

What does West Sussex worry about? Technical report for qualitative analysis methodology

Public Health and Social Research Unit

April 2026



Authors

Catherine Wells, Kirsty Bennett, Lizzie Vass, Matthew Dorey, Ruth Burns and Tim Martin.

Background

At a national level, there is some evidence that the COVID-19 pandemic impacted many aspects of health and wellbeing and widened health inequalities. In West Sussex, we wanted to understand health and health behaviours at a local level, following the disruption of the pandemic. We were particularly interested in understanding more about the health of people in the most disadvantaged areas. To do this, we conducted the West Sussex Your Health Matters (YHM) survey, using deprivation as the key criterion for our sampling strategy.

In the YHM survey, where possible, we used validated tools to explore topics spanning general health, health behaviours and the wider determinants of health. We also included a small number of free-text questions, which asked respondents to list their three main concerns and worries; what would help them to be more physically active; and where they would seek help or support for mental health and wellbeing.

This report outlines our methodological approach to analysis of the free-text responses to the ‘three main concerns and worries’ question. An overview of the sampling strategy and data collection methods for the full survey is provided for context, following by an in-depth description of our approach to analysis.

Methodology

This survey used a mixed-methods design to gain further information on health and health inequalities across the county, through postal, online and face-to-face surveying.

The research team and working group met regularly to develop the methodology and the sampling framework for the survey. Question topics were discussed with Public Health leads and refined to produce a draft survey questionnaire. Once this was developed and finalised, it was approved by the Public Health leadership team. Progress with project was monitored by the working group through weekly meetings.

Sample

A market research company, QaResearch, was contracted to undertake the data collection. Residential postal addresses in the county are split into five groups (called ‘quintiles’), according to deprivation, as defined by the Index of Multiple Deprivation (IMD; 2019). The IMD measures how deprived one area is compared with another. The aim was to have equal samples across the deprivation quintiles¹.

To achieve this, a total of 31,500 survey packs were sent out to residential addresses across West Sussex. This number was based on QaResearch’s recommendation that a 12% response rate (equivalent to 3,750 surveys) was a realistic and feasible expectation for this type of survey. A lower response rate among the most deprived quintile was anticipated, therefore, to achieve equal samples across deprivation quintiles, quintile 1 was oversampled. A random stratified sample was used to select the required volume of addresses from the Postcode Address File (PAF)² for each IMD quintile. Table 1 shows the number of postal surveys distributed by deprivation quintile.

Table 1: Number of postal surveys distributed by deprivation quintile

Deprivation quintile	Number of postal surveys distributed
Quintile 1 (most deprived)	8,000
Quintile 2	5,875
Quintile 3	5,875
Quintile 4	5,875
Quintile 5 (least deprived)	5,875

¹ NB: the population in West Sussex is not evenly split between the IMD quintiles. Relatively few people live in the most deprived areas; around eight-times as many people live in the least deprived quintile as the most deprived quintile.

² The Postcode Address File (PAF) is maintained by Royal Mail and is the most accurate, up-to-date and complete address database in the UK, containing 32 million addresses.

Data collection

Market research panel

To gather an initial subset of survey completions in a quick and cost-effective way, and to examine any issues with questions, a commercial access panel was recruited with existing panel members who had already consented to participate in research. The online survey did not involve setting any demographic quotas and in the final sample certain demographics (including IMD quintile) were over or under-represented. The market research panel survey was open 16-27th January 2024.

Postal survey

Postal surveys were the primary method for this project. Each survey pack contained a survey and covering letter which included a link and QR code so residents could submit their responses online, if preferred. To maximise the response rate, a prize draw was offered as an incentive with prizes of £750, £150 and £100. Residents could opt-in to the draw or decline participation. Surveys were sent on 1st February 2024, with returns accepted until 6th March 2024.

