GUIDANCE ON DATA CLEANING FOR THE NHS COMMUNITY MENTAL HEALTH SERVICE USERS SURVEY 2013

THE CO-ORDINATION CENTRE FOR THE NHS SURVEYS PROGRAMME



Making patients' views count

Last updated: 8th July 2013

Contacts

The Patient Survey Co-ordination Centre Picker Institute Europe Buxton Court 3 West Way Oxford OX2 0BJ

Tel:01865 208127Fax:01865 208101E-mail:mentalhealth.data@pickereurope.ac.ukWebsite:www.nhssurveys.org

Key personnel

Chris Graham (Director of Research & Policy)

Esther Howell Kelsey Flott Caroline Killpack Jenny King Katherine Körner Sheena MacCormick Nick Pothecary Stephen Sizmur Alexandra Skew Mark Waters

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 Co-ordination Centre website at:

www.NHSSurveys.org

Questions and comments

If you have any questions or concerns regarding this document, or if you have any specific queries regarding the submission of data, please contact the Co-ordination Centre:

By e-mail: Mentalhealth.data@PickerEurope.ac.uk

By phone: 01865 208127

Contents

1	Mental Health Survey 2013 – data cleaning	1
1.1	Introduction	1
1.2	Definitions	1
2	Submitting raw ('uncleaned') data	3
3	Editing/cleaning data after submission	4
3.1	Approach and rationale	
3.2	Filters	4
3.3	Dealing with multiple response questions	5
3.4	Demographics	7
3.5	Usability and eligibility	В
3.6		
3.7	Non-specific responses	8
4	Appendix A: Example of cleaning10	C
4.1	Incorrectly followed routing10	C
5	Appendix B: Non-specific responses	2

1 Mental Health Survey 2013 – data cleaning

1.1 Introduction

Once fieldwork for the 2013 Community Mental Health Service Users survey has been completed, participating trusts and contractors are required to submit data to the Co-ordination Centre in a **raw** (**'uncleaned'**) format. Following this, data for all trusts in the survey is collated and the full dataset is cleaned together to ensure that cleaning taking place on data for each trust is comparable.

This document provides a description and specification of the processes that is used by the Coordination Centre to clean and standardise data submitted by contractors and trusts as part of the 2013 Community Mental Health Service Users survey. By following the guidance contained in this document, it should be possible to recreate this cleaning process.

If you have any comments or queries regarding this document please contact the Co-ordination Centre on **01865 208127**, or e-mail us at <u>mentalhealth.data@pickereurope.ac.uk</u>.

1.2 Definitions

Definitions of terms commonly used in this document, as they apply to the Community Mental Health Service Users survey are as follows:

Raw/uncleaned data: 'Raw' or 'uncleaned' data is data that has been entered verbatim from completed questionnaires without any editing taking place to remove contradictory or inappropriate responses. All responses ticked on the questionnaire should be included in the data entry spreadsheet¹ (see <u>Section 2: Submitting raw ('uncleaned') data</u>, for detailed guidance on submitting raw data). 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.

Data cleaning: The Co-ordination Centre uses the term 'data cleaning' to refer to all editing processes applied to survey data once the survey has been completed and the data has been entered and collated.

Routing questions: These are items on the questionnaire which instruct respondents either to continue on to the next question or to skip past irrelevant questions depending on their response to the routing question. For the 2013 Community Mental Health Service Users survey, the routing questions in the questionnaire are **Q9**, **Q11**, **Q15**, **Q17**, **Q19**, **Q22**, **Q24**, **Q28**, **Q34**, and **Q35**.

Please note that although Question 1 is technically a routing question, any respondents who tick option 7 ('I have never seen anyone from NHS mental health services') should be recoded as '5' (ineligible).

¹ Except where: a) multiple responses have been ticked - set these to missing (the **exceptions** to this are for the 'tick all that apply' question **Q52** where respondents may tick more than one response option); b) year of birth has been entered in incorrect format - if the service users' *intended* response is unambiguous from the questionnaire, then enter this.

Filtered questions: Items on the questionnaire that are not intended to be answered by all respondents are referred to as 'filtered' questions. Whether individual respondents are expected to answer filtered questions depends on their responses to preceding routing questions. For the 2013 Community Mental Health Service Users survey, the filtered questions in the questionnaire are **Q10- Q16, Q18, Q20-Q21, Q23-Q27, Q29-Q33, Q35-Q36.**

Non-filtered 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 7 at Q1). For the 2013 Community Mental Health Service Users survey, the non-filtered questions are **Q1-Q9**, **Q17**, **Q19**, **Q22**, **Q28**, **Q34**, **Q37-Q55**.

