Comprehension and use of UK nutrition signpost labelling schemes

Scientific rationale and design

September 2008

Prepared by:

BMRB Social Research with input from the Food, Consumer Behaviour and Health Research Centre at the University of Surrey

Prepared for:

Food Standards Agency
## Table of Contents

Introduction ................................................................. 1  
   Detailed objectives and scope of study .............................. 1  
   Achieving a robust research design .................................. 2  
   Factors influencing research design ................................. 4  
   Arrangement of this document ........................................... 6  

1 Executive Summary ...................................................... 7  
   1.1 Introduction ............................................................. 7  
   1.2 Label format ............................................................ 8  
   1.3 Context of label presentation ...................................... 9  
   1.4 Measures of objective understanding ........................... 10  
   1.5 The testing design ................................................... 11  
   1.6 Other factors influencing comprehension .................... 13  
   1.7 Data collection and sampling method ......................... 13  

2 Elements of FOP nutrition signpost label - formats ............. 15  
   2.1 Label format: content and presentation ....................... 15  
   2.2 Background to content-related elements ..................... 17  
      2.2.1 Nutrient levels ...................................................... 17  
      2.2.2 Traffic light (TL) signposting ................................ 17  
      2.2.3 Interpretive text-based signposting ....................... 17  
      2.2.4 % Guideline daily amount (GDA) signposting ........... 17  
      2.2.5 Energy ............................................................. 18  
   2.3 Evidence on the performance of schemes using the different content-related elements ........................................ 18  
      2.3.1 % GDA and TL colour .......................................... 18  
      2.3.2 Interpretive text .................................................. 20  
      2.3.3 Energy ............................................................. 20  
   2.4 Evidence on the performance of schemes using the different presentation-related elements ................................. 21  
      2.4.1 Direction of presentation ...................................... 21  
      2.4.2 Colour used to draw attention to particular parts of the label .................................................... 21  
      2.4.3 The order in which nutrients and other information appears ...................................................... 22  
      2.4.4 Shapes within label ................................................. 22  
      2.4.5 Position of information within FOP label ................ 23  
   2.5 Implication for inclusion of FOP label elements in study design .... 24  
      2.5.1 Content issues in the research design .................... 25  
      2.5.2 Presentation issues in the research design ............... 26  
   2.6 Summary of priorities for format elements .................... 28  

3 Elements of FOP nutrition signpost label - context ............ 29  
   3.1 FOP nutrition signpost label presentation .................... 29  
      3.1.1 Summary of label presentation decision .................. 32  
   3.2 Product category ................................................... 32  
      3.2.1 Evidence .......................................................... 32
7.2.3 Sample selection ................................................................. 70
7.2.4 Maximising response rates .................................................. 70
7.3 The questionnaire and questionnaire development .................. 71
  7.3.1 Field piloting ................................................................. 72
  7.3.2 Cognitive testing ............................................................. 72
7.4 Fieldwork ........................................................................... 73
7.5 Data processing ..................................................................... 73
7.6 Ethical considerations ............................................................. 74
7.7 Summary of data collection and sampling method .................... 74
8 Appendix ............................................................................... 76
  8.1 Appendix A. Evaluation of tests used in previous studies .......... 76
  8.2 Appendix B: Final product and nutrient allocation for tests ...... 85
  8.3 Appendix C – criteria for signposting ..................................... 88
9 References ............................................................................. 89
10 Annex 1 – Initial results of phase 1 qualitative work .................. 94
11 Annex 2 – Results of cognitive testing stage ............................. 95

Copyright: survey findings and deliverables are normally intended for use within the Client's organisation or its consultants and other associate organisations such as advertising agencies. Should the Client intend wider circulation of the survey findings and deliverables, the Client should inform BMRB prior to such disclosure and agree the form and content with BMRB. The client should acknowledge BMRB as the source of the information with wording acceptable to BMRB.
Introduction

In 2008 the Food Standards Agency (FSA) commissioned a programme of research to evaluate the impact of front of pack (FOP) nutrition signpost labelling schemes on purchasing behaviour and consumer knowledge. This evaluation was intended to address two main questions:

1. How well do individual signpost schemes enable consumers to correctly interpret levels of key nutrients? While the impact of, e.g. time constraints, on comprehension should be considered in this part of the research, it will not involve testing comprehension in real life contexts.

2. How do consumers use front of pack labels in the retail environment and at home? The aim of this part of the research is to explore use in real life contexts.

This document concerns the research designed to address first of these two aims. It contains a discussion of the evidence drawn upon, explains the thinking behind the decisions made and provides a rationale for the final design of the study.

Insights from the qualitative work (conducted to address the second aim) have been used to inform the design of the quantitative research undertaken to address the first aim (described in this document). Details of the qualitative research are available in an annex appended to this report.

Detailed objectives and scope of study

The required outcome of the quantitative research is to determine what scheme(s), or what combination of elements of the different schemes, best facilitates the accurate interpretation of key nutritional information by consumers such that they are enabled to make informed decisions about the foods they consume.

The three main types of FOP labelling scheme to are:

1. Monochrome schemes providing information on percentage of Guideline Daily Amount (GDA);

2. Traffic light (TL) colour coded schemes indicating nutrient level;

3. Schemes which provide both a traffic light colour code and percentage of GDA.

More specifically the research objectives are:

1. To objectively assess the extent to which individuals are able to correctly interpret the nutritional information given on front of pack labels

2. To compare the comprehensibility of the three main formats

3. To identify the characteristics of a successful scheme, that is, one that enables consumers to make informed choices in relation to fat, saturated fat, salt, sugars and calories (where provided)

4. To investigate the impact of various social factors (e.g. socio-economic status, educational attainment, gender, ethnicity) on ability to interpret the information presented

In order to meet the above objectives, the study needs to focus on **objective understanding**, as distinct from subjective understanding and preference. Since subjective understanding and preference have been widely tested in earlier work (Grunert and Wills (2007), a small amount of additional information on this also needs to be collected in this study, to allow exploration of the relationship between subjective and objective understanding and to provide a basis for comparison with earlier research.

**Achieving a robust research design**

The research needs to produce data that is sufficiently robust to have scientific credibility. This means identifying specific research hypotheses to address, and designing both the tools and a robust data collection method to capture the information needed to address these hypotheses.

**The independent variables**

In order to do this we first need to **identify the various elements of the FOP labelling schemes that could influence comprehension**. These equate to the ‘characteristics’ of the FOP labels in the research aims above. In terms of hypothesis testing, these are the ‘independent’ variables and are discussed in Chapters 2 and 3.

The most systematic way of looking at the effects of these elements on comprehension is within a ‘full-factorial design’. This means testing comprehension of FOP labels which cover all possible combinations of the elements of front of pack labels in the design. This allows examination of both
main effects of each individual element on comprehension\(^2\) and interaction effects between different elements on comprehension\(^3\). So, for example, a full-factorial design could assess the effect on comprehension of just % GDA or Traffic Lights, and also any effect of interaction between the two.

The use of a full-factorial design places a limit on the number of elements that can be included (see section 2.1) so some prioritisation is needed; any elements not included as part of the full-factorial design need to be held constant on all labels shown, as any variation could influence the findings and risk making them unclear or unreliable. For example, if you did not wish to include the order in which the nutrients were presented as part of the full factorial design, but you then varied the order between two different labels shown, you would not know if any differences in comprehension were the result of the elements chosen as part of the full factorial design, or the order of the nutrients.

**The dependent variables**

Once the priority label elements are identified, we also need to develop a way of measuring how accurately people are interpreting key nutritional information (this produces the ‘dependent variables’ of comprehension (Chapter 4)). These final measures will be referred to as the ‘tests’ in this document.

**The design**

Having determined the label elements and measures for inclusion, we then need to put together a design which will allow us to test the hypotheses (see Chapter 5). Within this design hypotheses could be tested either ‘within’ or ‘between’ respondents. ‘Within’ respondent means the same person gets asked all the questions needed to test a hypothesis. This means you can be confident that any differences are real (i.e. that a person finds it easier to understand one type of label than for another). This is clearly the ideal approach. However, it is often the case that there are too many questions to ask each individual (the questionnaire would become over-long and the burden on the respondent would be too great). Instead a ‘between respondents’ design can be used, where you ask one set of questions of one group of respondents, and another set of questions of a separate group of respondents. In order to test the hypothesis you need to compare the answers to the two sets of questions. To be confident that there is a real difference you need to be confident that the two sets of people are fundamentally the same. As long as the sampling approach is robust this is not a problem. However, it does mean that hypotheses may need to be prioritised, with those

---

\(^2\) ‘the effect of a single independent variable on one or more dependent variables’ (Grimm & Yarnold, 1995)

\(^3\) ‘an interaction occurs when the effect of an independent variable on some dependent variable depends on the level of another independent variable’ (Grimm & Yarnold, 1995).
that are key being tested ‘within respondent’, and those that are secondary addressed ‘between respondents’.

In addition we need to:

- Identify any other factors that could influence comprehension (e.g. social and demographic factors – Chapter 6);
- Design a data collection method and tools to administer the research design and collect information on other factors that could influence comprehension (Chapter 7).

Factors influencing research design

Beyond the need for robustness, there are a number of other factors influencing the choice of research design. These include addressing the more detailed objectives of the research (as described above), taking into account evidence from previous research and other information available, and balancing the ideal design with considerations of feasibility and practicality. In addition the design needs, of course, to meet the usual ethical standards for the conduct of research. These factors are discussed below in more detail.

i. Evidence and information available

The sources of evidence considered are:

1. Learnings from previous work: A review of the relevant scientific literature was conducted, focused on recently published reviews (European Heart Network, 2003; Cowburn & Stockley, 2005; Grunert & Wills, 2007) and the studies they covered, together with any relevant studies published more recently.

2. Discussion with relevant experts: Five of the authors of relevant studies were consulted about the types of tests and other design considerations.

3. Findings from phase 1 of the qualitative stage of this research (see annex): This phase focussed primarily on actual use of FOP labels in the market place and in home, but it also generated some valuable data on comprehension issues.

4. Discussion with key actors in the project: This includes the Project Management Panel (PMP), the contractors (BMRB and the University of Surrey) and relevant officials from the Food Standards Agency (FSA). In addition, the views of the Advisory Group and the Nutrition Strategy Steering Group (NSSG), on which key external stakeholders of the project are represented, were also sought.
5. External peer review of the draft scientific rationale (following which modifications to the research design and revisions to this document were made).

**ii. Feasibility and practicality**

There are a number of practical considerations that limit the scope of this research including:

1. The number and nature of hypotheses that can feasibly be addressed in the study, given budget and timetable restrictions.

2. The type, complexity and duration of tests that it is feasible and reasonable to expect respondents to endure.

3. A balance between isolating the effect of different elements of the labels on comprehension, and ‘real life’ considerations about how labels are currently used in the marketplace so that the results can be applied to real life situations.

In relation to the last of these points, it is important to note that this aspect of the research is not intended to test use in real life contexts (e.g. at point of purchase), or to assess to what extent labels motivate consumers to read them. Furthermore, it is not intended to determine how well interpretation of the information on the labels was translated into personal choice (e.g. a purchase decision). In the context of an interview, levels of motivation and concentration are likely to be higher than in a real life context, which means the results are likely to reflect the maximum possible level of comprehension, rather than levels that may be achieved in real life situations. Furthermore, whilst (where possible) the design will reflect real life use of labels (and predominance in the marketplace), design decisions may be made that prioritise the theoretical above real life situations. This is to ensure that the objectives can be fully addressed. For example, the research needs to isolate the impact of the key characteristics (elements) of front of pack labels, which may mean removing other packaging information to avoid distraction effects. The parallel qualitative study will help to place the results of this study into context.

Addressing all of these considerations means establishing priorities. Final decisions reflect a balance of priorities for the different elements of the research design to achieve a design which can test robustly the hypotheses identified as key within reasonable time and resource constraints.
iii. Ethical considerations

There are always ethical considerations for any research project. BMRB (who will be carrying out the data collection) is an MRS Company Partner, which commits the company to upholding industry standards and complying with self-regulation.

Key ethical considerations for this project include:

1. Obtaining informed consent
2. Data confidentiality
3. Avoiding respondent burden (i.e. restricting the amount of time the interview takes, and the number of tests presented to each individuals)
4. Avoiding causing stress or anxiety by presenting tests.

Arrangement of this document

This introduction is followed by a summary of the design decisions taken. The main body of the document contains a more detailed discussion of the evidence available, the issues taken into account and the rationale for the final design decisions.
1 Executive Summary

1.1 Introduction

The overall aim of this study is to establish which front of pack (FOP) labelling scheme(s), or which elements of such schemes, best facilitate the accurate interpretation of key nutritional information by consumers, such that they are enabled to make informed choices about the foods they purchase. The document sets out the evidence and arguments which informed the design of the quantitative stage of the project. This part of the study is not intended to explore use in any real life context, as this is covered in the accompanying qualitative stage of the study.

The key challenge for this study is to achieve a robust research design. There are a number of issues considered in determining the choice of design, including:

- The objectives of the study (see introduction for details)
- The evidence available, covering:
  - Previous work in relevant fields (including nutrition, psychology, consumer science etc.) from published papers;
  - Discussion with relevant experts in the field;
  - A tailored analysis of phase one of the qualitative element of the research; and
  - Discussion with the key actors in the project (including a steering group of key externals stakeholders).
  - External peer-review
- Issues of practicality and feasibility, covering:
  - The practical limitations of administering such a study; and
  - Balancing real life considerations with the need for scientific rigour. This included consideration of marketplace issues, and in particular, the predominance of a number of scheme types in the market place; and
- Ethical considerations such as burden, confidentiality and consent.

Decisions are needed to establish priorities for three key aspects of the design:

- Label format (both content and presentation) – Chapter 2
- Label context (including product category) – Chapter 3
- Measures of objective understanding – Chapter 4

These then need to be combined into a workable test design (Chapter 5). An assessment of whether there are other factors influencing comprehension is needed to fully address the research objectives, (Chapter 6). Finally, appropriate and robust data collection and sampling methods and tools need to be designed (Chapter 7).
1.2 Label format

Three content-related elements are key to the stated research objectives, and the evidence suggests they are most likely to influence comprehension of the nutritional information provided on FOP labels. These are:

1. % GDA / no % GDA signposting
2. Traffic Light (TL) signposting / no TL signposting
3. Interpretive text (high, medium, low) / no interpretive text (referred to as ‘text’ throughout the report)

A full-factorial design is recommended for these three elements. This means testing labels showing all eight possible combinations of the three elements. This allows examination of the impact of each individual element on comprehension, and also any interactions between them.