Face-to-face survey

To ensure the final sample for each quintile matched the demographic profile of West Sussex residents living in each quintile as far as possible, a follow-up face-to-face 'top up' survey of residents from under-sampled demographic groups (residents in quintiles 1-4, those aged 16-24 and 25-36 and males) was carried out.

Regardless of the survey method, the questionnaire took an average of 15 minutes to complete.

Survey response rate

Across the above data collection modes (excluding the open access online survey), 6,440 respondents completed the survey. Response rate by method of survey completion and deprivation quintile is shown in Table 2.

Table 2: Final sample achieved, by completion route and deprivation quintile

Deprivation quintile	Market research panel	Invited and returned by post	Invited by post and completed online	Face-to-face boosted survey	Total
1 (most deprived)	14	961	193	55	1,235
2	71	789	165	49	1,146
3	118	927	175	83	1,468
4	90	1,023	200	39	1,510
5 (least deprived)	108	1,099	253	28	1,681
Total	401	4,799	986	254	6,440

Worries and concerns item

The YHM survey included a small number of free-text questions. One question, and the focus of this report, asked about respondents top three worries and concerns:

At the moment, what would you say are your three main concerns or worries? These concerns don't have to be about your health, we are interested in what's important to you and causing concern.

Respondents were able to provide up to three responses, or to select 'Nothing' or 'Prefer not to say'. This question was situated approximately halfway through the survey, within a section of closed questions asking about mental health, loneliness and sleep. Previous questions included closed questions about the respondent's household and home, general health, alcohol, and smoking and vaping. Although this question

stated that concerns did not have to be health-related, it is possible that the health focus of the survey may have primed respondents towards thinking, at least in the first instance, of health-related worries.

Response rate to worries and concerns question

Of the 6,440 individuals who completed the survey, 15.3% (n=987) responded 'Nothing' and 5.1% (n=331) responded 'Prefer not to say'. A further 2% of respondents did not leave a response (n=129).

The number of respondents to the worries and concerns question are shown in Table 3. In total, 4,993 respondents left at least one free-text response (77.5%), totalling 13,193 responses. A small number of these responses were excluded because they were illegible or incomplete (n=97), leaving 4,986 respondents who provided at least one worry or concern (77.4%), and 13,096 free text-responses included in the analyses. A large proportion of respondents who provided at least one worry, provided multiple worries: 4,473 (69.5%) of all survey respondents provided two worries and 3,637 (56.5%) provided three worries.

Table 3: Number of respondents (and proportion of overall number of respondents)

All respondents n (%)				Final number of respondents included in the analysis* n (%)		
Responded 'Nothing'	Responded 'Prefer not to say'	No response	Provided at least one response	Provided one worry	Provided two worries	Provided three worries
987 (15.3%)	331 (5.1%)	129 (2.0%)	4,993 (77.5%)	4,986 (77.4%)	4,473 (69.5%)	3,637 (56.5%)

*Responses which were illegible or incomplete were excluded from the analyses

Research team

The research team consisted of eight public health researchers/analysts, five female and three male, across a range of ages. We are all permanently employed in salaried local government roles and come from a range of health-related backgrounds including public health, social care, health psychology and statistical modelling. Some members of the research team identify as mixed-methods researchers, with others identifying predominantly as quantitative researchers. We have experience across a range of public health topics including mental health, housing, tobacco control, ageing, child health, social care and sexual health. Our identities and experiences shape how we approached this research and to address potential biases, we engaged in rigorous and reflexive practices which are detailed later in this report.

Data analysis

Data were analysed using thematic analysis.

Due to the large size of the dataset and the short, concise nature of responses, we used a Microsoft Excel workbook to organise our data and structure our coding. An inductive and iterative approach was used to develop a thematic codebook, grounding the identified themes in the data and allowing continual refinement of these themes as the analysis progressed.

