Understanding your free text data

This document provides guidance around two free text questions (Q21a and Q21b) that were included in the core 2020 NHS Staff Survey. It is designed to be used by NHS organisations looking for more advice and information about the questions, data coding and analysis, and how the results can be interpreted and used. It is divided into two sections. 1) guidance to help you understand your data 2) frequently asked questions. Further information on the analytical process has been included as a separate document. This has been produced by Hertzian, a company specialising in free text analysis.

1. Guidance to help you understand your data

1.1. Background

Every year, we carry out a national survey of NHS staff and whilst this past year has been very different for the NHS, we asked many of the same questions, in the same way, as in previous years. This was to ensure that we could measure the impact of the pandemic on the experiences of our staff by comparing it to previous years.

In recognition that this year has not been typical, changes were made so that we could understand our NHS people's experience of working through the COVID-19 pandemic more fully. We know that the last year will not have been the same for everyone; and it's important that at a national and local level, we can uncover the best practice that really made a difference to people, and also get a clear picture of what could have been done differently.

In 2020 NHS Staff Survey we included the following two free text questions.

- Thinking about your experiences of working through the COVID-19 pandemic, what lessons should be learned from this time?
- What worked well during COVID-19 and should be continued?

The questions were intentionally open-ended to allow our NHS people to tell us what was most important to them. The questions were asked in the Autumn of 2020 and provide a valuable resource to reflect on the achievements and learnings from the time and to inform recovery. They were not designed to measure and compare how well different organisations were performing but can be used to provide insight around the different experiences that staff had.

We are grateful to our NHS people for taking the time to feedback their experiences. In total we received c.700,000 comments across the two questions. c.360,000 made a comment at Q21a (60% of all respondents), and c.340,000 made a comment at Q21b (57% of all respondents). The average length of the comments was 46 words at Q21a and 21 words at Q21b. The comments added up to 23.3 million words in total.

1.2   Organising and presenting the responses

The number and size of the comments is too large for each comment to be individually coded by an analyst. Instead we adopted a machine learning approach where computer algorithms are used to decide what the data is telling us. The algorithms improve automatically through experience. Some manual checking takes place to ensure the coding is logical. Further details of the methodology used are available in the Hertzian report which has been provided separately.

Respondents' comments have been organised into the following:

- 13 'topics',
- over 150 'subtopics',
- 4 'attributes',
- and 'sentiment'.

The topic and subtopics tell you what a comment was referring to. For example, 'health and wellbeing' is a topic and 'morale' is a subtopic of that.   Attributes provide more context by telling you who is mentioned (such as people from a particular staff group) and where (a location such as hospital or ward); and the kinds of care being discussed (such as intensive care). Finally, the sentiment tells you whether the comment has a positive, negative, or neutral tone. Please note, many of our NHS people wrote about more than one thing in their response to the questions. In these cases, their response has been broken down into more than one comment.

1.3.   The reporting outputs

Your organisational outputs will include the following.

- The 2020 NHS Staff Survey Free Text Report for your organisation. This is a static PDF document that includes visual summary, in chart form, of the breakdown of all the free text comments received in response to Q21a and Q21b by topic/subtopic and sentiment (positive/neutral/negative).

- An interactive Excel workbook which details the organisation's full redacted free text comments alongside the topics, subtopics, sentiments, and attributes.

In addition, from 27 May you will be able to access the National 2020 NHS Staff Survey Free Text Commentary providing a summary overview of the results. This is a static PDF document that includes:

- background information about the free text questions and size of the response,
- an overview of the results including details of the most mentioned topics and who we heard from,
- an outline of the potential application of the findings,
- a deeper dive into specific topics of interest, and
- an appendix including a visual summary, in chart form, of the breakdown of all the free text comments received in response to Q21a and Q21b by topic/subtopic and sentiment (positive/neutral/negative).

Please note, the data has been organised at a national level as this provides efficiencies of scale. Firstly, one code frame has been developed in total rather than one per organisation which saves duplication, and secondly machine coding avoids the need for a team of analysts to read and code every comment in a consistent way.

1.4. Using the findings

The free text questions were designed to provide insight around the different experiences that staff had of working through the pandemic and their suggestions for improvements. This information adds vital context to the experiences that have been collected and measured within the NHS Staff Survey. The questions were not designed to measure and compare how well different organisations were performing. For this reason, the organisational results will not be published.