Energy (also referred to as calories) has been identified as a secondary priority. The qualitative work revealed energy to be used as a proxy for other nutrients for judging the healthiness of products. Energy will only be included as part of the full-factorial design in tests that involve judging the healthiness of a product, otherwise energy will be held constant and present (as it is present on most labels in the marketplace).

Two presentational elements were identified in the qualitative work and previous work as having a potential impact on comprehension. These are the use of a circular presentation format (e.g. the Sainsbury wheel) and the use of non-TL colour to differentiate between nutrients (e.g. the Tesco pastel coloured label). Both have a well established position in the marketplace.

Inclusion of direction and non-TL colour in the full factorial design would increase the number of label combinations to be tested to 48, which is too many to include within the practical constraints of the research. Instead we decided to include just two further labels, approximating those used by Sainsbury and Tesco. These can each be compared with one of the eight labels in the full factorial design, differing from that label by only one element (direction/colour). Comparing comprehension on these FOP labels will allow limited evaluation of the impact of a circular presentation and of non-TL colour, but only in terms of the way they currently appear in the marketplace. They are not part of the full-factorial design, and so it is not possible to look at interactions of direction or of non-TL colour with %GDA, TL or text.

All other presentational elements is to be held constant as there is no evidence from the literature or the phase 1 qualitative work that these affect comprehension (including direction of presentation, shape, colour, order of
nutrients, positioning of text). The most common current label format in marketplace will be used to inform the decision over format.

**Hypotheses for label format**

- There are individual and interactive impacts on comprehension for the presence or not of %GDA, TL and text
- The presence or not of energy impacts on the accuracy judgements of product healthiness
- The use of a circular presentation impacts on comprehension (only tested for the format in which it is currently used in the marketplace)
- The use of non-TL colour to differentiate nutrients impacts on comprehension (only tested for the format in which it is currently used in the marketplace)

**1.3 Context of label presentation**

The qualitative work showed that a wide variety of the other information on the packets is used to inform judgements about products, such as picture, health claims, food descriptions etc. The inclusion of any such additional information in the label presentation for this research risks the ability to isolate comprehension of FOP labels and determine which elements of the information on the FOP label affect comprehension. In light of this and the objectives of the study which focus on the effects of variation within the FOP labels on objective comprehension, label presentation is to be held constant, and will include details of **product name, weight, number of portions above the generic FOP label format showing information per portion**.

The selection of products may affect the way the different signposting schemes are interpreted. Nutrients in products that are consumed as small portions (snacks, breakfast cereals etc) will be low in terms of %GDA, yet can still appear as high in terms of TL, as this is calculated per 100g. There would be less of an apparent discrepancy for products consumed in larger portions (e.g. main meals). It is hypothesised that this could impact on people’s comprehension. Policy interest in the effect of portion size issues leads us to include two product categories in the design (food which represents a smaller part of a meal and food which represents a larger part of a meal).

It is also considered important to cover a range of products within these categories in the tests. The products to be used for the tests are breakfast
cereals, yoghurt and crisps in the ‘smaller part of a meal’ category and a ready meal, a sandwich and soup in the ‘larger part of a meal’ category. These cannot be included as part of the factorial design for practical reasons but products will need to be distributed randomly to ensure no bias is introduced (see Chapter 5).

**Hypothesis for label presentation**

- There is a greater difference in comprehension between %GDA and TL for products which form a smaller part of a meal than for larger meal products.

### 1.4 Measures of objective understanding

Tests are needed to test objective (not perceived) understanding, to test the different elements of the schemes, and (where possible) reflect the way FOP labels are used in reality, without placing an undue burden on respondents.

Three test formats are to be included in the design, covering:

- Test 1 - Evaluation of the level of a single nutrient in a product (two nutrients included per product)
- Test 2 - Evaluation of the healthiness of a single product
- Test 3 - Comparison of two or more products in terms of healthiness.

These have been identified as being most likely to discriminate between the different types of FOP signposting, and to reflect the most common use of FOP labels.

Tests used in previous studies have been evaluated, and three potential tests developed for use in this study. These have been tested cognitively and refined to ensure they truly test the required aspect of comprehension (as described above).

Each test must have a pre-defined correct answer against which respondents’ answers can be judged. There is no clearly defined process for producing a totally objective measure. FSA has, instead, conducted a survey of nutritionists and dieticians to define the correct answer for selected products for each of the tasks. Only products where good agreement between individuals is reached are included in the survey.

Alongside accuracy of response, time taken to respond will also be taken into account. A label which takes a long time to interpret is not likely to be useful in
reality. The time taken to respond to each individual test presentation will, therefore, be recorded during the interview.

- Using these three tests allows the hypotheses identified to be tested in terms of three different aspects of comprehension:
  - Evaluation of the level of single nutrients in a product
  - Evaluation of overall product healthiness
  - Comparisons of overall product healthiness

### 1.5 The testing design

The decision on label format produces eight label versions within the full factorial design, and two further labels for inclusion outside the full factorial design (10 labels in total).

Testing comprehension of these label elements is to be carried out within respondent, meaning the same respondent is presented with all label versions for any test they carry out. This gives the best level of confidence that any differences in comprehension between the different types of signpost label are real.

**Energy** (calories present or not) will be included as a factor, but only for overall single product judgments. For all other tests it will be present and held constant. This will be carried out between respondents, with one set of respondents seeing the labels with energy, and a second set seeing those without. Random selection of respondents will ensure that the two groups are fully comparable for this purpose.

Each of the three tests will be shown for all 10 labels except for test 1 (individual nutrient evaluation) which will only be shown for the 8 labels in the full-factorial design.

All elements of presentation of the label not under test will be held constant as follows:

- Label generically reproduced
- Horizontal direction (except for label 10)
- FSA recommended nutrient order
- Nutrient information within rounded rectangle. GDA at bottom, text at top; white background where no TL (except for label 9)
- Generic product name, with weight and number of portions. Label shows information per portion.
- Where energy is present, this will be presented to the left of other nutrients, and signposted.

Each test will be presented for two **products categories**: P1 (a meal sized product portion) and P2 (smaller portion or snack products). This will also be carried out **within respondent**, with each respondent presented with each label for both P1 and P2 products.

The **final design** is shown in the grid below.

<table>
<thead>
<tr>
<th>Group of respondents</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test 1</td>
<td>Test 2</td>
<td>Test 2</td>
<td>Test 3</td>
</tr>
<tr>
<td></td>
<td>P1</td>
<td>P2</td>
<td>P1</td>
<td>P2</td>
</tr>
<tr>
<td>TL Text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDA</td>
<td>L1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no GDA</td>
<td>L2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDA</td>
<td>L3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no GDA</td>
<td>L4</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>No TL Text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDA</td>
<td>L5</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no GDA</td>
<td>L6</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDA</td>
<td>L7</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no GDA</td>
<td>L8</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>No TL no text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDA</td>
<td>L9*</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>no GDA</td>
<td>L10*</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TL no text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no GDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>16</td>
<td>16</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

- * - has non-TL colour – label approximates to that used by Tesco
- * - has circular presentation – label approximates to that used by Sainsbury
- * - E indicates energy is NOT included on the label Energy will be present on other labels.

The design gives a total of 92 test presentations. These will be split into four groups of tests, each going to a randomly selected quarter of respondents. Each group will see the full range of labels once for each test, for both P1 and P2. Within each group, the actual product shown with each label type will be rotated, and the order in which the tests are shown will be randomised. This will avoid any effects from ordering or product selection. At test 1, where possible, the same two nutrients will be asked about for P1 and P2 within each label type.

Tests will be self completion, administered on a laptop computer. All tests will be timed individually, with a maximum of 20 minutes spent on tests.

For the eight labels included in the fully factorial design, logistic regression will be used to look at individual and interaction effects of TL, GDA, Text and product on comprehension (i.e. the ability to give correct answer).
As a secondary measure ANOVA will be used to look at effects of TL, GDA, Text and product on the time taken to give an answer. This will be used to help distinguish between label types with very similar levels of comprehension.

In addition, logistic regression will be used to establish which key demographic and behaviour variables have an impact on comprehension of the eight label types (see section 1.6 for details of variables).

Logistic regression will also be used to compare label 9 with label 7, and label 10 with label 4, to establish any impact of using a circular rather than horizontal direction of presentation, and of the use of non-TL colour (but ONLY within the formats in which these are currently used in the marketplace).

1.6 Other factors influencing comprehension

Information on the following will be collected, as evidence suggests each may influence comprehension:

Key demographics:

- Age
- Sex
- Household composition
- Ethnicity
- Terminal education age
- Social grade
- Postcode (to allow matching to geodemographic information)
- Colour blindness

In addition we will collect information on:

- Shopping habits
- Dietary needs
- Attitudes to labelling and healthy eating
- Basic numeracy and literacy (in relation to food labels)

1.7 Data collection and sampling method

We will collect data using random probability sampling, with in-home face-to-face interviews, using Computer Assisted Personal Interviewing to allow labels to be shown on screen and the tests to be self-completed. We anticipate a response rate of around 65%, ensuring a representative sample. We will interview 3000 UK

---

4 The number of respondents from minority ethnic groups will not be sufficient to explore differences in any detail, as ethnicity was not identified by the qualitative research as a key factor influencing comprehension of FOP labels.
main shoppers, providing a robust basis for statistically testing respondents’ comprehension of the different labelling approaches. The interview will last around 30 minutes on average, including a maximum of 20 minutes on the tests.

The method and questionnaire has been tested in a small scale field pilot of 25 respondents, and a cognitive interviewing stage of around 100 respondents. This latter stage used qualitative techniques to ensure that the test questions (and other key questions) are designed to address the hypotheses.
2 Elements of FOP nutrition signpost label - formats

The first task is to identify the various elements of the FOP labelling schemes that could influence comprehension (the independent variables). Two main areas were identified for consideration:

1. Label format (information and presentation within the label itself) (discussed in this chapter)
2. Label context (the context within which the label is shown, including the product to which the label applies) (discussed in Chapter 3)

2.1 Label format: content and presentation

The Food Standards Agency’s (2007) current technical guidance on FOP signpost labels recommends the inclusion of fat, saturated fat, sugars and salt in the label information, with energy optional. There are a variety of label formats in the UK: most (although not all) include all four nutrients and others include further nutrients and energy (presented as calories).

Information on FOP signpost labels can be described as either content or presentation related. FSA recommendations currently cover content and not presentation. Content related elements provide information about nutrients and calories. This always includes the level of nutrient present (in grams), and can additionally include the use of interpretive elements within the signpost design to provide further information about the level.

The different signposting methods in current use include:

1. Red, amber or green coding to provide information on the levels of nutrients, also known as traffic light (TL) colour use;
2. % Guideline daily amounts (GDAs);
3. Text referring to the levels of nutrients, i.e. low, medium and high (this currently only appears in conjunction with traffic light colours)

Presentation-related elements can also be used to aid interpretation of the content related elements. These can cover presentation within the signpost label (discussed in this section), and issues relating to the context within which the label is presented (see Chapter 3). Within-label presentation-related elements have the potential to increase the salience of one or more pieces of the content-related information. The extent to which content is supplemented with additional presentational cues varies considerably.

The various elements of FOP labelling schemes, both content and presentation-related (within label), are set out in Table 1. The level of each nutrient (in
grams) is not included in the table below, as it appears in all labels and is not, therefore, a variable element. Each element is an independent variable and each has at least two levels. There are potentially more elements and levels than are shown in the table.

**Table 1: Overview of main label format elements currently in use**

<table>
<thead>
<tr>
<th>Elements</th>
<th>Levels</th>
<th>Description of levels</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nutrient Content</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traffic light (TL) signposting</td>
<td>2</td>
<td>Without TL signposting</td>
</tr>
<tr>
<td>% GDA signposting</td>
<td>2</td>
<td>Without % GDA signposting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With % GDA signposting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With TL signposting</td>
</tr>
<tr>
<td>Text-based signposting</td>
<td>2</td>
<td>Without text-based signposting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With text-based signposting</td>
</tr>
<tr>
<td>Inclusion of energy (calories)</td>
<td>2</td>
<td>Without energy (calories)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With energy (calories)</td>
</tr>
<tr>
<td><strong>Presentation (within label)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direction of presentation</td>
<td>3 (more possible)</td>
<td>Linear – horizontal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear – vertical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Circular</td>
</tr>
<tr>
<td>Colour in presentation</td>
<td>2</td>
<td>Non TL colour e.g. nutrient specific colours</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monochrome</td>
</tr>
<tr>
<td>Shapes within which label information appears</td>
<td>3 (many more possible)</td>
<td>Oval</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hexagon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘Finger nail’</td>
</tr>
<tr>
<td>Order of presentation</td>
<td>6 (many more possible if more label elements are used)</td>
<td>Nutrient name – nutrient level - signpost</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nutrient level – signpost - nutrient name</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Signpost - nutrient name - nutrient level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nutrient name – signpost – nutrient level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nutrient level - nutrient name – signpost</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Signpost - nutrient level - nutrient name</td>
</tr>
<tr>
<td>Position of information within FOP label</td>
<td>3</td>
<td>Above left</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Below right</td>
</tr>
</tbody>
</table>

Taking all possible combinations of the content and presentation elements in Table 1, there are over 5000 possible label combinations (although most combinations do not exist in practice). As the intention is to use a full-factorial design (see introduction), we need to establish the priorities for inclusion in this piece of research. For example, including TL and % GDA would result in a 2X2 = 4 label design (since each has 2 levels). Adding in either text-based signposting or energy would increase this to 2X2X2 = 8 labels. The inclusion of the direction of presentation would increase this three-fold and so on. Consideration of the
evidence and objectives has enabled us to produce a list of first, second and third level priorities for consideration.

The discussion in Section 2.2 draws on the evidence in coming to conclusions about these priority elements for inclusion in the study.

### 2.2 Background to content-related elements

#### 2.2.1 Nutrient levels

In all cases, FOP labelling includes the levels of the nutrients e.g. grams fat, grams sugar. The research will therefore include this information on all of the labels as a constant.

#### 2.2.2 Traffic light (TL) signposting

Red, amber and/or green signposting: This coding provides ‘at a glance’ interpretive information on the level (i.e. whether high, medium or low) of individual nutrients in the product. The Food Standards Agency’s (2007) current guidance provides the nutritional criteria that determine the colour coding for fat, saturated fats, sugars and salt. See section 2.2.5 for discussion of colour coding of energy (calories).