Analytic approach in brief:

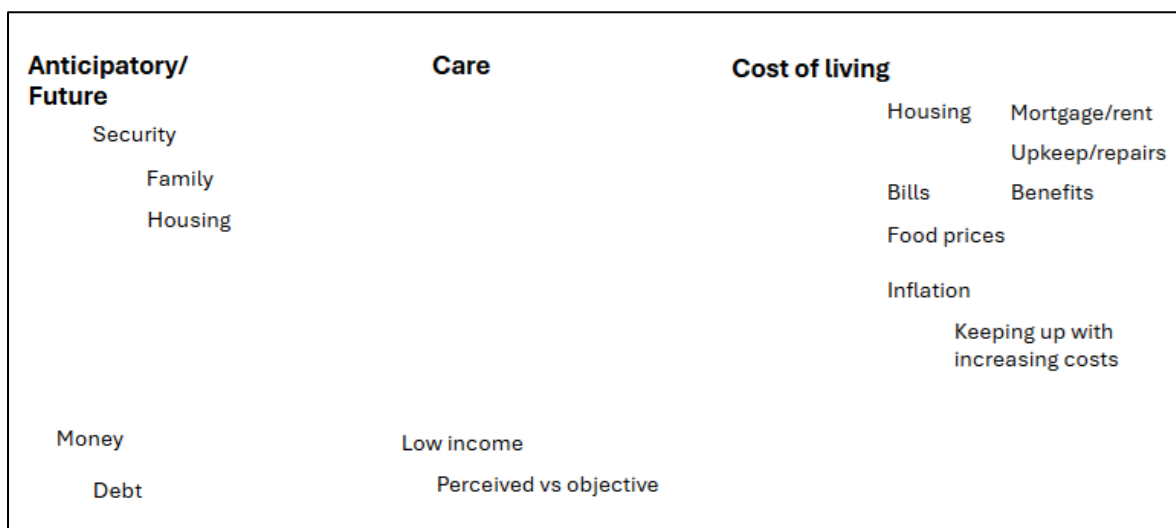
1. Developing a draft coding frame of themes and subthemes using a small sample of the full dataset.
2. Testing this coding frame against a larger sample of the full dataset and refining as necessary.
3. Applying the revised coding frame to the full dataset.
4. Checking inter-rater reliability
5. Narratively summarising thematic themes and sub-themes
6. Quantitatively analysing prevalence of themes/sub-themes across respondents by key characteristics (e.g., deprivation)