Out-of-range data: This refers to instances where data within a variable has a value that is not permissible. For categorical data – as in the case of the majority of 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). For scalar data – eg year of birth – data is considered to be out-of-range if it specifies a value that is not possible (for instance, year of birth as 983 or 2983). Out-of-range responses entered into the dataset **should not** be automatically (eg, algorithmically) removed prior to submitting the data to the Co-ordination Centre (see <u>Section 2:</u> <u>Submitting raw ('uncleaned') data</u>.

Non-specific response: This is a loose 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", which indicate a failure to recall the issue in question. 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 support" or "I do not have any caring responsibilites". A full list of such responses for the 2013 Community Mental Health Service Users survey can be found in <u>Appendix B: Non-specific responses</u>.

2 Submitting raw ('uncleaned') data

For the 2013 Community Mental Health Service Users survey, trusts and contractors are required to submit raw ('uncleaned') data to the Co-ordination Centre. 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 (eg, where service users answer questions that they have been directed to skip past, these responses should still be entered)
- Where a respondent has ticked more than one response category on a question, this should be set to 'missing' in the data. The exception to this is the 'multiple response' question, Q52, where respondents may tick more than one response option (i.e. 'tick all that apply')
- iii) Where a respondent has crossed out a response, this should not be entered in the data. Where a respondent has crossed out a response and instead ticked a second response option, this second choice should be entered into the data
- iv) Where a respondent has given their response inconsistently with the formatting of the questionnaire but where their intended response is nonetheless unambiguous on inspection of the completed questionnaire, then the respondent's *intended* response should be entered. For example, where a patient has written their *date* of birth in the boxes for Q49, but written their *year* of birth in at the side of this, then the respondent's year of birth should be entered.
- v) For the year of birth question, unrealistic responses should still be entered *except* following iv) above. For example, if a respondent enters '2013' 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
- vi) 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 users' 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 Co-ordination Centre where they are found to have been entered inconsistently with the respondent's intended response.

3 Editing/cleaning data after submission

3.1 Approach and rationale

The aim of the Co-ordination Centre in cleaning the data submitted to us 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 do this in a relatively permissive way so as to enable as many responses as possible to contribute to the overall survey results.

3.2 Filters

Some of the questions included in the survey are only relevant to a subset of respondents, and in these cases filter instructions on the questionnaire are used to route respondents past questions that are not applicable to them. For example, people who tick "no" to **Q28** ("In the last 12 months, have you had a care review meeting to discuss your care?") are instructed to skip all further questions on the care review (**Q29-Q33**).

It is necessary to clean the data to remove inappropriate responses where filter instructions have been incorrectly followed. In such cases, participants' responses to questions that were not relevant to them are deleted from the dataset. Responses are only deleted where respondents have answered 'filtered' questions despite ticking an earlier response on a routing question instructing them to skip these questions (eg a respondent ticking "No" to **Q28** but then answering the five questions about the care review as in the example above). Responses to 'filtered' questions are not removed where the response to the routing question is missing. For example, **Q29-Q33** are filtered by the response to **Q28** (e.g. if **Q28=3 or 4**), but if a respondent does not answer **Q28**, or if the **Q28** response is missing for any reason, then responses to **Q29-Q33** should not be removed.

<u>Figure 1</u> (overleaf) shows a summary of all routing questions, and the filtered questions they relate to, that are included in the 2013 Community Mental Health Service Users survey. Please note that these instructions should be followed in the order provided to be consistent with the procedures applied by the Co-ordination Centre.

	ROUTING QUESTION		RESPONSE VALUES		FILTERED QUESTIONS
if	Q1	=	7	Recode as outcome=5 (inelig	nible)
if	Q9	=	2	then delete responses to:	Q10 - Q16
if	Q11	=	2 OR 3	then delete responses to:	Q12 – Q14
if	Q15	=	2	then delete responses to:	Q16
if	Q17	=	2	then delete responses to:	Q18
if	Q19	=	2 OR 3	then delete responses to:	Q20-Q21
if	Q22	=	4	then delete responses to:	Q23-Q26
if	Q22	=	5	then delete responses to:	Q23-Q27
if	Q24	=	3	then delete responses to:	Q25
if	Q28	=	3 OR 4	then delete responses to:	Q29-Q33
if	Q34	=	2 OR 3	then delete responses to:	Q35-Q36
if	Q35	=	2	then delete responses to:	Q36

Figure 1 - List of routing/filtering instructions

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 <u>Appendix A: Example of cleaning</u>.