#### 2.2.3 Interpretive text-based signposting

This also refers to the levels of nutrients, categorising them as low, medium and high. It essentially presents the same information as TLs but in words rather than colours. In the current UK marketplace, it is only used as a cue in addition to the TL colour coding.

#### 2.2.4 % Guideline daily amount (GDA) signposting

% GDA has a more complex history. In 1996, MAFF published a leaflet entitled Use your Label: Making Sense of Nutrition Information (Williams & Rayner, 1996), the first official publication to provide Daily Guideline Intakes (DGIs) for five key nutrients: fat, saturated fat, sodium, sugar and fibre. These were derived from the population dietary goals in two official COMA reports: Dietary Reference Values for Food Energy and Nutrients for the United Kingdom (Department of Health, 1991) and Nutritional Aspects of Cardiovascular Disease (Department of Health, 1994).

Daily Guideline Intakes (DGIs) became more widely referred to as Guideline Daily Amounts (GDAs) as a result of qualitative research carried out by the Institute of Grocery Distribution (IGD 1997) which explored consumer preferences for different terms for DGIs. The resulting publication, Voluntary Nutritional Labelling Guidelines to Benefit the Consumer (IGD, 1998), set GDAs for fat, saturated fat
and calories based on the 1991 COMA report. In 2005 these GDAs were reviewed and expanded to include GDAs for carbohydrates, sugars, protein, salt and fibre for adults and specific GDAs for children for defined age groups (IGD, 2005), also based on the COMA report, although the GDA for salt was based on the Scientific Advisory Committee on Nutrition’s (SACN) recommendations for salt and health (SACN, 2003).

Whilst GDAs were originally developed to communicate nutrient levels that individuals should use as a target if the population is to achieve agreed nutrient goals (Rayner 2003), GDA values for fat, saturates, sugars and salt typically represent upper limits for consumption (Denny, 2006).

### 2.2.5 Energy

Energy (as calories) is included in most of the FOP signpost labels which appear in the marketplace. However, the way in which it is presented varies. It is part of the standard % GDA system and is signposted in terms of the percentage of GDA energy supplied by the product. In order to keep the presentation of energy constant across TL and %GDA it will be necessary to agree the criteria for TL colour coding energy. As previous FSA research (FSA 2005b) showed that TL colour coding single nutrients enabled the consumer to make healthier choices, it was not necessary to TL colour code calories and TL criteria were not developed for calories by the FSA. In view of there being no recommended criteria for this interpretive element, it is proposed that the criteria reflect how they are used in the marketplace by retailers.

Including energy on the FOP label offers the possibility of contextualising the portion of food within the overall diet.

### 2.3 Evidence on the performance of schemes using the different content-related elements

#### 2.3.1 % GDA and TL colour

*Work covering both GDA and TL colour*

In their review of research in this area, Grunert and Wills (2007) concluded that consumers feel that both % GDA and TLs are generally easy to understand. On measures of objective understanding, TLs were found to outperform various versions of % GDA schemes, when participants were asked to indicate whether a food was high, medium or low in a nutrient (FSA, 2005b; Which? 2006). A combination of GDA and TLs were most successful when respondents were asked to compare similar products and say which product was higher in a particular nutrient (FSA, 2005b), but only GDA and TLs also elicited a high level of correct response. Grunert and Wills (2007) concluded that, where less processing of
information is required by the participants (as on TL labels), they are more likely to interpret the information on the label accurately. Grunert and Wills (2007) also concluded that multiple traffic lights seem to outperform GDA-based systems in simplicity, but the GDA-based systems were less likely to make consumers feel coerced or that a decision is being made for them that they do not understand.

Which? (2006) found that TL and % GDA systems did not differ in their effect on consumers classifying single products as healthy, but TLs were more effective than % GDA systems when the test was to compare products.

**Work focused on GDA**

A recent review (FSANZ) concluded that while %DV/GDA information potentially provides consumers with useful information with which to assess products, many find the concept difficult and are confused by some aspects of its application to product evaluations. Research by Cereal Partners Worldwide (2006) found that objective measures of understanding are easier for the % GDA format when compared to the same information presented as a bar chart.

Cowburn and Stockley (2005) concluded from their review that the inclusion of a benchmark such as % GDA seems to help consumers place an individual product in the context of their overall diet (also reported in European Heart Network, 2003). Studies using traffic lights were not included in these reviews.

FSA (2007a) concluded that whilst GDA schemes were an improvement over the nutritional information alone, they still needed to be read, and nutrients were not flagged in an immediately noticeable way.

Phase 1 qualitative findings from this study (see Annex) indicated that consumers were confused about whether and how % GDAs applied and whether there were different GDAs for different people. Also, some people did not understand the term ‘GDA’; this tended to put them off using FOP labels. There was also some misunderstanding about what ‘% GDA’ meant. Those who did understand GDA labels were likely to be those who have a good understanding of nutrition generally.

**Work focused on TL colour**

TL colour was found to have advantages for those who cannot read or for whom English is not their first language, but can have disadvantages for those who are colour-blind (with respect to red, amber and green). Kaufman-Scarborough (2000) noted the advantage of using text in addition to colour for those with difficulties in perceiving colour. However, colour can have different connotations; while red is associated with fear and anger in Anglo-Saxon cultures, it is associated with love and happiness in Chinese culture (Aslam, 2000). Given the
common association of red, amber and green with traffic lights in the UK, it might be expected that most of the target population would share this connotation.

Phase 1 qualitative findings (see annex) found that although there was considerable recognition of TL labels, respondents did not always realise that TL colours had a specific meaning or had false assumptions about what it meant. Various assumptions were made, such as the colours being intended to make the labels stand out or the nutritional elements being colour-assigned, as in GDA labels that used pastel colours specific to each nutrient. Once they did understand TLs, however, people found them useful. For those who understood them, TL labels enabled instant recognition and the easy viewing of nutrient levels via the TL colours. TL labels were valued by people who were short of time or were generally unwilling to make a decision based on gram weights of nutrients alone. In these cases, TL colours were used to make quick decisions. This reinforced the findings reported by the FSA (2007a) that TLs were seen as quick and easy to use, allowing ‘at a glance’ evaluation.

In the phase one qualitative (see annex) work hybrid labels (showing both TL and % GDA) tended to be well received by respondents (subjectively), judged as allowing people to take whatever element they wished and use what they were comfortable with. A hybrid design with text supporting TL colours was considered by some to contain all the necessary information and be especially helpful for people who needed extra guidance. In the work conducted by the FSA (2007a) colour coded GDA labels performed most strongly in comparison tasks.

### 2.3.2 Interpretive text

Cowburn and Stockley (2005) concluded from their review that the inclusion of verbal banding, such as high, medium and low, aids interpretation of numerical information but other forms of bar charts and star ratings were found to be more confusing. The European Heart Network (2003) review concluded that verbal descriptions such as high, medium and low can help people in placing a food in the context of their overall diet. There are no studies that have looked at the effect of including or excluding text based signposting information in addition to either TL or % GDA signposting.

Phase 1 qualitative findings (see annex) indicated that text was used by people who did not understand that colours in TL labels had meaning. There were also examples where respondents spontaneously suggested that the terms ‘low/medium/high’ should be included on labels to aid understanding.

### 2.3.3 Energy

Cowburn and Stockley’s (2005) review and recent FSA studies (2005b) indicated that calories are among the most frequently looked-at information on food labels
in Europe. Recent European qualitative research (van Kleef et al., 2007) suggested that highlighting energy on the front of the pack is a promising platform of communication.

Phase 1 qualitative findings (see annex) indicated that respondents understood calorie intake limits and were used to thinking about and calculating calories. They were often used when making comparisons between products. It was not unusual for calories to be consulted as a proxy for the overall nutritional value of products, especially where people had difficulty understanding other FOP label elements. FSA (2007b) also found widespread use of calorie information, with some groups (such as older women) looking almost exclusively at calories.

2.4 Evidence on the performance of schemes using the different presentation-related elements

2.4.1 Direction of presentation

The nutrients in the FOP signpost label can be presented vertically, horizontally or in a circle. Other formats are possible but are not currently used. The majority of labels in the marketplace use a horizontal approach, though one major supermarket (Sainsbury’s) uses a circle and some other companies use a vertical display.

There is good evidence that, in European cultures, information is read left to right and top to bottom and that information presented at the top of a display is afforded more attention (Corney et al., 1994) but no evidence was identified that explores whether horizontal or vertical presentation of information is better on FOP signposting labels. There have been studies examining the effectiveness of the orientation of graphical information (Fischer et al., 2005) but that does not directly relate to the type of numerical and verbal information presented in FOP signposting. They found comprehension was faster for vertical bar graphs than for horizontal bar graphs.

The phase 1 qualitative research (see annex) found that respondents appeared to find horizontal presentation easiest to read of the three alternatives.

Earlier research (FSA 2007a) identified some reported problems with comprehension of circle presentation and the phase 1 qualitative findings also indicated that some respondents believed it to be a Pie Chart, with the size of wedge having meaning.

2.4.2 Colour used to draw attention to particular parts of the label

One major supermarket (Tesco) uses pastel colours to differentiate the individual nutrients in the FOP signpost label. This means exposure to the use of colours to draw attention to the different nutrients is widespread in the market place. Other
manufacturers use a single colour as the background for all nutrients (e.g. a pale blue).

The literature review did not identify any data on the effect of nutrient based colour, though the phase 1 qualitative findings (see annex) identified some confusion among respondents. Where manufacturers use a single colour across all nutrients there were examples of respondents ascribing values to the chosen colour, such as interpreting blue as an indication that that level was ‘low’ for all nutrients. On labels assigning different colours to each nutrient there was a similar tendency to ascribe value meaning to the colours. Where this happened, the ‘cooler’ colours – blue and green were assumed to indicate nutritional element content was low (being equated to green in the TL scheme). In both cases, this caused problems when comparisons were being made between products with % GDA and TL labels. This reflected the earlier findings of the FSA (2007b).

Visibility is linked to the use of colour. Respondents also reported that monochrome labels could ‘blend into the background’ of packaging and therefore were not always easily spotted. The phase 1 qualitative findings (see annex) found that some people deemed FOP labels too small to see or to be noticed. Monochrome labels and, exceptionally, labels using a variety of pastel colours were sometimes judged hard to see on packaging.

2.4.3 The order in which nutrients and other information appears

The order in which nutrients are listed in FOP signpost labels varies, and there are 120 possible ways in which calories and the four nutrients could be ordered. Calories (if present) are usually presented first, but the nutrients are then presented by different manufacturers in a number of different orders. FSA recommends the order of fat, saturated fat, sugar and then salt.

No evidence on whether the order of information on FOP labels affects ease of understanding was identified in the literature review and it was not identified as an issue in the qualitative work of this study.

2.4.4 Shapes within label

The information on each nutrient in FOP signpost labels is usually included within an outline shape such as oval, hexagon, or ‘finger-nail’ shaped.

No evidence on whether the use of different shapes affects the ease of comprehension of FOP signposting information was identified in the literature review and nor was it identified as an issue in the qualitative work of this study.
2.4.5 Position of information within FOP label

The position of the information within the FOP labels also varies, e.g. whether the words high, medium or low are to the left or right, or above or below the other information, or where the word “fat” is positioned in relation to the information on level of fat.

The literature review did not identify any evidence on the effect of the position of information within the FOP label on ease of comprehension of information in FOP labels. It was not identified as an issue in the qualitative work of this study.
2.5 Implication for inclusion of FOP label elements in study design.

Table 2: Pros and cons of including each element in the study design

<table>
<thead>
<tr>
<th>Nutrient Content</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>% GDA signposting</td>
<td>Exploring its effect is a key aim of this study. Clear evidence of impact on comprehension</td>
<td></td>
</tr>
<tr>
<td>Traffic light (TL) signposting</td>
<td>Exploring its effect is a key aim of this study. Clear evidence of impact on comprehension</td>
<td></td>
</tr>
<tr>
<td>Text/no text signposting</td>
<td>Inclusion would allow a test of whether it is the TL or text that is best at providing information or if there is an interaction effect between the two. Could address evidence on problems in interpreting colour across cultures, and for colour blind Evidence that can aid comprehension of numerical information</td>
<td>Text only exists along with TLs in the market place, so inclusion would introduce a number of purely hypothetical label types</td>
</tr>
<tr>
<td>Energy/no energy</td>
<td>Evidence indicates that people like and use energy information on FOP labels. May be useful in contextualising within the overall diet. Evidence that it is used as a proxy for overall nutritional value of products.</td>
<td>Evidence suggests that energy is a second order issue. No FSA criteria for the TL colour coding of energy. Energy is presented in different ways on labels (e.g. sometimes not traffic lighted, sometimes traffic lighted)</td>
</tr>
<tr>
<td>Presentation (within label)</td>
<td>Would allow comparison of some widely used systems (e.g. Sainsbury circle, Waitrose vertical designs). Evidence that wheel causes some comprehension problems</td>
<td>Increases number of label types in the factorial design threefold, and the majority of these additional labels are hypothetical, i.e. there are no examples in the market place.</td>
</tr>
<tr>
<td>Nutrient-specific colour in presentation (e.g. pastel)</td>
<td>Wide exposure in the market; may be a familiarity effect based on exposure to these in the market place. Evidence that there are problems with comprehension both with pastel colours and (to a less extent) pale monochrome labels.</td>
<td>Increases the number of label types in the design, including a large number of hypothetical cells, as only in use by one manufacturer at the moment</td>
</tr>
<tr>
<td>Order of presentation</td>
<td></td>
<td>Greatly increases the number of cells in the design. No empirical or theoretical justification available</td>
</tr>
<tr>
<td>Shapes</td>
<td></td>
<td>Greatly increases the number of cells in the design. No empirical or theoretical justification available.</td>
</tr>
<tr>
<td>Position of information within FOP label</td>
<td></td>
<td>Increases the number of cells in the design. No empirical or theoretical justification available.</td>
</tr>
</tbody>
</table>
Table 2 summarises the evidence for and against including each of the content and presentational factors in the design. Given the objectives of the survey (see introduction) and on the basis of the evidence it was concluded that label content is fundamental to comprehension of FOP labels, and as such the research design gives prominence to content.