Approach

1. Developing the initial coding frame

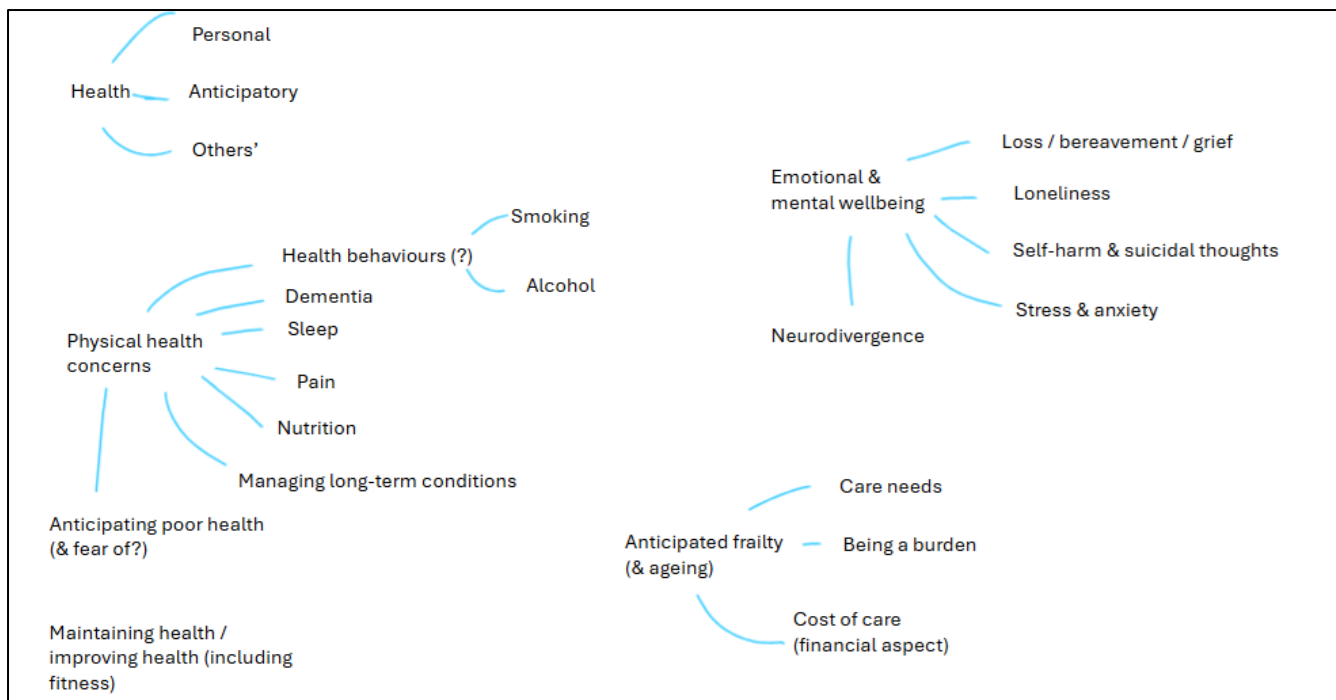
- We took an initial random sample of 150 respondents from the full dataset. This sample comprised 386 responses (as some respondents provided more than one response).
- Throughout the coding process, where a respondent had left more than one response, all responses were kept together and reviewed by the same researcher. While most responses could be coded in isolation, in some instances, respondent's other responses provided useful contextual information as to how the response should be coded. While coding, we did not have access to demographic information (such as age and gender) to ensure we remained impartial and were not influenced by any existing assumptions.
- We individually familiarised ourselves with this sample dataset and noted initial thoughts and ideas of topics that respondents had repeatedly raised.
- We then met to review this sample dataset and to collectively generate initial broad themes that described the most frequently raised worries and concerns. This was an iterative process until we were satisfied that all responses were captured by the themes. The initial 386 responses were grouped into five broad themes: 1) Finance, 2) Health, 3) Service access, 4) Environment, and 5) Geopolitical, which were used to structure further coding of the data.
- For the Finance theme, we, as a group, systematically reviewed each response in the initial sample and inductively generated codes describing any worries that seemed to fit within this initial thematic area. Our initial codes for the Finance theme are shown in Figure 1.

Figure 1: Initial generated codes for the Finance theme.



- For the four other initial broad themes, we individually reviewed the sample dataset, with each researcher following a similar coding process as used for the Finance broad thematic area to each develop a thematic map. An example of a thematic map developed by one researcher for the 'Health' theme is shown in Figure 2.

Figure 2: Thematic map for the 'Health' theme



- We then met several times, as a group, to discuss each researchers' interpretation of the dataset and to develop an initial coding frame for each broad thematic area. Throughout this process, we continually reviewed our initial themes and sub-themes. Where there was significant overlap in themes/sub-themes, or areas of the data that we felt were not described in the existing themes/sub-themes, we refined these to better delineate or join concepts, or to create new sub-themes or themes.
- We 'tested' our initial coding frame against the sample dataset and began to develop a codebook in Microsoft Excel, which provided a definition of each theme/sub-theme and inclusion and exclusion criteria. We began this process by testing and developing a codebook for one of the largest broad thematic areas (Health), initially testing the coding frame as a group and then continuing testing and codebook development individually. We then met to discuss and agree a complete draft codebook for this broad thematic area. For the remaining themes, we individually tested the coding frame and began codebook development and then met to review and agree a complete draft codebook.