3.3 Dealing with multiple response questions

For most questions, each column corresponds to one survey question. However, there is one exception to this rule. For the multiple response question **Q52** that gives the instruction 'Tick all that apply', each response option is treated as a separate question.

Example

Q52.Are you currently in paid work? TICK ALL THAT APPLY

- ¹ Yes, I am working between 1-15 hours a week
- $_{2}$ Tyes, I am working 16 or more hours a week
- ³ No, I am retired
- ⁴ No, I do voluntary work
- $_{5}$ \checkmark No, but I am a full time student
- $_{6}$ \Box No, other reason

Responses to each part of this question are coded: 1 if the box is ticked 0 if the box is not ticked²

Q52 takes up six columns in the data file, labelled as follows:

Column headings	Q52_1	Q52_2	Q52_3	Q52_4	Q52_5	Q52_6
Codings for this example	1	0	0	0	1	0

.....

² Please note: if a respondent does not answer any part of a multiple response question, (ie does not tick any of the response options) then it should be left blank or coded as a full stop (.)

3.4 Demographics

Basic demographic information is included in the sample section of the data, but the 'About You' section at the end of the questionnaire also asks respondents to provide this information. In a minority of cases, the information provided from the sample frame and by the respondents does not correspond on age and gender – for example, the sample may identify an individual as male only for them to report being female (e.g. **Q48=2**).

Because of this, and because questions about demographics tend to produce relatively high item non-response rates, it is not appropriate to rely on either source of data alone.

Where responses to demographic questions are present, it is assumed these are more likely to be accurate than sample frame information (since it is assumed that respondents are best placed to know their own gender and age)³. Where responses to demographic questions are missing, however, sample data are used in their place.

For demographic analysis on groups of cases, then, it is necessary to use some combination of the information supplied in the sample frame and by the respondents. To do this, we first copy all valid responses to survey demographic questions into a new variable. Where responses are missing we then copy in the relevant sample information (note that for a very small number of service users demographic information may be missing in both the sample and response sections; in such cases data must necessarily be left missing in the new variable)⁴.

Certain demographic variables require special consideration during data cleaning:

Age (Q49)

A common error when completing year of birth questions on forms is for respondents to accidentally write in the current year – thus responses to **Q49** of '2013' will be set to missing during cleaning. Out-of-range responses will also be set to missing⁵. For the 2013 Mental Health Service Users survey, out-of range responses are defined as **Q49**≤**1892 OR Q49**≥**2013**.

³ Respondents will *not* be considered ineligible for the survey solely on the basis of their response to the year of birth question indicating that they are under the age of eighteen. This is because of the difficulty of inferring the source of errors when year of birth from sample and response sections are mismatched – in other words, we cannot be certain whether this mismatch occurs *[see overleaf]* due to an error in the sample file, an error in the patient's completion of the questionnaire form, or an error in data entry.

⁴ While a combination of sample and response information is used for demographic analysis on groups of cases, **only the sample information is used to calculate response rates by demographic groups**. Unlike subgroup analysis of variables which is only carried out for respondents, response rates are calculated using all patients in the sample. Because response rates vary between groups with differing demographic variables, 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 information should be used to calculate response rates by demographic groups.

⁵ The majority of out-of range responses present in data relating to year of birth questions result from errors in data entry (for example, not keying one of the digits – so '1983' may become 983, 183, 193, or 198). In such cases it is important that the responses be checked against the completed questionnaire forms, and data corrected if necessary, **prior** to submission of data to the Co-ordination Centre.

3.5 Usability and eligibility

Sometimes questionnaires are returned with only a very small number of questions completed. For the Mental Health Service Users survey, questionnaires containing fewer than five responses are considered 'unusable' – we will delete all responses pertaining to such cases, and outcome codes of 1 ('returned useable questionnaire') relating to these cases will be changed to 6 ('questionnaire not returned'). The number of responses per questionnaire will be counted after all other cleaning, and that all responses remaining at this stage will be counted (including responses to the demographic questions)⁶. This should affect only a very limited number of cases, and so should not have a significant impact on response rates.

Outcome codes for respondents will also be changed if the respondents are believed to be under the age of eighteen and therefore ineligible for participation. Since the sample files for the survey are checked by the Co-ordination Centre prior to mailing, this is unlikely to affect more than a handful of cases throughout the survey, as service users coded as being aged under 18 will be identified and removed from the sample before the start of the survey. Sample members will not, however, be removed from the sample if data on their year of birth is missing from the sampling frame. If sample information on a respondent's year of birth is missing, though, and their response to Q49 indicates that they are under 18 (specifically, if Q49>1995) then the outcome code for that service user should be recoded from 1 ('returned completed questionnaire') to 5 ('ineligible for participation in the survey'). This should only be done where sample information is missing. If sample information indicates a patient is aged 18 or over, but this is contradicted by the service user's response, then the service user's survey outcome should remain as 1. This is to avoid removing legitimate responses because of an overly conservative approach to assessing eligibility; in other words, where the service user'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.