Your organisation's 2020 NHS Staff Survey Free Text Report provides numerical information around the topics and subtopics that people in your organisation mentioned in their answers to the free text questions. This tells you how often topics and subtopics were mentioned and whether the sentiment expressed was positive, negative, or neutral in tone.

The interactive Excel workbooks includes all the comments made. This is valuable for understanding the topics and subtopics, providing more detailed insight, and identifying opportunities for learning and improvement. We would recommend involving staff in the process of discussing and interpreting the findings as this will increase buy in and understanding.

The National 2020 NHS Staff Survey Free Text Commentary provides numerical information about the national picture. This includes results for all our NHS people as well as a deeper dive into specific topics. There is also information about interpreting and using the data.

There is very limited value in comparing numerical findings from your organisation to the national results as the context will vary by organisation and staff group. For

example, the experiences of our people working in COVID-19 wards in areas which were particularly badly hit are likely to be different to those working in different circumstances.

1.5 The benefits and limitations of the machine learning approach

The machine learning approach taken to organise the data offers users a way to find and analyse comments by key areas of interest. If you wished to explore an area of interest in more detail you could conduct a thematic analysis of relevant comments to identify the implications for your organisation.

Machine learning is useful for very large volumes of data which cannot be practically organised by analysts. We are confident the computer algorithms have performed well as they have been assessed through several iterations of improvement and multiple quality assurance checks. However, accuracy with these methods is never 100% and there may be examples where the machine has not made the same decision that an analyst would. Keep in mind that machine learning operates on a range of rules rather than an understanding of context. An example of this is that the topic 'continuity' mostly includes comments around continuity of care but also captures some comments around business continuity. The same is true of the process of automatically interpreting the sentiment of a comment whereby the machine's decision may not align always to the interpretation of an analyst.

1.6. Additional data cleaning and analysis

You may wish to re-code portions of your data when you come across examples where an analyst would disagree with the machine's allocation of a topic, sub-topic or sentiment. We encourage you to do so and would love to hear your feedback on how well you feel the machine learning approach has performed in terms of its accuracy. You can do this by sending your feedback to us at nhsi.staffsurveyengagement@nhs.net.

You may also want to cut the data in different ways to look at the comments by groups of people with different characteristics or opinions. For example, you could look at the comments made by colleagues with specific protected characteristics to ensure their views are being heard. To receive a cut of the data for your organisation please contact your external supplier. If you wish to do this on a national level please email nhsstaffsurvey@surveycoordination.com. Please note that the data files may be very large due to the large number of comments received.

If you do decide to conduct additional data cleaning and analysis we would really appreciate it if you could tell us what you have done and how it has worked so we can share learning across the NHS. Please could you email nhsi.staffsurveyengagement@nhs.net to tell us about your experiences or arrange a meeting to discuss this further.

2. Frequently asked questions

We have provided a list of questions and answers in the following table.

Question(s)	Response
How do the local and national outputs differ?	Each organisation has received a report on their numerical organisational results as well as all their redacted free text comments. These are not published. The National Free Text Commentary provides numerical information about the national picture. This includes results for all our NHS people and some specific audiences as well as a deeper dive into specific topics. There is also information about interpreting and using the data. This report will be published on 27 May.
What is the difference between the data provided on the free text questions and data from other (closed) staff survey questions? How valid are the results of free text questions compared to closed questions?	The free text responses provide a valuable resource to reflect on the achievements and learnings from the time and to inform recovery. The questions were not designed to measure and compare how well different organisations were performing but can be used to provide insight around the different experiences that staff had in different organisations. A key benefit of the free text questions is that our NHS people were able to comment on their own priorities instead of selecting an answer from a set of options. All the other core questions on the national staff survey are closed questions. These provide less nuance but work better for measuring the views and experiences of our NHS people. Data from the free text questions can be presented numerically and as a set of comments. The numerical data is statistically robust but needs to be interpreted with caution as the topics and sentiment scores are an interpretation of individual comments rather than directly selected options as you would find in a closed question. It is also worth noting that a machine learning approach involves coding decisions being made based on a set of rules rather than an understanding of the context. The full set of comments provided by the organisational level interactive Excel workbooks provide deeper insight around staff views and experiences.
How do I know if I have a good result?	The free text responses provide a valuable resource to reflect on the achievements and learnings from the time and to inform recovery. The questions were not designed to measure and compare how well different organisations were

