code of ‘98’ is used at Q49 if respondents have circled more than one answer and it cannot be determined which number they wanted to circle.

## Multiple questionnaire responses:

The 2023 Adult Inpatient Survey has used both postal and SMS reminders. All patients receive the first mailing letter, those aged under 80 receive a letter with a link and QR code to access the survey online, whilst those over 80 also get a paper copy of the questionnaire. After 5 days those that haven't completed the survey get a text reminder. Then at 8 days, they receive a reminder letter. At 12 days, they receive a final text reminder. The final reminder is at 22 days when they receive a letter with the option of a paper questionnaire to complete.

Though the reminders are only sent out to participants who have not completed the survey, there is the potential for an individual respondent / patient to complete multiple questionnaires, such as completion of the paper questionnaire and the online if there has been an overlap in mailings.

## Sample data:

Patient data that is provided by the trust as part of the sampling process. This includes: gender, year of birth, ethnicity, date of admission and discharge, length of stay, treatment function code, admission method code, ICD-10 chapter code, NHS site code for both admission and discharge, and a virtual ward indicator.

## Response data:

Data from the completed questionnaire which is provided from the patient. This includes answers to Q1 through to Q61.

## 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 SCC. A full list of in-range responses for the 2023 Adult Inpatient Survey can be found in [Appendix B: In-range data](#AppB).

## Outcome:

An outcome code is given to each patient 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 patients are coded appropriately. The coding for outcome is as follows:

Outcome 1: Returned completed questionnaire (including accessible versions)

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

## 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 did not need help to eat meals” or “I did not bring medication with me to hospital”. 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.

**Please note, contractors should submit raw ‘uncleaned’ data to the SCC, as per the ‘entering and coding data prior to submission’ section below.**

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.
* 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.

# Entering and coding data prior to submission

For the 2023 survey, 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 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). The exceptions to this are for the ‘multiple response’ questions (e.g., Q6), where respondents may select more than one response option and for Q49 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). 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 (“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 ‘2023’ 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 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 patient.
* Free text comments given in the final three questions in the survey (“Was there anything particularly good about your hospital care?”; “Was there anything that could be improved?”; and “Any other comments?”) should be submitted within the data entry spreadsheet.

## 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 2023 Adult Inpatient 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 data used are selected according to a priority order, and the earliest questionnaire received (either online or paper) 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, priority will be given to the response completed online. | **Third** |

## Outcome code priorities

As patients are offered the option of completing the questionnaire online or in paper 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 2023 Adult Inpatient Survey

|  |  |
| --- | --- |
| **Outcome Code** | **Priority** |
| Outcome 1: Returned completed questionnaire | **1st** |
| Outcome 7: Deceased prior to fieldwork | **2nd** |
| Outcome 3: Deceased during fieldwork | **3rd** |
| Outcome 5: Ineligible | **4th** |
| Outcome 4: Opted out | **5th** |
| Outcome 2: Undelivered / moved house | **6th** |
| Outcome 6: Unknown | **7th** |

# 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.

## 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.[[1]](#footnote-3) See table 3 for a list of all routing questions included in the 2023 Adult Inpatient Survey, the response values that require cleaning, and the appropriate filtered questions to recode as ‘998’.

Table 3. Appropriate cleaning for routing questions in the 2023 Adult Inpatient Survey

|  |  |  |
| --- | --- | --- |
| **Routing question** | **Response values requiring cleaning** | **Filtered questions to be recoded to 998** |
| **Q1** | 2 | **Q2 - Q4** |
| **Q7** | 3, 4 | **Q8** |
| **Q12** | 5, 6 | **Q13 - Q15** |
| **Q32** | 2,3 | **Q33, Q34** |
| **Q39** | 2, 3 | **Q40** |
| **Q45** | 4 | **Q46** |
| **Q52** | 17, 18 | **Q53** |

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 [Appendix A: example of cleaning](#_Appendix_A:_Example).

Responses are only recoded as 998 where respondents have answered filtered questions despite ticking an earlier response on a routing question instructing them to skip these questions.

## Example 1:

A screenshot of a medical survey

Description automatically generated

In the example above (example 1), the response to Q2 would be recoded to ‘998’ because according to their answer from Q1 (the routing question), they were supposed to skip Q2.

Responses to filtered questions are not removed where the response to the routing question is missing.

## Example 2:

A screenshot of a survey

Description automatically generated

In the example above (example 2), the response to Q13 would remain as code 5 and Q12 would be coded as missing (999).

## Cleaning multi-code questions – Incompatible answer codes

Where participants have answered two incompatible codes in a multi-code question, these should be removed, as it is not possible for both those answers to be correct. For example, at Q6 participants cannot select both "Noise from staff" and "I was not prevented from sleeping" as reasons for being prevented from sleeping at night.