2.5.1 Content issues in the research design

A key objective of this stage of the research is to explore the effects of TL and % GDA on relevant tests of objective understanding and the evidence discussed above shows clearly that TL and % GDA have some impact on comprehension. There are two levels to each of these elements, giving four possible combinations (i.e. four types of label).

There is also fairly strong evidence that the other two content-related elements, text and energy, have some influence on comprehension. Each of these has two levels. Inclusion of either element would mean that the study has eight possible combinations of element levels, while the inclusion of both gives 16 combinations.

The addition of the ‘text’ element would allow the hypothesis that text presented with TLs makes a difference to comprehension to be tested. It is also possible that text information alone may be as effective as TL information. Inclusion in the full factorial design would allow assessment of the effect of text alone, and in combination with TL, with GDA and with both TL and GDA. If text were not included in the design, there would be a danger of concluding that either TL or GDA was better, when the real difference may be the result of the inclusion of text alongside TL. Given the qualitative evidence on text as an aid to comprehension, text was selected as a priority area for the full-factorial design.

The inclusion of energy (calories) as an element offers the possibility of exploring whether the comprehension of healthiness, for example, is different for labels which include energy. Indeed, evidence suggests energy may be used by consumers to decide on the healthiness of a product. There is no evidence that the presence (or not) on a label of energy would affect people’s ability to judge the level of an individual nutrient, but it is possible that judgements on overall product healthiness will be different if energy is presented, compared with labels where energy is not shown. This is based on evidence from the qualitative work that energy can be used as a proxy for healthiness by some people.

This said, energy appears on most FOP labels in the market place and since energy is only likely to affect some measures of comprehension, this suggests energy should be a second level priority for inclusion.
2.5.2 Presentation issues in the research design

As explained above, content is prioritised over presentational elements. Based on the evidence available, presentation variables which could be considered for inclusion in the design are: the choice of a set background colour on monochrome labels, direction of presentation, and the use of nutrient-specific colour. The primary reason for considering these elements was, initially, their place in the market, but this was supplemented by the findings from qualitative research which identified problems with comprehension of both nutrient-specific colour and a circular presentation.

The evidence that the use of a specific nutrient background colour on a monochrome label can influence comprehension comes from the stage 1 qualitative work. This suggested that the use of a cool colour (e.g. pale blue) may be interpreted as meaning low levels of all nutrients. There are, however, a number of colours currently in use in the marketplace, and many more that could potentially be used. To test this element fully would require the inclusion of a number of different colours, which would greatly increase the number of label types to be tested and there is little evidence available to inform the choice of colours. Further qualitative exploration of how a range of different colours are interpreted would illuminate this issue.

Adding a direction of presentation variable would increase the number of label combinations by three. Most of the labels in the marketplace are in a horizontal format and this was the direction that the qualitative work suggested was easiest to use. Inclusion of this element in a full factorial design would mean the creation of many further theoretical labels, not currently in use in the market place, and only one that is being used. Since the qualitative work did suggest that the Sainsbury-style circular presentation could cause some confusion and given its exposure in the marketplace, it may be worth further exploration.

A further label format candidate for inclusion in the design is the use of nutrient-specific colour to draw attention to the different nutrients, as currently used by Tesco. Inclusion of this as a variable within a full factorial design would double the number of label types, and all but one would be hypothetical label designs. However, given the evidence on the confusion it causes and its exposure in the market place, we feel it is worth exploration.

Table 3 shows the potential number of labels in the design if all of the 3 priority content elements, and the two lower priority presentation elements were included in a full-factorial approach. Where a label type is currently in the marketplace, an example is given of where it is used.
Table 3: Pros and cons of including each element in the study design

<table>
<thead>
<tr>
<th>DIRECTION OF PRESENTATION/ USE OF NON-TL COLOUR</th>
<th>Horizontal</th>
<th>Vertical</th>
<th>Circle</th>
</tr>
</thead>
<tbody>
<tr>
<td>No colour</td>
<td>Colour</td>
<td>No colour</td>
<td>Colour</td>
</tr>
<tr>
<td>LABEL 1 e.g. Asda</td>
<td>LABEL 9</td>
<td>LABEL 17 e.g. McCain</td>
<td>LABEL 25</td>
</tr>
<tr>
<td>LABEL 2 e.g. Londis,</td>
<td>LABEL 10</td>
<td>LABEL 18 e.g. Waitrose</td>
<td>LABEL 26</td>
</tr>
<tr>
<td>LABEL 3 e.g. M&amp;S</td>
<td>LABEL 11</td>
<td>LABEL 19</td>
<td>LABEL 27</td>
</tr>
<tr>
<td>LABEL 4 e.g. Co-op</td>
<td>LABEL 12</td>
<td>LABEL 20</td>
<td>LABEL 28</td>
</tr>
<tr>
<td>LABEL 5</td>
<td>LABEL 13</td>
<td>LABEL 21</td>
<td>LABEL 29</td>
</tr>
<tr>
<td>LABEL 6</td>
<td>LABEL 14</td>
<td>LABEL 22</td>
<td>LABEL 30</td>
</tr>
<tr>
<td>LABEL 7 e.g. Morrisons</td>
<td>LABEL 15 e.g. Tesco</td>
<td>LABEL 23</td>
<td>LABEL 31</td>
</tr>
<tr>
<td>LABEL 8</td>
<td>LABEL 16</td>
<td>LABEL 24</td>
<td>LABEL 32</td>
</tr>
</tbody>
</table>

Since the three content elements are the top priority, this gives a minimum of 8 labels (labels 1 to 8). The scale of including direction and use of non-TL colour is illustrated in the table (raising the number to 48 labels). It also shows the number of labels this would introduce that are NOT currently in use in the marketplace (all but five). If energy were also included as an element, the number of label types would double to 96. Again many of these additional labels would be hypothetical since the majority of manufacturers include energy on their labels.

It is possible to include the Sainsbury-style wheel (label type 36) and the Tesco-style pastel label (label type 15) in the design without including ALL possible combinations of direction, and use of non-TL colour (i.e. outside of the full-factorial design). This would reduce the number of labels in the design from 48 to 10 (labels 1 to 8 plus labels 15 and 36). The performance in the test of label 15 could be compared with that of label 7 from which it differs ONLY by the inclusion or not of non-TL colour. Likewise, the performance of label 36 can be compared with that of label 4 from which it differs only by direction.

This approach does not allow us to look at the interaction of direction or non-TL colour with any other elements. For example, If label type 36 (the Sainsbury style label) differed in comprehension from label type 3 we would not be able to tell if this difference was the result of the circular vs horizontal presentation, or the presence/absence of GDA, or some interaction of the two.
This approach would, therefore, only allow a limited evaluation of the impact of the circular presentation, and the use of nutrient specific colour. It would, however, allow consideration of how these two elements impact on comprehension in the way they are currently used in the marketplace, compared with labels that are otherwise the same in terms of content.

### 2.6 Summary of priorities for format elements

Since all labels currently in the marketplace include the amount of each nutrient in grams, this will be held constant as present in all labels.

Given the research objectives, and the evidence available, the **top priority** for inclusion in the full-factorial design are three content-related elements:

1. % GDA / no % GDA signposting
2. TL signposting / no TL signposting
3. Interpretive text (high, medium low) / no interpretive text

The hypothesis is that each of these three would have a different impact on comprehension, and that there may be some interactive effect between them.

The **secondary priority** is the fourth content-related element

4. Energy / no energy

Here the hypothesis is that the presence or not of energy would have some impact on the ability to make overall product judgements (this judgement would be different when energy is present, from that made in its absence).

The **third level of priority** includes two presentation-related elements

5. Direction of presentation (horizontal, vertical, circular)
6. Use of non-TL colour / no use of non-TL colour

Here the hypothesis is that a circular presentation, and the use of non-TL colour would adversely affect some measures of comprehension.

No other elements are considered priorities given the evidence available. Where an element is not included in the design, it will be held constant. The way it is most frequently used in the marketplace will be used to guide the decision on to form to be used across all labels in this case.

The final decision on which elements to include in the design (and how) will be discussion in Chapter 5 in the context of other design decisions.
3 Elements of FOP nutrition signpost label - context

The context of presentation for labels also needs consideration. There are two key dimensions of label context: label presentation and food product.

3.1 FOP nutrition signpost label presentation

Label presentation deals with the context within which FOP signpost labels will be presented to study participants. This can vary from actual product labels as they occur in the market place through to generic product names along with the FOP signposting. It would be possible to systematically vary the label presentation to test impact on consumer understanding, but this would tell us nothing about the impact of FOP signposting itself on objective understanding (the key aim of this project). It was, therefore, decided to hold label presentation constant.

Decisions on the presentation to be used were informed by the evidence. The study by Levy (1996) provides a key example of a ‘thin’ label presentation context. The control condition label presentation context simply presented a product name and macronutrient information in metric units with no interpretational aids. Other conditions presented this information, plus the interpretational aids of interest and/or percentage information. Other studies have ‘thickened’ the context by using a short verbal description of the food with or without a simple line drawing of it. Extrapolating from consumer choice studies would suggest that stimulus material using verbal descriptions functions as well as actual pictures (e.g. Vriens et al., 1998; McCabe & Nowlis, 2003).

A different approach was taken by Feunekes et al. (2008). Here respondents were shown pictures of products with a FOP labelling format with branding information removed. An enlarged picture of the FOP labelling format was shown below the product picture to assess subjective comprehension measures (alongside measures of perceived healthiness and liking etc). It would not be possible to make all these judgements simply on the basis of the FOP label.

Other than the FSA (2005b) study, no studies were identified from the review of literature that used actual packs as the context in which to look at the effect of FOP labels (or any elements of these) on objective tests of understanding. Information on how labels are presented in comprehension tests is not always readily available; for example, Grunert & Wills (2007) and some of the original study reports, do not include this information.

Table 4 presents a range of options for the context within which the FOP signposting label could be presented, along with the implications of using each one.
Table 4: Possible types of label presentation context

<table>
<thead>
<tr>
<th>Types of label presentation context</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
</table>
| 1. Real packets/picture of real packets | As they appear – full packet or full packet front.  
+ May seem more ‘real’ to consumer and most closely approximates real world context for labels.  
- This would introduce far too much ‘noise’ to extract any meaningful information in the tests. Branding and other information will vary and could influence test answers meaning this would not be a fully objective test of the FOP label.  
- Does not allow for hypothetical labels to be included in the study |
| 2. Real packets/picture of real packets with some information altered/obscured | A ‘mock up’ of a real product. There are various sorts of information that can be removed (brand, picture, serving suggestion; weight; name; health claim; endorsement etc). At its most minimal this could go down to just showing the picture, name of product, weight/portion information and FOP labels.  
+ May seem more ‘real’ to consumer and most closely approximates real world context for labels to some extent  
+ Enables introduction of ‘theoretical’ FOP labels on real products.  
- People may still be able to identify the brand from the remaining information. Bias towards those familiar with the product.  
- Any information left on the packaging (pictures, colour schemes) could influence the test answers, meaning this would not be a fully objective test of the FOP label.  
- Needs rationale for deciding what to remove. Removing different components of the product will have different effects. |
| 3. Product name along with FOP label reproduced generically | Bears minimal relation to a real product. It could include product name, weight/portion information and FOP label.  
+ Enables introduction of ‘theoretical’ FOP labels.  
+ No danger that other information will drown out the FOP label  
- Is unlikely to seem ‘real’ to consumer and does not reflect what people actually come across in the market place.  
- Will not look like any particular product (this could be seen as an advantage as consumers will have fewer preconceptions about the product). |
Type 2 above represents a continuum between real labels at type 1 and the generic labels at type 3. Types 2 and 3 also introduce the possibility of manipulating nutritional information to produce greater variation of products for testing, or could be used based purely on real products. If a great enough variation of real products can be found, then real products should be used to ensure the final tests reflect the kinds of judgement people would need to make in reality.

The key issue in deciding how to present the FOP signpost label is the aim of the study. This is to identify the key elements (as discussed in Chapter 2) of FOP labels that best enable people to understand them. Any information that is additional to the key elements is likely to obscure the effects of any variation of the key elements.

On the basis of the review of evidence, presentation of the FOP label, along with the name of the product and portion information only, is most likely to maximise the chances of finding differences in the effectiveness of FOP labels. The evidence also suggests that verbal information is as useful as a picture, so there seems no need to include a picture which may influence judgements depending on the picture chosen. The more information that is included, e.g. picture of food, serving suggestion, product description, marketing information, health claims, brand, the more likely it is that any effects of FOP label format are likely to be ‘drowned out’ by this other information. The phase 1 qualitative findings (see annex), and an early pilot survey also supported this.

The phase 1 qualitative work found that it was not unusual for people to use ‘background packaging’ information, such as nutrition and health claims, branding, pictures on the packaging, back of pack information and appearance of the product through the packaging, before or instead of FOP labels. It also identified a high regard for and a belief in ‘healthy’ ranges. When asked to select a ‘healthy choice’, respondents who habitually bought products from such ranges tended to make an assumption that products in the range were healthier than those not and expressed surprise when this was not the case. The pilot work found that even including brief descriptions of the product (e.g. cheese sandwich, tuna sandwich) could influence judgement of healthiness. It was therefore decided that labels should be reproduced generically and presented with the minimum of background ‘noise’ in order to maximise the possibility of identifying differences in label comprehension resulting from the FOP label alone.

The phase 1 qualitative findings also produced evidence of confusion over portion size on all label types, both when comparisons between products were being made, and for single items. Examples of this are: not knowing whether, for example, 30g of cereal was a ‘normal’ portion for them; and assumptions that nutritional information related to the whole of the packaged item (such as fizzy
drink, ready meal or sandwich) when sometimes it only related to a portion of the packaged item.

This suggests that the information on portion size needs to be clear and consistent in the label presentations to prevent this being a source of confusion. We would recommend using labels that give the information per portion to ensure consistency, with the number of portions in the packet clearly marked.

All of the evidence clearly suggests that the most minimal form of presentation is needed (option 3). The main potential disadvantage of this choice is that it will not look like a real product in the marketplace. Given the aim of the study, we do not feel that this is a problem. This part of the study is intended to unpick the elements of labels that facilitate comprehension, and not to look at how labels are used in practice.

### 3.1.1 Summary of label presentation decision

Label presentation will be held constant.

Since the objectives of the study focus on the effects of variation within the FOP labels on comprehension, and the wide variety of other information on the packets would obscure this, the selected presentation format is in line with option 3 in Table 4 – product name, weight, number of portions and FOP label showing information per portion, reproduced generically.