3.6 Missing responses

It is useful to be able to see the numbers of respondents who have missed each question for whatever reason. Responses are considered to be missing when a respondent is expected to answer a question but no response is present. For non-filtered 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 to the survey 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 Co-ordination Centre codes missing responses in the data with the value 999⁷. For results to be consistent with those produced by the Co-ordination Centre, missing responses should be presented but should not be included in the base number of respondents for percentages.

3.7 Non-specific responses

⁶ The multiple choice question Q52 is only counted once. So if, for example, Q52_1 and Q52_4 are ticked, this would count as only one response for the purpose of determining if a questionnaire is usable.

⁷ This is an arbitrary value chosen because it is 'out-of-range' for all other questions on the survey.

As well as excluding missing responses from results, the Co-ordination Centre also removes nonspecific responses from base numbers for percentages. The rationale for this is to facilitate easy comparison between institutions by presenting only results from those service users who felt able to give an evaluative response to questions. For a full listing of 'non-specific' responses in the 2013 Community Mental Health Service Users survey, please see <u>Appendix B: Non-specific</u> <u>responses</u>.

.....

4 Appendix A: Example of cleaning

4.1 Incorrectly followed routing

		i igui e z. e	латріе та	uutu			
Record	Outcome	Q28	Q29	Q30	Q31	Q32	Q33
Patient Record Number	Outcome of sending questionnaire (N)	In the last 12 months have you had a care review meeting to discuss your care?	Were you told that you could bring a friend, relative or advocate to your care review meetings?	Before the review meeting, were you given a chance to talk to your co- ordinator about what would happen?	Were you given a chance to express your views at the meeting?	Did you find the care review helpful	Did you discuss whether you needed to continue using NHS mental health services?
A	6						
В	1	2	1	2	1	2	1
С	1	1	1	1	2		1
D	4						
E	1	3	3	3			
F	6						
G	1	4	2	2	3	3	3
Н	1	1	3	2	2	3	1

Figure 2: Example 'raw'/'uncleaned' data

<u>Figure 2</u> shows hypothetical raw/uncleaned data for eight sample members, five of whom have responded to the survey. It can be seen from this data that some of the respondents have followed filter instructions from routing questions incorrectly:

Respondent 'E' has reported that they have not had a care review meeting in the last 12 months (**Q28=3**) and Respondent 'G' has reported that they cannot remember if they have had a care review meeting (**Q28=4**), but have both responded to filtered questions: 'E' has answered two questions after the filter (**Q29 and Q30**) before skipping the remaining questions, whilst 'G' has answered **Q29-Q33**).

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

if Q28 = 3 or 4 then delete responses to:	Q29-Q33
--	---------

In accordance with this, all responses for Q29, Q30, Q31, Q32 and Q33 must be removed in cases where the respondent has ticked Q28=3 or 4 (i.e. not had a care review in the last 12 months). This will lead to two responses being removed for respondent 'E' (Q29 and Q30) and five responses being removed for respondent 'G' (Q29-Q33) who have for whatever reason followed the routing instructions incorrectly and continued to answer the section on care review meetings.

<u>Figure 3</u> (below) shows how the data would look following cleaning by the co-ordination centre to remove responses to filtered questions that should have been skipped – cells where responses have been removed are shaded.

Figure 3: Data from Figure 3 following cleaning

Record	Outcome	Q28	Q29	Q30	Q31	Q32	Q33
Patient Record Number	Outcome of sending questionnaire (N)	In the last 12 months have you had a care review meeting to discuss your care?	Were you told that you could bring a friend, relative or advocate to your care review meetings?	Before the review meeting, were you given a chance to talk to your co- ordinator about what would happen?	Were you given a chance to express your views at the meeting?	Did you find the care review helpful	Did you discuss whether you needed to continue using NHS mental health services?
A	6						
В	1	2	1	2	1	2	1
С	1	1	1	1	2		1
D	4						
E	1	3	•				
F	6						
G	1	4					
Н	1	1	3	2	2	3	1

5 Appendix B: Non-specific responses

The following table lists all 'non-specific responses' included in the 2013 Mental Health Service User survey. Numbers in the final column indicate the response options that should be considered non-specific. Where the 'non-specific responses' column contains only a dash, the relevant question has no such response options.