Table 4. List of multi-code questions and answer codes that can only be single-coded



|  |  |
| --- | --- |
| **Condition for multi-code questions** | **Recode** |
| **Q6 any of option 1 to 7 = 1 AND Q6 option 8 = 1** | **Q6 = 997** |
| **Q16 option 1 = 1 AND Q16 any of option 2 3 4 5 = 1** | **Q16 = 997** |
| **Q16 option 5 = 1 AND Q16 any of option 2 3 4 = 1** | **Q16 = 997** |
| **Q41 any option 1 2 3 4 5 = 1 AND Q41 option 6 = 1** | **Q41 = 997** |
| **Q41 any option 1 2 3 4 6 = 1 AND Q41 option 5 = 1** | **Q41 = 997** |
| **Q52 any of option 1 to 16 = 1 AND Q52 option 17 = 1** | **Q52 = 997** |
| **Q52 any of option 1 to 16 = 1 AND Q52 option 18 = 1** | **Q52 = 997** |
| **Q52 option 17 = 1 AND Q52 option 18 = 1** | **Q52 = 997** |
| **Q54 any of option 1 to 3 = 1 AND Q54 option 4 = 1** | **Q54 = 997** |

Where participants have selected incompatible answer codes at a multi-code question, or have selected more than one answer code at a single code questions, these responses should be recoded as 997.

## 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 *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 patient’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 4.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Original outcome code** | **Sample data** | **Response data** | **Eligibility** | **Final outcome code** |
| 1 | YoB ≤ 2008 | Q55 > 2008 | Eligible | 1 |
| 1 | YoB ≤ 2008 | Q55 ≤ 2008 | Eligible | 1 |
| 1 | YoB ≤ 2008 | Q55 = missing | Eligible | 1 |
| 1 | YoB ≤ 2008 | Q55 = out-of-range | Eligible | 1 |
| 1 | YoB = missing | Q55 > 2008 | Ineligible | 5 |
| 1 | YoB = missing | Q55 = missing | Ineligible | 5 |
| 1 | YoB = missing | Q55 = out-of-range | Ineligible | 5 |

## Demographics

In a small number of cases, sample data and response data does not correspond for age and gender/sex. 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 patient 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/sex 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)[[2]](#footnote-4).

## 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 **Q55** would be considered as an out-of-range response. For the 2023 Adult Inpatient Survey, out-of-range responses for **Q55** are recoded as ‘997’. The out-of-range responses for **Q55** are defined as **Q55 ≤ 1900 or Q55 ≥ 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 also be set for invalid responses to all other questions in the survey. The out-of-range responses will depend on the number of response options given for each question. For instance, all questions with three response options (e.g., Q1, Q18 or Q23) with response data of **≤ 0 or ≥ 4** would be set to missing.

A list of in-range responses for the 2023 Adult Inpatient Survey are listed in Appendix B: In-range data.

## Usability

Sometimes questionnaires are returned with only a very small number of questions completed. For the 2023 Adult Inpatient Survey, 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, Q52 would be counted as one response in the below scenario.

## Example

A screenshot of a questionnaire

Description automatically generated

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 responses

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

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

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

The SCC codes missing responses in the data as ‘999’[[3]](#footnote-5). 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.

The SCC will suppress results at both national and trust level for questions that have fewer than 30 respondents[[4]](#footnote-6)[[5]](#footnote-7) 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 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.

List of non-specific responses is detailed in the [data mapping document](https://nhssurveys.org/surveys/survey/02-adults-inpatients/).

# Weighting

Weighting is used to ensure trusts are comparable with one another, standardising for demographic differences, and to take into account non-response, to ensure results are representative of the populations being measured.

### National data weights

When calculating national data, weights are used to account for non-response, by weighting the completed population back to the original sample. To do this, the sample is split into strata by gender, age band and route of admission (elective or emergency) from the initial sample. A weight is then calculated to ensure each stratum is the same size in the completed responses as in the initial sample, for each trust. For example, if men aged 16-35 admitted in an emergency made up 10% of the initial sample in one trust, but only 5% of their responses, these respondents would be given a weight of 2, so this group would now be twice the size and make up 10% of responses. This weight is capped at 5 to ensure that no excessive weights are used.

An additional weight is also applied per question to make sure each trust has the same number of weighted respondents. This is done to ensure no trust is over- or underrepresented in the national results.

To calculate the national data, the two weights are multiplied together and applied before the data is run, meaning each trust is an equal size and the results reflect the sampled population.

### Trust data weights

To calculate trust scores, a weight is used to standardise the trusts by age band, gender and route of admission (elective or emergency). This is to ensure that trusts do not appear to be performing better or worse than one another, simply because they are serving a different population or providing more of a different type of care.

Unlike with the national data weights, the strata are calculated using a combination of sample data and questionnaire data. These strata are then calculated to match the overall population of responses to the survey at a national level, for every trust. Therefore, if 10% of respondents to the survey were women aged 66+ admitted for elective care, then each trust would be weighted to ensure 10% of their responses were from this group. This ensures every trust has a consistent population.

### Other data weights

Separate weights are used for site level, medical and surgical analysis. These follow a similar process as the trust weight, but at site level or medical or surgical level within a trust, rather than at overall trust level.