2. Testing the draft coding frame

- Once we had agreed an initial coding frame and codebook that described the full sample dataset, we again 'tested' this against a new, larger randomly selected sample of respondents. Each researcher was assigned a previously uncoded random sample of 150 respondents from the full dataset, in addition to the initial 150 respondent sample. With seven researchers in the initial research team, this equated to a further 1050 respondents. Once individual researchers had coded the responses in their individual sample dataset, they reviewed the responses they had assigned to each sub-theme to check the internal homogeneity of responses within sub-themes and broad themes (that responses within each theme and sub-theme were similar to each other) and the external heterogeneity between sub-themes and broad themes (that there was a clear distinction between themes and sub-themes).
- Following individual 'testing', we met several times to discuss the fit of the draft codebook against the larger sample. We agreed that, although we had reached data saturation with the initial 150 respondent sample, this sample was not entirely representative of the full dataset. Throughout this process, we developed several new themes/sub-themes to address the identified gaps in the draft codebook, repeated individual testing of the codebook against the larger sample, and further refined the codebook.

3. Applying the codebook to the full dataset

- Upon completion of a refined codebook, the 'coding' of the most frequent one- or two-word responses (e.g. 'health', 'money', 'family') was automated (using RStudio). Automation of such responses

included variants of each one- or two-word response, to account for spelling errors, plurals and synonyms.

- To apply the refined codebook to the full dataset, each researcher was assigned a further randomly selected portion of the full dataset and coded the responses using a standardised excel spreadsheet with columns for each theme and sub-theme. Each researcher was assigned the responses of approximately 625 respondents. NB: At this stage, an additional researcher joined the research team.
- Generally, most responses could be coded as single, isolated responses. However, where the meaning of a response was unclear or ambiguous, researchers considered the response in the context of any other responses the respondent had made.
- Throughout the process of applying the refined codebook to the full dataset, the research team met regularly to discuss difficulties or uncertainties in how some responses should be placed, and to clarify any remaining ambiguities in the codebook. Group discussion and review identified some areas where researchers felt there may have been inconsistencies in how the codebook had been applied to the full dataset, in some cases due to conceptual overlap of some themes/sub-themes in the codebook.
- Where some small changes to the codebook were agreed, researchers individually made revisions to their subset of the data.

4. Checking inter-rater reliability

- To check the consistency of coding by individual researchers, several themes/sub-themes were selected for review, to check internal homogeneity and external heterogeneity. The themes/sub-themes that had stimulated the most discussion or difficulties in coding were selected for review (including 'life changes', 'living with long-term health conditions', 'emotional wellbeing and mental health', 'giving and receiving care', 'ageing', 'future health and treatment', 'independence', the housing sub-themes) and the 'miscellaneous' theme.
- For each theme/sub-theme selected, two researchers reviewed and discussed the responses coded to the themes. Where the researchers felt responses had not been coded according to the codebook or should have been coded differently, these were flagged for group discussion and coding changes suggested.
- We met as a group to discuss the consistency of coding within the reviewed themes and agree any necessary revisions. Where necessary, additional themes and sub-themes were created (e.g. a new theme of 'coping' was created for non-specific worries about the ability to cope and 'perimenopause and menopause' and 'pregnancy' were created as separate sub-themes of the theme 'life changes and transitions').
- Once revisions had been agreed by the research team, the coding of responses was amended. Where a response had been amended, this was noted in a separate column in the Excel workbook.

5. Narratively summarising thematic themes and sub-themes

- Following the inter-rater reliability review and subsequent revisions made to the coding of responses, the research team narratively summarised each high-level theme and sub-theme.
- For each theme and sub-theme, we provided the number of respondents (rather than the number of responses) who had left a comment relating to that theme as we felt this provided a more accurate representation of the prevalence of a theme. While thematic analysis does not aim to quantify responses, we felt it was useful to provide information on how frequently a theme was mentioned.
- For each theme and sub-theme a member of the research team reviewed all the responses in that theme. The range of responses within each theme and sub-theme were described in a narrative summary, along with illustrative quotes.