### 3.2 Product category

The decision on product categories depends on whether there is any reason to expect that product category will affect comprehension. It would clearly be impractical to include all product categories in the research as this would entail asking far too many questions in the survey. Given that the products will be presented only as names, rather than detailed packs, the choice of product category can be expected to have less effect on comprehension of FOP label format, than if an actual product were presented. This is because there are likely to be fewer consumer expectations for the generic name of the food than for a full product description, with all the additional information this provides.

### 3.2.1 Evidence

**Comprehension:** The literature review did not identify any evidence to show that objective tests of comprehension (the focus of this study) vary depending on the product category although subjective (perceived) understanding has been shown to vary across product category (Feunekes et al, 2008). The differences
appear to be minor, although they were statistically significant. In the first study reported by Feunekes et al. (2008), ice cream, dairy drink and spreads were included. The differences in rated healthiness were greatest for dairy drinks and least for spreads. For spreads there was greater differentiation for particular signposting schemes (in this case between TLs which gives information on individual nutrients and products, and Smileys and Stars (the other systems under test) which give an overall assessment of healthiness of a food). In the second study, although two product categories were included, no differences were reported for product category.

**Products on which labels are used:** Qualitative work showed that health logos were more likely to be used by shoppers where the health status of a product was unclear (FNLI, 2005), while other studies found that consumers saw nutrition information as less relevant for certain product categories, e.g. fresh meat, canned vegetables (Directorate General for Health and Consumer Protection, 2005). FSA (2005b) found that consumers would like to see FOP signposting on foods such as burgers, sausages, ready meals, breakfast cereals and pizzas, rather than on food categories such as fruit and vegetables, rice, pasta (which were thought of as ingredients or which they understood the nutritional value of) or treats and snacks. The qualitative findings conducted for this study also highlighted the importance of ‘healthy’ branding in product selection.

In support of this, the phase 1 qualitative findings (see annex) suggested that labels were unlikely to be used:

- Where products are categorised as ‘treats’ (e.g. crisps, biscuits, confectionary). The justification was that there was no point in checking because treats were highly likely to contain high levels of fat, sugar and salt.
- Where products were understood to be ‘healthy’. Examples given were couscous and vegetables.
- Single ingredient items, such as fish, butter, cream, oats etc. These were seen as ‘innocent’ as they were unlikely to be processed to any degree.

They were most likely to use the labels on convenience foods and highly processed foods. This was also the finding of the FSA in earlier research (2007a).

**‘Healthiness’ ratings:** Bech-Larsen and Grunert (2003) investigated the impact of health claims and found that product category had a larger effect on perceived healthiness than the health claims included on the packaging. Likewise FSA (2005b) found that product category (rather than FOP label) is used to form a judgement about healthiness in single product assessments.
3.2.2 Criteria for selection of products

There is little evidence available of systematic differences in the effectiveness of FOP labels between product categories. Since the evidence shows that people’s prior knowledge of different products could affect their judgement, it was felt that a small range of products should be included in the study, but gave no clear basis on which to select specific products. One possible deciding factor is the degree to which products vary in the nutrients shown within the FOP labels, since the labels may play more of a role where there is more variation. A further factor is the evidence that labels are more likely to be used on products where the ‘healthiness’ status is less clear to the consumer. In addition we need to test out the labels on products for which FOP labels may be used in real life.

Further discussion within the project team revealed policy interest in the hypothesis that there may be some systematic difference between GDA and TL labels resulting from the way the information in the two different signposting schemes is presented. TLs are based on amount of nutrient per 100g in the main, whilst GDA are related to proportion of daily intake. This means that a food with a small portion size (e.g. crisps) may be coloured red for fat using TLs, but would have a low % GDA given their small portion size as part of the daily diet. In contrast, a product which is intended to be eaten in large amounts (e.g. ready meal) may show less discrepancy between the two signposting methods. This difference could produce differences in comprehension between GDA and TL-based labels, both in terms of overall product assessments (e.g. Is this product healthy? Which product is healthier?) and for single nutrient judgements (e.g. Is this product high in fat?).

Products were therefore split into two groups: those which form a smaller meal part (including snacks) and those which form a larger meal (intended as full meals). These two groupings gave the potential to vary the different nutrients, and to cover the types of foods for which labels are currently used.

Since it would not be possible to cover the full range of products in the two categories, it was decided to cover a small range of products to provide some variation for respondents and to provide a range of potentially easier and more difficult judgments.

Table 5 shows the three examples to be included in each of the two categories selected for the study. These were selected to ensure a range of variation across different nutrients.
Table 5: Product categories selected for study and likely variation in nutrients.

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Fat</th>
<th>Sat fat</th>
<th>Sugar</th>
<th>Salt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product grouping P1</td>
<td>A Readymeal</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
</tr>
<tr>
<td></td>
<td>B Soup</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
</tr>
<tr>
<td></td>
<td>C Sandwich</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
</tr>
<tr>
<td>Product grouping P2</td>
<td>D Yoghurt</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E Breakfast Cereal</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
<td>Varies</td>
</tr>
<tr>
<td></td>
<td>F Crisps</td>
<td>Varies</td>
<td>Varies</td>
<td></td>
<td>Varies</td>
</tr>
</tbody>
</table>

3.2.3 Varying product between respondents

A key decision is whether we can use a ‘within respondent’ design to test for difference between P1 and P2. This means presenting the same test on the same label for both P1 and P2 to the same respondent (see introduction for explanation of a ‘within respondent’ and ‘between respondent’ designs). This may be constrained by the number of questions any one individual can answer within the survey, and it may be necessary to look at product category on a between subject basis (some subjects get P1 and some get P2). Ideally we would use a ‘within respondent’ design as it is more robust, and means each respondent gets to see a wide range of products, reducing fatigue.

There is no intention to analyse by the three product types within each of P1 and P2. These are introduced purely to provide variation of nutrients and to reduce respondent fatigue. The final design will, of course, need to ensure that the products presented with each label do not bias the tests in any way, using a random design. This will be discussed further in Chapter 5.
3.2.4 Summary of product category decision

Two product categories: large portion (P1) and small portion (P2) will be included in the research design, with three examples of each to reflect variation in different nutrients.

The hypothesis is that the differences in the way GDA and TL are calculated will result in differences in comprehension for P2 products, with less difference for P1.

The two product categories will be presented using a ‘within respondent’ design if possible (see Chapter 5 for discussion), for reasons of robustness and avoiding respondent fatigue.

Within product category, the three product types will be distributed randomly (see Chapter 5 for details) to ensure no bias is introduced. There is no intention to look for differences between the three product types within P1/P2.

The products to be used for the tests are ready meals, sandwiches and soups for P1 and breakfast cereals, yoghurts and crisps for P2. These are included purely to provide variation for respondents.
4 Measures of objective understanding (the ‘tests’)

The required outcome of this research is to determine what scheme(s), or what combination of elements of the different schemes, best facilitates the accurate interpretation of key nutritional information by consumers, such that they are enabled to make informed decisions about the foods they consume. More specifically the research is required to assess objectively the extent to which individuals are able to correctly interpret the nutritional information given on front of pack labels.

In order to measure understanding we first need to clarify what we mean by understanding in the context of FOP labels. In addition we need to identify how labels can be used, and how they are used in reality, to guide the choice of measures for this project.

4.1 Subjective and objective understanding

Understanding includes both subjective (the meaning the consumer attaches to the information, or perceived understanding) and objective (whether the consumer’s understanding of the message is compatible with the intended meaning) components. The focus of this research is objective understanding. This chapter will discuss different types of objective understanding (Sections 4.2 and 4.3) before recommending measures for this study.

The intended meaning of the information being conveyed by the different forms of signposting varies. For example, providing % GDAs gives consumers ‘benchmarks’ against which they can judge the amount of a nutrient a portion of the food would provide relative to a recommended amount (Rayner, Scarborough & Williams, 2003). TL signposting also conveys information about the levels of nutrients relative to recommendations, with TL colour determined on a combination of information per 100g and per portion. Given the difference in information being communicated, different tests of objective understanding may favour different methods of signposting. Since the labels are all intended to provide consumers with the information they need (albeit in different ways) it is crucial that the tests of comprehension used in this study are even-handed and do not favour any particular signposting methods.

4.2 What is health numeracy?

Use of FOP labels involves dealing with numbers. Evans (2000) defines numeracy as ‘the ability to process, interpret and communicate numerical, quantitative, spatial, statistical, even mathematical information, in ways that are appropriate for a variety of contexts, and that will enable a typical member of the culture or subculture to participate effectively in activities that they value’. More recently Golbeck & colleagues (2005) combined this definition with one of health literacy
(U.S. Department of Health and Human Services, 2000) to establish the following definition of health numeracy – ‘the degree to which individuals have the capacity to access, process, interpret, communicate, and act on numerical, quantitative, graphical, biostatistical, and probabilistic health information needed to make effective health decisions’. Health numeracy includes:

**Basic health numeracy:** sufficient basic skills to identify numbers, and make sense of quantitative data requiring no manipulation of numbers (e.g. identifying a number as it appears on the label). This is also known as the ability to **replay** information (e.g. able to identify the amount of fat in a product).

**Computational health numeracy** which involves the ability to count, quantify, compute, and otherwise use simple manipulation of numbers, quantities, items, or visual elements in a health context so as to function in everyday health situations. This is basically the ability to **compute** a nutrient level based on information that appears on the label (e.g. able to equate a GDA of 20% with a maximum of 5 portions a day).

**Analytical health numeracy** which involves a higher level of literacy than the previous levels, i.e. the ability to make sense of information and involves higher-level concepts such as inference, estimation, proportions, percentages, frequencies, and equivalent situations (e.g. understanding the level of a nutrient). This is basically the ability to use the information to **evaluate** a product (e.g. able to judge how healthy a product is).

**Statistical health numeracy** which involves an understanding of basic biostatistics involving probability statements, skills to compare information presented on different scales (probability, proportion, percent), the ability to critically analyse quantitative health information such as life expectancy and risk, and an understanding of statistical concepts such as randomization and a ‘blind’ study. FOP labels are not intended to facilitate this level of numeracy among consumers, and this is the role of specialists.

**4.3 How can labels be used?**

The purposes for which food labels are used determine what format features are likely to be helpful (Levy, Fein & Schucker, 1996). The intent of nutrition labels is to allow meaningful comparisons between and across foods and to encourage the consumption of foods with potential to improve dietary intake and reduce the risk of chronic disease (Taylor & Wilkening, 2008). To be eating healthily the Government advise that people reduce the level of fat, saturated fat, salt and sugars in the foods they eat, and FOP labels contain the information to enable people to do this.
Front-of-pack (FOP) signpost labels have the potential to assist with the tasks consumers might commonly undertake. Cowburn & Stockley’s review (2005) distinguishes the following decisions where consumers might make use of food labels:

- Identify the amount of a specific nutrient a product contains. This is equivalent to *replay* in section 4.2.
- Assess what counts as a low or high amount of the nutrient. This is equivalent to *evaluation* of an individual nutrient in section 4.2.
- Decide the overall healthiness of a product. This is also *evaluation*, but of a product overall.
- Compare a specific nutrient content (or the overall nutrient content) of a product with one or more similar products or between different types of products. This takes replaying one step further into the ability to *compare two or more products*.
- Calculate the amount of a nutrient eaten in a serving. This is an aspect of *computing*; and
- Assess the product in the context of a meal choice or daily intake. Again, this is an aspect of *computing*.

This list covers the three main types of numeracy described in section 4.2. It also sets out the difference between tasks that involve a *single nutrient judgement*, and those that require an *overall product judgement* to be made, pulling together the information on multiple nutrients.

More recently, the review carried out by Grunert and Wills (2007) lists four types of objective understanding which are similar to those identified by other sources.

- replay part of the label information,
- evaluate a given product based on a given nutrient,
- evaluate the overall healthiness of a product, and
- compare two (or more) products based on a given nutrient

### 4.3.1 Categorisation of possible label uses

Combining the information on numeracy and the ways in which labels can be used, produces six major possible uses of the information in FOP labels, as shown in Table 6.
4.4 How are labels used?

There has been little work conducted on the use of labels in everyday life. Research has shown that consumers tend to use information in the form in which it is provided, rather than transform it (Bettman & Kakkar 1977; Jacoby, Chestnut & Silberman 1977; Klopp & MacDonald 1981, Levy & Fein, 1998), suggesting computation tasks are unlikely to be used commonly. Higginson & colleagues (2002) suggest that consumers only naturally use labels in the first two ways (comparing products and evaluating levels of single nutrients). FSA (2007a) asked people how they used labels but this relied on recall rather than observation, which can lead to over-stating use.

The most comprehensive recent evidence on how the labels are used in reality is that collected in phase 1 of the qualitative work, carried out as part of this project (see annex for initial findings). As explained in Chapter 3, this found that FOP labels were most likely to be used by consumers on convenience foods, and processed foods, and unlikely to be used on foods seen as treats, foods seen as healthy, and single ingredient items. Evidence was found of labels being used in the following ways:

*Making comparisons between products:* It was clear from the phase 1 qualitative findings that making comparisons between the same FOP labels could sometimes be difficult, for example if product portion sizes varied or the nutrients were not in the same format. There could also be difficulties in comparing between FOP labels (see the earlier discussion about misinterpretation of colour in Chapter 3). Generally, comparisons were made between similar products, for example,

---

**Table 6: Possible uses of FOP labels**

<table>
<thead>
<tr>
<th></th>
<th>Single nutrient</th>
<th>Overall product</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Replay</strong></td>
<td>e.g. how many grams of fat in this product?</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Compute</strong></td>
<td>e.g. how many portions could I eat a day and no have too much fat?</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Evaluate</strong></td>
<td>e.g. how high is this product in fat?</td>
<td>e.g. how healthy is this product?</td>
</tr>
<tr>
<td><strong>Compare products</strong></td>
<td>e.g which product is higher in fat?</td>
<td>e.g. which product is healthier?</td>
</tr>
</tbody>
</table>
looking at different ready meals. Parents made comparisons for children’s food, for example comparing breakfast cereals and cereal bars. Respondents also made comparisons when they were looking to ‘trade down’ from a branded to own brand product. They tended to assume own brand products, though cheaper, would be less healthy. Those who were just generally ‘health aware’ compared products using factors such as nutrients, taste and price.