# 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 ‘003’ and ‘005’ have answered questions about their virtual ward experience (Q33 and Q34) even though they said they did not or did not know / could not remember being admitted to a virtual ward in Q32, meaning that they should have skipped Q33 and Q34, and gone straight to Q35. Respondent ‘008’ has followed the routing correctly.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Record** | **Outcome** | **Q32** | **Q33** | **Q34** | **Q35** |
| Patient Record Number | Outcome of sending questionnaire (N) | When leaving hospital, were you admitted onto a virtual ward, also known as hospital at home? | Were you given enough information about the care and treatment you would receive while on a virtual ward? | Before being admitted onto a virtual ward, did hospital staff give you information about the risks and benefits of continuing your treatment on a virtual ward? | To what extent did staff involve you in decisions about you leaving hospital? |
| 001 | 6 |  |  |  |  |
| 002 | 1 | 1 | 1 | 2 | 1 |
| 003 | 1 | 2 | 4 | 2 | 1 |
| 004 | 4 |  |  |  |  |
| 005 | 1 | 3 | 3 | 4 | 4 |
| 006 | 6 |  |  |  |  |
| 007 | 1 | 1 | 3 | 3 | 4 |
| 008 | 1 | 2 |  |  | 1 |

Figure 1. Example of raw / uncleaned data

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Routing question** | | | **Response values requiring cleaning** | **Filtered questions to be recoded** |
|  | **Q32** |  | 2 or 3 | **Q33, Q34** |

In accordance with this, all responses for **Q33** and **Q34** must be recoded as ‘998’. In cases where the respondent has answered **Q32 = 2 or 3** (i.e., had not been admitted to a virtual ward).

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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Record** | **Outcome** | **Q32** | **Q33** | **Q34** | **Q35** |
| Patient Record Number | Outcome of sending questionnaire (N) | When leaving hospital, were you admitted onto a virtual ward, also known as hospital at home? | Were you given enough information about the care and treatment you would receive while on a virtual ward? | Before being admitted onto a virtual ward, did hospital staff give you information about the risks and benefits of continuing your treatment on a virtual ward? | To what extent did staff involve you in decisions about you leaving hospital? |
| 001 | 6 |  |  |  |  |
| 002 | 1 | 1 | 1 | 2 | 1 |
| 003 | 1 | 2 | 998 | 998 | 1 |
| 004 | 4 |  |  |  |  |
| 005 | 1 | 3 | 998 | 998 | 4 |
| 006 | 6 |  |  |  |  |
| 007 | 1 | 1 | 3 | 3 | 4 |
| 008 | 1 | 2 |  |  | 1 |

Figure 2. Example of cleaned data

# Appendix B: In-range data

The [data mapping document](https://nhssurveys.org/surveys/survey/02-adults-inpatients/) published for 2023 Adult Inpatient Survey indicates the in-range values for each survey question. The below table (Table 5) covers the in-range data for sample information, or any information completed during fieldwork.

Table 6. In-range sample data

|  |  |
| --- | --- |
| **Sample Variable** | **In-range data** |
| Mobile number indicator | 0,1 |
| Year of Birth | ≥ 1900 or ≤ 2008 |
| Gender | 0, 1, 2, 9 |
| Ethnic category | Anything except I, O, Q, T-Y |
| Day of Admission Day of Discharge | ≥ 1  ≤ 31 |
| Month of Admission Month of Discharge | ≥ 1  ≤ 12 |
| Year of Admission | 2022, 2023 |
| Year of Discharge | 2023 |
| Length of Stay | >0 ≤ 9999 |
| Treatment Function Code (on discharge) | See list of Treatment Function codes by [clicking here](https://www.datadictionary.nhs.uk/attributes/treatment_function_code.html). |
| ICD-10 Chapter Code | I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIV, XV, XVI, XVII, XVIII, XIX, XX, XXI, XXII, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 |
| Admission method | See list of Admission method codes by [clicking here](https://www.datadictionary.nhs.uk/attributes/admission_method.html). |
| NHS Site code-Admitted | See list of NHS Site Codes by [clicking here](https://files.digital.nhs.uk/assets/ods/current/ets.zip). |
| NHS Site code-Discharged | See list of NHS Site Codes by [clicking here](https://files.digital.nhs.uk/assets/ods/current/ets.zip). |
| Virtual Ward Indicator | 0, 1 |
| Day Questionnaire Received | ≥1  ≤ 31 |
| Month Questionnaire Received | ≥ 1  ≤5 |
| Year Questionnaire Received | 2024 |
| Outcome Code | ≥1  ≤ 7 |

1. Code ‘998’ is an arbitrary value chosen because it is out-of-range for all other questions on the survey. [↑](#footnote-ref-3)
2. 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-4)
3. This is an arbitrary value chosen because it is ‘out-of-range’ for all other questions on the survey. [↑](#footnote-ref-5)
4. This does not include the demographic items included in the ‘about you’ section of the questionnaire. [↑](#footnote-ref-6)
5. Non-specific responses (i.e., “Don’t know / can’t remember”) are excluded from the count. [↑](#footnote-ref-7)