6. Quantitatively analysing prevalence of themes/sub-themes across respondents by key characteristics (e.g., deprivation)

- The prevalence of themes and subthemes were analysed quantitatively using R. Uncoded responses and those that were illegible or incomplete were excluded. The demographic characteristics of

respondents and the frequency of respondents and responses across themes and subthemes were summarised. These analyses [are described elsewhere](#).

Rigour and reflexivity

A number of factors are thought to define rigorous qualitative research, including transparency, validity or credibility, reliability and dependability and reflexivity³. Transparency refers to thoroughly describing the research process. During the analyses we documented in detail the steps that we took and the decisions we made. These are comprehensively summarised in this document. By being transparent about our procedures and processes, if other researchers wanted to try and replicate the research, we are confident that they would come to similar conclusions.

A valid and credible study requires data to be represented fairly and accurately. When narratively summarising each theme and sub-theme we have presented the full range of responses, regardless of how frequently they were mentioned. To increase the reliability of the study, multiple coders coded the initial subsets of the sample, with similar themes and sub-themes generated. We also met regularly to discuss the coding of responses to ensure that interpretations were aligned across coders. Furthermore, to check the consistency of coding between researchers, as detailed above, inter-rater reliability was carried out for several themes and sub-themes.

We acknowledge that our backgrounds and experience may have shaped this research and that our assumptions and values may have influenced the analyses and interpretation. To address this, we revisited codes iteratively, remained attentive to disconfirming data, and discussed the identified themes as a team to ensure that interpretations were grounded in participants' accounts rather than preconceived assumptions.

Strengths and limitations

This analysis benefitted from a large sample size and a rigorous, reflexive process. A team of researchers collectively developed the approach and undertook the analysis, with regular peer discussion, debriefing and review. Throughout each stage of data analysis, a thorough record of the development of the analytical approach, decisions and accompanying rationale was kept, alongside researcher reflections on the dataset and process.

However, despite the collective development of the codebook and in-depth discussion of all themes, the relatively large size of both the research team and the dataset itself meant that there remained differences in interpretation of the data between researchers. We performed inter-rater reliability testing (as detailed above). This returned mixed results, with some relatively large discrepancies between coders in some sub-themes, although this may be partially attributed to our selection of the most discussed and divisive themes for inter-rater reliability review. Changes were made in coding, following inter-rater review, although there remained inconsistencies in coding, which were revealed during the subsequent narrative summary of the dataset.

There were also some changes in the research team during the analysis, including two members leaving the team and one other joining. Whilst the initial codebook and framework was developed by the same core team, changes in the research team during later stages of the analysis may have led to inconsistencies in coding. There were also some changes made to the codebook after the full dataset had been coded, which may not have been completed by all members of the research team (particularly where researchers left the team).

For some themes and sub-themes, there is a remaining conceptual overlap. This meant that some responses could be 'double-coded' to multiple themes, where it was not clearly apparent which of similar themes they fit in. This may be attributed to the ambiguity inherent in short, concise responses and the discursiveness across responses that were contained in a very large dataset, alongside the inter-related nature of many of the issues discussed in the responses. Where there was ambiguity in a response, the research team were careful to avoid making assumptions and discussed in depth how to approach such responses. However, despite a no-

³ [Sage Research Methods - The SAGE Encyclopedia of Qualitative Research Methods - Rigor in Qualitative Research](#)

assumptions principle during coding, some level of assumption was inevitable, simply due to the human nature of qualitative research.

The Your Health Matters Survey was a health survey and it is possible that because of this respondents may have been more likely to mention a health-related worry or concern. While Health was the most frequently occurring theme, mentioned by 61.2% of respondents, a number of other themes were identified and several of these were frequently mentioned such as Financial (mentioned by 44.1% of respondents) and Family, relationships and social connectedness (mentioned by 31.4% of respondents).

In terms of the population representativeness of the subset of respondents providing at least one worry or concern, the demographic characteristics of these respondents were similar to the characteristics of respondents to the survey overall.