Comparing products on particular nutrients: In making comparisons, people tended to focus on one or two particular nutrients, for example salt and sugar in breakfast cereals. Nutrients checked could vary; for example, those on weight loss/control diets tended to look at calories, fats and saturated fats and those with specific medical conditions might focus on salt or sugar (depending on condition).

Single product single nutrient judgements: Phase 1 qualitative findings (see annex) identified different ways FOP labels were used to make decisions on a single product:

- To check against a claim on the packaging, e.g. ‘70% less sugar’;
- To check on specific dietary requirements e.g. calories/fat content, for those on weight loss/maintenance programme; those on a Weightwatchers programme also used back of pack information to convert to the Weightwatchers points system. Similarly those with medical conditions used both FOP and back of pack information.
- To check levels of certain nutrients, for those who wished to eat healthily. These people tended to have their own ‘cut-off’ point on levels of nutrients, regardless of other factors, like price, taste etc and would take an ‘accept or decline’ approach. This included checking specific nutrients in particular products, such as the level of salt in a particular cheese, or the level of sugar in a juice.

Single product overall judgments: there were respondents who showed signs of brand loyalty – tending to buy the same products over time. Amongst these there were people who would notice if their usual purchase had a ‘new improved flavour’, for example, and would then check to see whether/how the nutrients had changed. This was not necessarily done with a view to comparing the product with another, but purely for information/interest purposes.

Related to this, the qualitative findings found no evidence that use of labels varied by education level, but there were indications that level of numeracy (particularly how comfortable people felt with numbers) was an important factor. It also found that people thought they had to do some mathematical manipulation in using %GDA labels and they found this off-putting; this was especially evident in relation to foods packaged as multiple portions. Basic misunderstanding of GDA caused problems. For example, some consumers thought that if the label said a
product contained 20% of your GDA for fat, that this actually meant the product was made up of 20% fat.

There was little evidence of computing techniques (e.g. calculating the number of portions per day, or using labels in the context of meal or diet choices). Earlier qualitative research (Food Standards Agency, 2004b) suggests that some consumers foresee some confusion when signposting information is used to choose whole meals or weekly shopping baskets. The recent qualitative work suggested that computation only seemed to occur if there was a very specific dietary need, such as members of WeightWatchers calculating ‘points’, or diabetics calculating the amount of sugar they could consume. The other uses of labels were considerably more common than computation.

4.5 Possible measures for inclusion in the survey

A review of the literature (using recent reviews as a starting point including European Heart Network, 2003, Cowburn and Stockley, 2005 and Grunert and Wills, 2008) revealed a large array of tests that have been used in previous studies. These tests were separated into the six categories of test identified in Section 4.3.1, Table 6. Full details of the tests are given in the Appendix to this document. These tables cite the pros and cons of each test, including the information they are likely to elicit, and whether they are likely to reveal any differences between different label types to be included in this study.

It should be noted that two issues are cited regularly as ‘cons’ in these tables; in practice, these are issues that need to be addressed, rather than absolute disadvantages:

1. Single nutrient-based tests: In order to fully judge a label, some tests would need to be repeated for each nutrient, which implies a large number of tests. In practice, it is possible to restrict the number of nutrients, for example by focussing on those which vary most for the product.

2. Objective criteria for the correct response to a comprehension test need to be established in order to assess responses to comprehension tests. These objective criteria need to be transparent, defensible and should not favour any label type.

4.5.1 Replay tests

The first set of tests do not go beyond replaying some of the label information (Table A1 in Appendix A). Grunert & Wills (2007) concluded that, usually, a majority of respondents can replay information on an individual nutrient correctly, and so such tests are not likely to discriminate between the different types of FOP signposting schemes. In some tests, text signpost information (in the form of ‘low’, ‘medium’ and ‘high’) forms part of the evaluation and thus become replay measures if presented in conjunction with text signposting. Tests that merely ask
consumers to replay part of label information are not true tests of understanding, as they do not reflect any interpretation of the nutrient levels and have therefore not been included in the research design.

Rothman & colleagues (2006) found poor label comprehension to be highly correlated with low-level literacy and numeracy skills. A separate assessment of basic literacy and numeracy will be included as part of the research and this will identify anyone who is not capable of replay measures (see Section 6.4).

4.5.2 Computation tests

Tests that merely involve carrying out computations using numbers appearing on the label (Table A2 in Appendix A) are also unlikely to discriminate between the different types of FOP signposting schemes, because the signposting information is not needed to carry out the task. Most of the tests involve simple numerical tasks and these can not be regarded as key tests of understanding, as they do not reflect any interpretation of the nutrient levels. They are therefore excluded from the research.

Some computation tests (Tests 2.2-2.4, 2.6 & 2.7) involve the manipulation of nutrient levels in relation to % GDA information. This is a more complex numerical task than merely multiplying (or dividing) nutrient levels based on the portion size as is the case in other tests. These tests will clearly only be possible with FOP labels that include % GDA. As stated earlier, it is important to avoid tests that favour a particular scheme. More importantly, the qualitative work and the literature both suggest that there is very little use of labels in this way.

As noted earlier, there appears to be very little use of computation in day to day life, particularly more complex computations. Since simple computations do not need to use signposting information, and the only element of signposting that lends itself to computation is % GDA, we recommend the exclusion of computation tests.

4.5.3 Evaluating products on individual nutrients

Tests that evaluate single products on a particular nutrient may also favour particular signposting methods (Table A3 in Appendix A). For example, Tests 3.1 and 3.7 ask consumers to assess nutrients on a scale that can easily be mapped onto the words high, medium and low used in text signposting. Test 3.8 only allows an assessment to be made in terms of % GDA.

The majority of the tests asked consumers to assess the level of a particular nutrient on a given scale. This was either a scale of healthiness or, more usually, a scale from ‘a little’ to ‘a lot’. A question that asks consumers to assess the amount of a nutrient (e.g from ‘a little’ to ‘a lot’) is, in principle, more straightforward for them to answer, than a question that asks them how healthy a product is in that nutrient, or how likely it is that the product has a healthy
content of that nutrient. The former question should have an objective answer, whereas the questions relating to healthiness requires an assessment of the impact on health of the amount eaten of a particular nutrient, and will also be subjective in terms of how each individual interprets the word ‘healthy’. The question should be as simple and straightforward as possible to ensure consistency of response.

The key to any question that asks consumers to evaluate a product with regard to an individual nutrient is to ensure that they can use signposting information in any format. This means the question must not rely solely on the use of % GDA, and the scale used for responses should not be divisible by 3 (as this would be easier for TL or text-based signposting than for GDA).

On this basis, it would seem sensible to include a question derived from Test 3.7, which asks simply “using the information on this pack would you say this product is high, medium or low in ...”, amended to use a scale that did not favour TL or text-based signposting. A scale similar to that used in Test 3.8 may be more suitable (a five point scale from a lot to hardly any). We recommend a scale of 5 points, since this is not divisible by 3, has a mid point (useful for analysis purposes) and people tend to find it harder to use scales with more than 5 points.

It should be noted that in order to evaluate a label fully, this question would need to be repeated for each nutrient. This is likely to be time consuming and too repetitive for respondents. We suggest restricting this question to two nutrients per FOP label, which have been selected in a way that does not bias the tests towards particular nutrients (see Section 5.3), or allow respondents to accurately guess the level from the name of the product alone, as this would not be a true test of label comprehension.

4.5.4 Evaluating products in terms of overall ‘healthiness’

Tests that evaluate overall healthiness of a product (Table A4, Appendix A) are not prone to favouring particular types of FOP signposting schemes. The scales used vary from how wise a choice for a healthy diet, to how nutritious, to how healthy a product is. The scale lengths also vary from seven to ten point scales. Since these questions only need to be asked once per FOP label this is unlikely to fatigue respondents.

All of the questions used in previous research fundamentally ask respondent to evaluate the product on a particular dimension using a multi-point scale. As for single nutrient evaluations, this should avoid multiples of three to avoid bias towards TL or text-based signposting.

In the context of this research, the key is for people to use the FOP labels to eat healthily. It would seem logical to use a scale of healthiness, with some explanation of what this means. To be eating healthily the Government advise
that people reduce the level of fat, saturated fat, salt and sugars in the foods they eat. We would again recommend a five point scale.

4.5.5 Comparing products on individual nutrients

All of the questions in Table A5 (Appendix A) ask respondents to compare two products in terms of individual nutrients. This tends to be looking for differences (essentially, which is higher or lower) although one test also requires a calculation based on serving size. This conflates two tasks and seems overly difficult. Two of the tests require the respondents to look at all of the nutrients in a product, and return an answer which identifies all of the nutrient differences in one response. This would be difficult to explain to respondents. It would be simpler to ask them to look at each nutrient in turn, as in Test 5.2. The answer scale used (product A, Product B, or neither) is the obvious scale for a comparison question.

This test is only one step beyond a replay measure, requiring respondents to extract the same piece of information from two labels, and compare the two to see which is higher. This would not necessarily require the use of signposting (it could be done using amount of nutrient alone), although signposting may make the task easier in some cases (e.g. different traffic light colours on two FOP labels or different %GDA’s on two FOP labels). Even with a relatively low level of numeracy, most people should be able to identify the correct answer, and the only differentiator is likely to be the time taken to respond. Given that this measure is unlikely to differentiate greatly between the different FOP labels, we do not feel it is a high priority for inclusion.

If this question was to be included, we would recommend using a simple comparison format such as Test 5.2 (‘which product is higher in a particular nutrient’).

This test would, of course, need to be repeated for more than one nutrient but, as for tests looking at individual nutrients on an individual product, we would recommend restricting the number of nutrients asked about on any individual label to two to avoid respondent fatigue.

4.5.6 Comparing products on overall ‘healthiness’

We have only identified two examples that have been used in previous research (Table A6, appendix A) in terms of overall product comparisons. One asks for a comparison of two products in terms of healthiness, and the other in terms of which is the better choice based on specific dietary instructions.

Given the recommendation of a healthiness scale for the individual product evaluation, we would suggest using healthiness for this comparison, again accompanied by an explanation of what we mean by healthiness in this situation.
As for individual nutrient comparisons, a scale of product A, product B or neither seems appropriate.

4.6 **Suggested tests**

Following the discussion above, Table 7 sets out the tests from the literature that we suggest using as a basis for this research. It also suggests refinements that would be needed to meet the specific needs of this project. There are four possible tests, as we discounted replay and computation tests in the above discussion.

**Table 7 – suggested tests and refinements**

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Test</th>
<th>Source</th>
<th>Refinement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation of single nutrient in single product</td>
<td>Test 3.8 - And thinking about your Guideline Daily Amount (GDA) for saturates/saturated fat [sugar], if you were eating one serving of this product as a main meal, do you think this amount of saturates/saturated fat [sugar] is...? Scale used: a lot; quite a lot; a reasonable amount; not very much; hardly any</td>
<td>3.8 - Institute of Grocery Distribution (2005)</td>
<td>Need to revise question wording not to favour GDA (change to 'using the information on this label' or similar), and consider scale wording to be more clearly an evenly spaced 5 point scale. Consider whether want to exclude wording &quot;as a main meal&quot; etc given the choice of product categories and possible multiple function of smaller portions.</td>
</tr>
<tr>
<td>Evaluation of overall product</td>
<td>Test 4.5 – Based on the picture and description you have just read, how would you rate the following product in terms of overall healthiness? Scale of 1-7 from extremely healthy to not at all healthy.</td>
<td>Cereal Partners worldwide (2006) (although all of the tests in Appendix A are very similar)</td>
<td>Simplify question to go with generic label e.g. “how healthy is this product”, develop 5 point scale</td>
</tr>
<tr>
<td>Comparison of products by single nutrient</td>
<td>Test 5.2 - Using the information on the packs, which, if any, of these products would you say has a higher ............ (insert Nutrient) content or would you say they are both the same? Scale used: x has higher content; y has higher content; Both the same; Can’t tell; Don’t know Scale used: Immediately (within 2 seconds); Within 5 seconds; Within 10 seconds; Within 20 seconds; Longer than 20 seconds)</td>
<td>FSA (2005b)</td>
<td>Adapt question for use with generic FOP label ('using the information on this label'). Need to consider whether to include ‘don’t know’ option⁵.</td>
</tr>
<tr>
<td>Comparison of products (overall)</td>
<td>6.2 Which of these foods do you think would be a wiser choice for a healthy diet? Scoring: Participants initial response and time took to make it was recorded and any reasons given.</td>
<td>Black and Rayner (1992)</td>
<td>Change wording to simpler healthiness judgement. Use same answer scale as single nutrient comparison test.</td>
</tr>
</tbody>
</table>

⁵ For all tests if a 'don’t know' option is included, a decision is needed on whether to mark this as incorrect, or not answered.
4.7 Judging the correct answers

All of the above tests require predefined correct answers. Criteria used in the tests included in the review were scheme-specific (e.g. Smiley face labels) but no information on the rationale behind these criteria was available.

For the purpose of this study examples of real life products in each product category have been identified. There is no pre-defined criteria to use to allocate each nutrient on a five point scale of high to low, and there is no pre-defined criteria as to what constitutes a healthy product. It was decided to conduct a survey of professional nutritionists and dieticians to generate data based on their individual assessments about the level of each nutrient, the overall healthiness of each product and the relative healthiness of selected pairs of products.

The responses to this survey will be collated and statistically analysed. Where there is a sufficient consensus among a majority of nutritionists about a product (or about nutrients in a product) this will be eligible for inclusion in the study. If there is no clear agreement, the product will be excluded. The consensus opinion will be used to provide the “correct” answer to each test (in the case of 5 points scales, this may be two adjacent answers).

Respondents in the main study will then be marked according to these answers on each test presentation. On all tests they will be given a mark of 1 points for each correct answer, and 0 point for any other answer. A ‘don’t know’ response will be considered equivalent to an incorrect answer. It was agreed that the scales used for this task would match those to be used in the main consumer study as closely as possible.

4.7.1 Speed of response

A number of tests included a measure of the time needed to perform the task. This is assumed to reflect the information processing effort required by the format (Russo, 1987). A measure of speed of label use for all objective tests of understanding is included as part of the research design; this will involve timing how long respondents take to provide an answer at each test.

4.8 Refining the tests

A review of the literature and the initial qualitative research has allowed us to produce a draft set of tests for the research. There is, however, little evidence from the literature on whether these tests measured what they were intended to measure. Furthermore, we have amended these tests to reflect our research aims. We have, therefore, conducted a stage of cognitive testing to trial these
tests (see Chapter 7 for details and annex for results). This is an approach that is regularly used in research to ensure that the questionnaire design addresses the research aims.