	performing but can be used to provide insight around the different experiences that staff had in different organisations.
How do local free text questions sit with these questions?	Some organisations have opted for additional open questions. These are likely to have been edited and presented in a different way. Organisations will need to talk to their supplier for advice about understanding and analysing these questions.
How does machine learning work, and what are the advantages and disadvantages of this approach?	A machine learning approach is appropriate to a data set of this size. One benefit is that data from all the NHS organisations who take part in the NHS Staff Survey has been looked at using the same set of rules. On the other hand, machine learning cannot take account of the context in the way that an analyst can. Re-editing and coding the topics, subtopics, attributes and sentiment may support evidenced-based decision making as the analyst can make judgement calls that could be missed by machine learning and conduct the work with a specific research question in mind. On the other hand, this work can take considerable time with the large quantities of data involved so should be used appropriately.
The way a response has been coded seems wrong. How might that have happened and how should I respond to this?	Coding data is a matter of interpretation. In a data set of this size it is possible that the interpretation taken through the machine learning may occasionally not be intuitive. This is because the data is looked at using a set of rules and machine learning cannot take account of context and nuance in a way that an analyst can. On the other hand, coding conducted by an analyst can be subject to differences in interpretation or human error. You may wish to re-code portions of your data when you come across examples where an analyst would disagree with the machine's allocation of a topic, subtopic or sentiment. We encourage you to do so and would love to hear your feedback on how well you feel the machine learning approach has performed in terms of its accuracy. You can do this by sending your feedback to us at nhsi.staffsurveyengagement@nhs.net. If you have noticed a problem that affects a large quantity of data, please contact nhsi.staffsurveyengagement@nhs.net so that an assessment can be made around any potential implications.
Is the data I am receiving ready to use, or is further	The information organisations receive has already been quality-assured, edited, and coded.

work needed before it can be analysed?	It is possible to edit the data further to allow additional analysis. • You may wish to re-code all or some of the comments if you feel they could be categorised in a more useful way. • You could develop new ways of categorising the data to better reflect the comments, or meet your needs, • You could cut the data in different ways to look at the comments by groups of people with different characteristics or opinions (based on data from the NHS Staff Survey), Please note, organisations will need to ask their external supplier to cut the data for them as they will have access to the full data set. If you wish to do this on a national level please email nhsstaffsurvey@surveycoordination.com. If you do decide to edit the data further, we would really appreciate it if you could tell us what you have down and how it has worked so we can share learning across the NHS. Please could you email nhsi.staffsurveyengagement@nhs.net to tell us about your experiences or arrange a meeting to discuss this further?
Is it possible to look at the results in a specific region, STP or ICS?	Our NHS people's individual comments have been provided at an organisational level to ensure that the amount of information is manageable and that learning and improvement can take place locally. There may be a need to look at comments within a region, STP or ICS level, particularly for areas that are primarily being managed at this level. We would recommend investigating specific areas of interest rather than looking at every recorded comment due to the size of the data. If you wish to see regional, STP or ICS data, please contact nhsi.staffsurveyengagement@nhs.net.
How can I find out how responses differ by role or background?	If you wish to do this at an organisational level, please contact your supplier. If you wish to do this on a national level please email nhsi.staffsurveyengagement@nhs.net.
Why has the data taken so long to arrive?	This has occurred because of the size of the data set (23.3 million words) and the checking and quality assurance needed to ensure it is as useful as possible.
Is the data anonymised?	Respondents were asked not to include personal details in their comments. The survey contractors also went through a process of checking comments and redacting any personal

Who has access to the data?	data. The data is presented anonymously. Individual comments are not published but participating organisations will receive their own redacted comments in their interactive Excel workbook. The national dataset will not be published but will be available to researchers on request, subject to an official data sharing process.
	
	The guidance states that before passing any free text comments to a participating organisation the contractor must review all comments and remove any identifying information. At a minimum, contractors must remove the following data from all comments that are to be passed to a participating organisation:
	     • names,
	     • dates of birth,
	     • ID numbers (e.g. a payroll number),
	     • addresses,
	     • email addresses,
	     • phone numbers