No.	Question	Non-specific responses
Q1	When was the last time you saw someone from the NHS mental health services?	6,7
Q2	Overall, how long have you been in contact with NHS mental health services?	6
Q3	Which of the following NHS healthcare workers or social care workers have you seen most recently for your mental health condition?	9
Q4	Did this person listen carefully to you?	-
Q5	Did this person take your views into account?	-
Q6	Did you have trust and confidence in this person?	-
Q7	Did this person treat you with respect and dignity	-
Q8	Were you given enough time to discuss your condition and treatment?	-
Q9	In the last 12 months, have you taken any prescribed medication for your mental health condition?	-
Q10	Do you think your views were taken into account in deciding which medication to take?	-
Q11	In the last 12 months, has any new medication (e.g. tablets, injections, liquid medicines, etc) been prescribed for you by an NHS mental health worker such as a psychiatrist or a community psychiatric nurse?	3
Q12	Were the purposes of the medication explained to you?	-
Q13	Were you told about possible side effects of the medication?	-
Q14	The last time you had a new medication prescribed for your mental health condition, were you given information about it in a way that was easy to understand?	-
Q15	Have you been on any prescribed medication for 12 months or longer for your mental health condition?	-
Q16	In the last 12 months, has an NHS mental health worker or social care worker checked with you about how you are getting on with your medication- i.e. have your medicines been reviewed?	-
Q17	In the last 12 months, have you received any of these sorts of talking therapies from NHS mental health services?	-
Q18	Did you find the NHS talking therapy you received in the last 12 months helpful?	4
Q19	Do you know who your Care Co-ordinator (or lead professional) is?	-
Q20	Can you contact your Care Co-ordinator (or lead professional) if you have a problem?	-
Q21	How well does your Care Co-ordinator (or lead professional) organise the care and services you need?	-
Q22	Do you understand what is in your NHS care plan?	4
Q23	Do you think your views were taken into account when deciding what was in your NHS care plan?	-
Q24	Does your NHS care plan set out your goals? It might include the changes you want to make to your life as your care progresses or the things you want to achieve	-

No.	Question	Non-specific responses
Q25	Have NHS mental health services helped you start achieving these goals?	-
Q26	Does your NHS care plan cover what you should do if you have a crisis (e.g. if you are not coping or if you may need to be admitted to a mental health ward)?	-
Q27	Have you been given (or offered) a written or printed copy of your NHS care plan?	4
Q28	In the last 12 months have you had a care review meeting to discuss your care?	4
Q29	Were you told that you could bring a friend, relative or advocate to your care review meetings?	3
Q30	Before the review meeting, were you given a chance to talk to your co- ordinator about what would happen?	3
Q31	Were you given a chance to express your views at the meeting?	-
Q32	Did you find the care review helpful?	-
Q33	Did you discuss whether you needed to continue using NHS mental health services?	-
Q34	Do you have the number of someone from your local NHS mental health service that you can phone out of office hours?	3
Q35	In the last 12 months, have you called this number?	-
Q36	The last time you called the number, did you get the help you wanted?	-
Q37	Has anyone in NHS mental health services ever asked you about your alcohol intake?	3
Q38	Has anyone in NHS mental health services ever asked you about your use of non-prescription drugs?	3
Q39	In the last 12 months, did anyone in mental health services ask you about any physical health needs you might have?	3
Q40	In the last 12 months, have you received support from anyone in NHS mental health services in getting help with your physical health needs?	4
Q41	In the last 12 months, have you received help from anyone in NHS mental health services in getting help with your care responsibilities (including looking after children)?	4,5
Q42	In the last 12 months, have you received support from anyone in NHS mental health services in getting help with finding or keeping work (e.g. being referred to an employment scheme)?	4,5
Q43	In the last 12 months, have you received support from anyone in NHS mental health services in getting help with finding and/or keeping your accommodation?	4
Q44	In the last 12 months, have you received support from anyone in NHS mental health services in getting help with financial advice or benefits (e.g. Housing Benefit, Income Support, Disability Allowance)?	4
Q45	Overall, how would you rate the care you have received from NHS mental health services in the last 12 months?	-
Q46	Have NHS mental health services involved a member of your family or someone else close to you, as much as you would like?	4,5
Q47	Who was the main person or people that filled in this questionnaire?	-
Q48	Are you male or female?	-
Q49	What was your year of birth?	-
Q50	In general, how is your mental health right now?	-
Q51	Have you been admitted to a hospital as a mental health patient in the last twelve months?	-

No.	Question	Non-specific responses
Q52	Are you currently in paid work	-
Q53	What is your religion?	-
Q54	Which of the following best describes how you think of yourself?	-
Q55	What is your ethnic group? (Tick ONE only)	-

.....