The aim of this stage was to determine exactly how respondents interpret and respond to each of the planned tests and a few other questions to assess what question wording will best provide the required measures of comprehension. Further details of the methods used are given in Chapter 7.

The wording used at this testing stage is given below. For each one a FOP label (or pair of FOP labels) was also shown:

**Introduction** (same for all tests):

In each of the next 10 questions you will see one or more labels. These are the types of labels you get on the front of food packaging. For each food label you see, you will be asked a question. Use the label information to help you answer the question. We want you to think about each food generally, from the point of view of someone who wants to eat in a healthy way and not necessarily in terms of your own eating habits or health needs. Tell the interviewer if you really don’t know which answer to choose, but try and use one of the answers on the questionnaire if you can.

**Test 1:** If you were eating one portion of this product, do you think the amount of (nutrient) is...? Please choose a number from 1 to 5 on the scale below where 1 is a little and 5 is a lot.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A little</td>
<td></td>
<td></td>
<td></td>
<td>A lot</td>
</tr>
</tbody>
</table>

**NOTE** – we also tested out the word ‘hardly any’ for point 1, and probed for descriptions for points 2 to 4.

**Test 2:** How healthy do you think this food is? Please choose a number from 1 to 5 on the scale below where 1 is very healthy and 5 is very unhealthy.

NOTE - to be eating healthily the Government advise that people reduce the level of fat, saturated fat, salt and sugars in the foods they eat.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very healthy</td>
<td></td>
<td></td>
<td></td>
<td>Very unhealthy</td>
</tr>
</tbody>
</table>

---

NOTE – we also probed for descriptions for points 2 to 4. We asked some early test questions without the accompanying note to see if this changed people’s approach to the tests.

Test 3: Which of these two products do you think would be the healthier choice?

1. A
2. B
3. No real difference between A and B

This work was carried out whilst the scientific rationale document was being finalised for publication. Full results of the cognitive stage together with recommendations for the final test wording are included in an annex to this report. Further details of the method used are given in Chapter 7.

4.9 Summary of objective measurement of understanding decisions

Existing tests have been modified to meet the specific needs of the study which is to measure comprehension.

Three measures of comprehension have been identified as being likely to discriminate best between the different types of FOP signposting schemes, and to reflect the most common uses of FOP labels. These are:

1. Evaluation of products on individual nutrients (restricted to 2 nutrients per product)
2. Evaluation of product in terms of overall healthiness
3. Comparison of two or more products in terms of overall healthiness

A fourth measure was identified as reflecting current use of labels, but was felt to have less capacity to differentiate between signposting schemes:

4. Comparison between products on single nutrient levels (restricted to 2 nutrients per product)

Tests used in the literature were evaluated, and four potential tests identified (one for each task), with suggested refinements to ensure they do not favour any signposting scheme. These have been further developed in a cognitive testing phase to ensure they are suitable measurement tools.

It is vital that all tests have a predefined correct answer against which to judge respondents for correctness. There is no clearly defined process for producing a totally objective measure. FSA will, instead, conduct a survey of professional
nutritionists and dieticians to generate data to define the correct answer for selected products for each of the tasks. Only products where a suitable consensus can be reached will be included in the survey.

A second measure to be used is time taken to respond. The time taken to respond to each individual test presentation will be recorded during the interview.

5 The design for the comprehension tests

The discussions in Chapters two to four produced a prioritisation of FOP label elements, and tests for inclusion in the design, and a decision on the context within which FOP labels would be presented, and the range of products that would be included. These needed to be combined into a workable design. This design is constrained by practical considerations of administering the survey, avoiding respondent fatigue, and realistic budgetary constraints on the size of the survey.

There is more discussion of the survey method in Chapter 7. This presents two main parameters, within which the design had to fit:

- A maximum of 20 minutes would be spent on tests for any one individual. Piloting work revealed that most people could respond to 20-25 test presentations within this time and without excessive fatigue. The number of tests does depend on the type of tests, as some take longer to respond to than others.
- A sample size of 3000 respondents, with a minimum of 750 respondents completing each test, meaning the sample could be split into up to four groups (each randomly selected, and therefore representative of all shoppers). This meant a maximum of 100 test presentations could be included in the survey (25 per group).

5.1 Establishing priorities of label elements, tests and products

Table 8 shows the priorities that were established.

Table 8 – Established priorities

<table>
<thead>
<tr>
<th>Priority</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label elements</td>
<td></td>
</tr>
<tr>
<td>PRIORITY LEVEL 1</td>
<td>% GDA / no % GDA</td>
</tr>
<tr>
<td></td>
<td>TL signposting / no TL signposting</td>
</tr>
<tr>
<td></td>
<td>Text / no text</td>
</tr>
<tr>
<td>PRIORITY LEVEL 2</td>
<td>Energy / no energy (in relation to overall product judgements)</td>
</tr>
<tr>
<td>PRIORITY LEVEL 3</td>
<td>Direction of presentation (horizontal / vertical / circular)</td>
</tr>
<tr>
<td></td>
<td>Use of non-TL colour / no use of non-TL colour</td>
</tr>
<tr>
<td>Products</td>
<td></td>
</tr>
<tr>
<td>PRIORITY LEVEL 1</td>
<td>P1 – full meal products / P2 – small portion/snack products</td>
</tr>
</tbody>
</table>
Tests

<table>
<thead>
<tr>
<th>PRIORITY LEVEL 1</th>
<th>Test 1 – evaluate level of individual nutrient in a product – repeat for 2 nutrients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test 2 – evaluate healthiness of individual products</td>
</tr>
<tr>
<td></td>
<td>Test 3 – compare healthiness of two products</td>
</tr>
<tr>
<td>PRIORITY LEVEL 2</td>
<td>Test 4 – compare level of individual nutrient between 2 products</td>
</tr>
</tbody>
</table>

Taking just priority level 1 into account, this would produce eight types of label (2X2X2) in a full-factorial design of the label elements. These would each need to be presented for P1 and P2 giving 16 labels.

These 16 would each need to be presented for the selected tests. In order to judge how many tests this would mean presenting, we need to make a judgement as to how many ‘question units’ each represented. We use test 2 (asking to judge the overall healthiness of one label) as our base unit.

**Table 9 – Test length in terms of question units.**

<table>
<thead>
<tr>
<th>Test</th>
<th>Assessment</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRIORITY 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test 1</td>
<td>Asking to judge the level of one nutrient on one label. Then asking to judge a second nutrient on the same label. This does not require familiarisation with a second label for the second nutrient, meaning this is less time consuming than two full questions.</td>
<td>1.5 units for two nutrient question.</td>
</tr>
<tr>
<td>Test 2</td>
<td>Base level question defined as 1 unit</td>
<td>1 unit</td>
</tr>
<tr>
<td>Test 3</td>
<td>This requires people to evaluate two labels before making a judgement, but only requires one judgement, so was felt to lie between test 1 and test 2 in likely time taken</td>
<td>1.25 units</td>
</tr>
<tr>
<td>PRIORITY 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test 4</td>
<td>This requires people to look at two labels, although only at one nutrient per label, and then repeat this task for a second nutrient. This was, however, a simple replay and comparison task, requiring no judgement to be made about levels, so it was felt that this would take a similar amount of time to Test 1.</td>
<td>1.5 units for two nutrient question</td>
</tr>
</tbody>
</table>

Presenting the 16 labels identified for the three priority tests would mean a total of 3.75 question units per label which is 60 question units in total (16 X 3.75).

We have identified a maximum of 100 units, meaning that some of the lower priority elements or tests could be included.

Since Test 4 is essentially a replay test, and is less likely to differentiate between the different signposting methods, this was established as a lower priority than the label element priorities. It was, therefore, discarded from the design.
The presence of energy or not was judged to be the next priority, given the evidence from the qualitative work that it is used as a proxy for healthiness. Including it in the full factorial design for all three tests would double the number of labels to 32 and the number of tests to 120, which exceeds the limit. However, the hypothesis under consideration is that energy will only come into play in overall judgements of products (it should not be a factor in judging the level of a single nutrient). This means it does not need to be included for test 1. This brings the number of tests down by 24 (1.5 X 16) to 96.

For other tests where the presence or not of energy was not being tested, energy will be held constant as present as this is the norm on most labels in the marketplace.

It is desirable to include the third priority of presentation and non-TL in some way. This can not be achieved if energy is included as an element in both Tests 2 and 3. Chapter 2 presented the possibility of including just two additional labels approximating to the Sainsbury label and the Tesco label on at least some of the tests (see Section 2.5 for discussion). In order to make room for this, the effect of energy on Test 3 could be controlled for if the number of calories was kept constant within each pair of labels: if both products under comparison have the same energy content, then it should not affect the judgement of which product is healthier.

The design was taken forward on the basis described above and the full array of test presentations is set out below (Table 10).

**Table 10 – Full array of test and label combinations, showing numbers of question units**

<table>
<thead>
<tr>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
<td>P2</td>
<td>P1</td>
</tr>
<tr>
<td><strong>TL</strong></td>
<td><strong>Text</strong></td>
<td>GDA</td>
<td>L1</td>
</tr>
<tr>
<td></td>
<td>no GDA</td>
<td>L2</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>GDA</td>
<td>L3</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>no GDA</td>
<td>L4</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>no text</strong></td>
<td></td>
<td>GDA</td>
<td>L5</td>
</tr>
<tr>
<td></td>
<td>no GDA</td>
<td>L6</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>No TL</strong></td>
<td><strong>Text</strong></td>
<td>GDA</td>
<td>L7</td>
</tr>
<tr>
<td></td>
<td>no GDA</td>
<td>L8</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>no text</strong></td>
<td></td>
<td>GDA</td>
<td>L9*</td>
</tr>
<tr>
<td><strong>TL</strong></td>
<td><strong>no text</strong></td>
<td>no GDA</td>
<td>L10*</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>

* - has non-TL colour – label approximates to that used by Tesco
# - has circular presentation – label approximates to that used by Sainsbury
At Test two, the columns headed E(P1) and E(P2) are those without energy present, with all other columns holding energy constant as present. The labels are called L1 to L10 as shorthand. L9 is the label equivalent to the Tesco label (this was label 15 in Table 3), and L10 is equivalent to the Sainsbury label (this was label 36 in Table 3). As explained in Section 2.5, by including these two labels in this way, we can compare comprehension of L9 with L7, and L10 with L4. Each label differs only by a single element (L9 is L7 with non-TL colour, L10 is L4 but with circular presentation). This means direction of presentation and non-TL colour are not included as part of the full-factorial design, but this still allows some evaluation of the effect of a circular presentation, and of non-TL colour just in the ways they are currently used in the marketplace. We cannot fully isolate the effect of these presentational issues, however, from Text, GDA, or TL colour.

As explained in the introduction, the ideal basis on which to test our hypothesis that the different label elements have an effect on comprehension, is to use a ‘within respondent’ design, i.e. presenting each test on all 10 label types to the same person. In effect, this means each column in Table of 10 becomes a set of tests that cannot be split up.

We would ideally also use a within respondent design to look at the effect of product on comprehension. This would mean within each test, the P1 and P2 columns would need to be presented to the same respondent. This causes a problem for test 1 as this would involve presenting 30 test units to a single respondent, which is above the set limit of 25. To keep this within the limit, it was decided to remove L9 and L10 from Test 1, as they were of lower priority.

For Test 2 there are 40 test units. These would clearly need to be split into 2 sets of tests for the survey (20 tests to each group of respondents). These could be split to give a within respondent design by product as for the other tests. This would mean one set of respondents getting test 2 for P1 and P2 with energy, and the other getting the tests for P1 and P2 without energy. Alternatively, they could be split to give a within respondent design for energy – one set of respondents getting all P1 products with and without energy, and the other getting all P2 products with and without energy. There is a third way of dividing up the tests (one group gets P1 with energy, and P2 without energy) but since this would not allow either energy or product to be looked at within respondents it was rejected.

There was no obvious rationale for prioritising either energy or product. There were, however, three arguments for prioritising product as the within respondent element:

- We plan to use a within respondent design on product for Tests 1 and 3
- Prioritising product means respondents will see a greater array of products, which should alleviate boredom.
The same product examples can be shown with and without energy within the survey, which gives a more balanced comparison. If one person saw labels with and without energy (a within respondent design for energy), we could not show the same products for each, as they may use their memory of seeing the product with energy to inform their evaluation when they see the label without energy.

On this basis, it was decided the within respondent design should be used for product. Energy could be assessed using a between respondent design. This approach is valid as long as each group of respondents is equivalent (see Chapter 7 for an explanation of how this will be achieved).

5.2 Final test design

The decisions described above give the test design outlined in Table 11. This shows the number of test questions that will be presented to respondents (rather than the numbers of question units this equates to in the questionnaire as shown in previous tables).

Table 11 – Final test design (with number of actual test questions shown)

<table>
<thead>
<tr>
<th>Group of respondents</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Test 1</td>
<td>Test 2</td>
<td>Test 2</td>
<td>Test 3</td>
</tr>
<tr>
<td>P1</td>
<td>P2</td>
<td>P1</td>
<td>P2</td>
<td>P1</td>
</tr>
<tr>
<td>GDA</td>
<td>L1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no GDA</td>
<td>L2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>GDA</td>
<td>L3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no GDA</td>
<td>L4</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>TL</td>
<td>Text</td>
<td>GDA</td>
<td>L5</td>
<td>2</td>
</tr>
<tr>
<td>no GDA</td>
<td>L6</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>GDA</td>
<td>L7</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>no GDA</td>
<td>L8</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>No TL</td>
<td>no text</td>
<td>GDA</td>
<td>L9*</td>
<td>1</td>
</tr>
<tr>
<td>No TL</td>
<td>no text</td>
<td>no GDA</td>
<td>L10*</td>
<td>1</td>
</tr>
<tr>
<td>TL</td>
<td>TL</td>
<td>no text</td>
<td>GDA</td>
<td>L10*</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* - has non-TL colour – label approximates to that used by Tesco
# - has circular presentation – label approximates to that used by Sainsbury

5.3 Allocation of product examples

As explained in Section 3.2, we plan to include three examples of each of P1 and P2 to provide some variety for respondents, in order to minimise fatigue. It is important that the selection of product for each test presentation does not bias the testing. For example, if (say) it is generally easier to make a judgement based on crisps, than on yoghurt then, if any label is only presented for crisps, it
will have an unfair advantage, compared to any presented just for yoghurt. At the next level, the judgement may be easier for one type of yoghurt, then for another. This means that a label presented for an ‘easy judgment’ yoghurt will be at a disadvantage, compared with the one presented for the more ‘difficult judgment’ yoghurt.

In order to overcome this problem, within each of P1 and P2, a set of specific products will be selected and each FOP label will be presented for the same set of products. For test 1, where there are 8 FOP labels, 8 products will be selected (for each of P1 and P2). For test 2 where there are 10 FOP labels, 10 will be selected (for each of P1 and P2). For test 3 where there are 10 FOP labels, 10 pairs of products will be selected (for each of P1 and P2).

The time available in the study only allows each FOP label to be presented once for each of P1 and P2. This means finding a way to randomise the presentation of product. The best way to do this is to split each of the four groups (of c.750 respondents) into 10 further groups (8 for test 1) and ensure each group sees the products and labels in a different one of the possible combinations. All possible combinations will be included. This produces a split into 38 different groups of respondents. Respondents will be allocated to one of these groups at the sampling stage to ensure they are evenly spread (allocated first to one of the four groups, and then within these to one of the smaller groups)\(^7\). For the final analysis, these groups will be recombined into the original four groups. There is no intention to analyse by the 38 groups (the sample size would be too small, in the region of 75 respondents).

The choice of product examples was discussed in Chapter 3. The product allocations to be used are reproduced in the Appendix to this document. This allocation means selecting at three or four different products for each of product examples A through to F (20 products in total). These will be real products, but will be restricted to products for which we have a suitable consensus on the correct answer from the nutritionists’ survey (see Section 4.7). In addition, further products may need to be selected to make up pairs at test 3. It is possible to make a full set of pairs from the 20 products selected, but this would require every pairing to have a consensus answer at test 3 (i.e. one is clearly healthier) and this is not guaranteed. We will select further products as necessary to make up suitable pairings. For the paired products we will hold energy constant (as discussed earlier) and we will use the mean amount of energy for the two products.

\(^7\) The alternative was to use the computer programme to randomly allocate the label/product combination and this is less likely to fall out evenly, as it would be done separately on each interviewer’s computer, rather than centrally across the whole sample.
Since we are selecting a range of products within each type (by type we mean A to F), we will select a range of difficulty levels (easier to harder judgments) within each type.

5.3.1 Allocation of nutrients at test 1

At test 1 each label will be presented twice to respondents, each time asking about a different nutrient. It is important that within this design each nutrient is allocated an even number of times, and the same number of times for each of P1 and P2. As far as possible we will also make sure that for each of Labels 1 to 10, a respondent will see the same nutrients for P1 and for P2. This is not essential as rotating the products within the design should ensure balance, but it will add to the robustness of the data. It is not possible to match nutrients at P1 and P2 completely as the different foods selected do not all vary by all nutrients (e.g. yoghurts do not vary by salt).

We have allocated nutrients as follows at test 1. This gives a totally even distribution of nutrients, and the same number of each nutrient allocated to each of P1 and P2. It also matches the nutrients shown for P1 and P2 at the same label (the products shown are always shown in the same pairs e.g. A1 and D1) in all but two cases (highlighted) because P2 product, yoghurts, do not vary by salt but their partner P1 product, ready meals, do.

<table>
<thead>
<tr>
<th>P1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1  a  c</td>
<td>D1  a  c</td>
</tr>
<tr>
<td>A2  a  d</td>
<td>D2  a  b</td>
</tr>
<tr>
<td>A3  b  c</td>
<td>D3  b  c</td>
</tr>
<tr>
<td>B1  c  d</td>
<td>E1  c  d</td>
</tr>
<tr>
<td>B2  a  b</td>
<td>E2  a  d</td>
</tr>
<tr>
<td>B3  b  c</td>
<td>E3  b  c</td>
</tr>
<tr>
<td>C1  a  d</td>
<td>F1  a  d</td>
</tr>
<tr>
<td>C2  b  d</td>
<td>F2  b  d</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P1</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ready Meal</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soup</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandwich</td>
<td>C n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yoghurt</td>
<td>D n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breakfast Cereal</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crisps</td>
<td>F n</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NUTRIENT</th>
<th>P1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat</td>
<td>a</td>
<td>4 4</td>
</tr>
<tr>
<td>Saturated fat</td>
<td>b</td>
<td>4 4</td>
</tr>
<tr>
<td>Sugar</td>
<td>c</td>
<td>4 4</td>
</tr>
<tr>
<td>Salt</td>
<td>d</td>
<td>4 4</td>
</tr>
</tbody>
</table>

5.4 Label appearance

The design above specified the elements of the labels that would vary between the different label types. Further decisions were needed on how to present the other aspects of the label that would remain constant. This was to be based on the most common use in the marketplace.

The context within which the label was presented had already been decided (see Section 3.1). The labels were to be reproduced generically, with just a generic
product name, weight and number of portions shown, with the FOP label showing information per portion.

Decisions were made as follows:

- **Direction** will be horizontal (except for Label 10) as the most used approach in the marketplace, and the approach best received in the qualitative work.
- Nutrients will be presented in the FSA recommended order of fat, saturated fat, sugar then salt from left to right (clockwise in Label 10) in the absence of any clear industry consensus.
- The **shape** of nutrient box will be a rounded rectangle, best approximating the most commonly used shapes.
- **% GDA** (when present) will be included at the bottom of box below a line – Just below the label will appear the words “of your guideline daily amount”. This reflects current use in the marketplace.
- **High, medium, low text** (when present) appears at the top of the nutrient box, above a line, to ensure it is no more or less visible than GDA in the generic labels.
- The **monochrome** nutrient box background (i.e. when no TL) will be white to avoid any bias from the colour chosen (the qualitative found that different colours were interpreted differently).
- **Product description** will be generic to avoid giving cues (e.g. ready meal, not “macaroni cheese”).
- **Energy** (when present) will be presented to the left of other nutrients as this is the most common placement in the marketplace. It will be presented with signposting (GDA, text, TL) as most labels signpost energy.

The last point means the need to develop criteria for the purpose of the study for signposting energy using TL. As there is no recommended criteria for this interpretive element, the criteria will reflect what is being used in the marketplace by retailers. These criteria are included in Appendix C, together with the criteria currently in use by FSA for signposting the nutrients on FOP labels.

Examples of the final FOP label formats to be used are given below.
5.5 Presentation of tests

Within the interview, the tests (labels and questions) will be presented on a computer screen. The labels will appear at a consistent size throughout. The questions will be completed by the respondents using self-completion wherever possible to avoid any interviewer bias. If a respondent is unable to type for themselves, the interviewer will enter the responses for them. The tests will be presented in a random order to avoid any ordering effects. The time take to complete each test question will be recorded, and the task will end when all tests have been completed, or after 20 minutes (whichever is soonest).

5.6 Analysis of data

This section gives an overview of the analysis process to be used.

5.6.1 Fully factorial design (labels 1-8)

To gain a full understanding of the differences between the labels we will use the following approach to analyse for each comprehension measure (tests 1 to 3) for the labels in the fully factorial design (labels 1 to 8):
TOP LEVEL ANALYSES

1. Look at the percentages of correct answers for each of label types 1 to 8 to get an initial impression of which combination of TL, GDA and Text appears to achieve the best levels of comprehension on that measure.

2. Use logistic regression techniques to look at the individual and interaction effects of TL, GDA, Text and Product (P1/P2) on comprehension (whether or not correct answer was given). This will use a backward stepwise approach, starting with all of the possible factors and combinations included in the regression, and then removing any that are not significantly linked to comprehension and re-running the analysis, continuing until a satisfactory solution is reached.

3. Use ANOVA to look at the interaction effects of TL, GDA, Text and Product (P1/P2) on time taken to answer each question. This will be used alongside the results from the logistic regression described above to determine which label is most effective. Correctness is obviously the most important way of judging comprehension, but if two label types facilitate similar levels of comprehension, then the label that facilitates the quickest answer will be judged more effective.

SECONDARY LEVEL ANALYSES

1. Look at percentages of correct answers within selected demographics and behavioural variable (see Chapter 6 for potential variables for inclusion). This stage can also be used to test specific hypotheses (e.g. numeracy levels will impact differently on ability to understand GDA and TL).

2. Add those demographic/behavioural variables that appear to impact on comprehension to the data and re-run the logistic regression model to look at any interaction between demographics and comprehension of the eight label types.

This approach will be used for each of the following:

1. Test 1 using the overall test data. Each respondent will have completed 2 tests for each label for each of P1 and P2.

2. Test 1, evaluated separately for each of the four nutrients (fat, saturated fat, salt and sugars) included in the tests.

3. Test 2 for the labels shown with energy.

4. Test 2 including energy/no energy as an additional variable (this is the one analysis that will be conducted between rather than within
respondent, as different respondents were asked the questions with, and without energy).

5. Test 3.

5.6.2 Labels 9 and 10

In order to evaluate label types 9 and 10 (see earlier discussion for limitations on this) the following approach will be taken:

1. Compare the proportion of correct answers for each label with the label that differs from it in on one dimension (label 9 paired with label 7, and label 10 paired with label 4) to assess any obvious differences in comprehension.

2. Use logistic regression for the results for each pair of labels to look for the impact of the single differentiating factor (direction of presentation/use of non TL colour) on comprehension and any interaction this has with product (P1/P2).

This will be carried out for each of:

1. Test 2 for the labels shown with energy.

2. Test 2 including energy/no energy as an additional variable.

3. Test 3.
## 5.7 Summary of design decisions

### Label content:

Three elements of the label content will be used in a full-factorial design, enabling us to identify the influence on comprehension of each element individually, and any effects of interaction between them. The elements are:

1. %GDA
2. TL
3. Text (high, medium, low)

All possible combinations of these three elements produce 8 labels. In addition, we will test two further label types, giving a total of 10 labels:

1. No text (high, medium, low), No TL, %GDA shown, nutrient specific colour (approximates to Tesco label)
2. TL colour, no text, no %GDA, circular presentation (approximates to Sainsbury label)

These will be compared against labels that only differ by one element (nutrient specific colour for the first, circular presentation for the second). This will enable the effect of this element to be assessed, but only in the format they are currently used in the marketplace. We will not be able to assess whether there has been any interaction between these elements and GDA, TL or Text (high, medium, low).

The impact on comprehension of the label elements will be carried out within respondent, with each respondent presented with all 10 labels for any test they carry out.

### Energy

In addition, energy (calories present or not) will be included as a factor, but only for overall single product tests. This will allow us to test the hypothesis that energy will influence comprehension on overall single product judgements. For all other tests energy will be held constant as present. For comparison tests (test 3) the average calorie value of the two products shown will be used.

This will be tested between respondents, with one set of respondents seeing the labels with energy, and a second set seeing those without. Random selection of respondents will ensure that the two groups are fully comparable for this purpose.

### Tests:

Three test types will be used:

- **Test 1. Evaluation of level of single nutrient on one product (2 nutrients tested per product label)**
Test 2. Evaluation of overall healthiness of one product
Test 3. Comparison of overall healthiness of two products

Each test will be shown for all 10 labels except for test 1 which will only be shown for the 8 labels in the full-factorial design.

**Label appearance**

All elements of the label not under test will be held constant as follows:

1. Label generically reproduced
2. Horizontal direction (except for label 10)
3. FSA recommended nutrient order
4. Nutrient information within rounded rectangle. GDA at bottom, Text (high, medium, low) at top; white background where no TL (except for label 9)
5. Generic product name, with weight and number of portions. Label shows information per portion.
6. Where energy is present, this will be presented to the left of other nutrients, and signposted.

**Product:**

Each test will be presented for two products categories to each respondent – P1 (a meal sized product portion) and P2 (smaller portion or snack products). This will also be tested within respondent, with each respondent presented with each label for both P1 and P2 products for the test in question.

Three product examples will be used within each category. P1 will be represented by soup, ready meals and sandwiches. P2 will be represented by breakfast cereals, fruit yoghurt and crisps. These are included to provide variety for respondents and the effect on comprehension of these different products will not be tested.

**Test administration**

The design gives a total of 92 test presentations. These will be split into four groups of tests, each going to a randomly selected quarter of respondents. Each group will see the full range of labels for each test, for both P1 and P2. Within each group, the product shown with each label type will be rotated, and the order in which the tests are shown will be randomised. This will avoid any effects from ordering or product selection. At Test 1, where possible, the same two nutrients will be asked about for P1 and P2 within each label type.

Tests will be self completion, administered on a laptop computer. All tests will be timed individually, with a maximum of 20 minutes spent on tests.
**Test analysis**

For the eight labels included in the fully factorial design, logistic regression will be used to look at individual and interaction effects of TL, GDA, Text and product on comprehension (i.e. the ability to give correct answer).

As a secondary measure ANOVA will be used to look at effects of TL, GDA, Text and product on the time taken to give an answer. This will be used to help distinguish between label types with very similar levels of comprehension.

In addition, logistic regression will be used to establish which key demographic and behaviour variables have an impact on comprehension of the eight label types.

Logistic regression will also be used to compare label 9 with label 7, and label 10 with label 4, to establish any impact of using a circular rather than horizontal direction of presentation, and of the use of non-TL colour (but ONLY within the formats in which these are currently used in the marketplace).
6 Other factors influencing comprehension

The remainder of the interview will be used to collect information on factors that may influence comprehension (these are factors outside of the labels themselves). The literature on previous work and the qualitative work carried out for this study suggested a number of areas for inclusion. These are discussed below.

There is limited room within the questionnaire for questions, since up to 20 minutes will be spent on the test, with additional time taken in introducing the test and the self completion method. The aim is to achieve an average length of 30 minutes to minimise respondent burden.

6.1.1 Demographics

Generally, research has shown that older age groups and those in lower social classes have performed less well on indicators of objective understanding (FSA, 2004a). Which? (2006) found that the lower socioeconomic groups had more difficulties classifying foods as high, medium or low in a particular nutrient when signposted using a % GDA scheme, whereas this effect was not seen for TL. Likewise, in a study examining daily reference information (similar to % GDA), Li et al. (2002) found those with higher levels of knowledge about nutrition labels performed better on objective measures. However, subjective understanding was not found to be influenced by level of education by Feunekes et al. (2008).

The qualitative findings (phase 1) found that there appeared to be no correspondence between any of the key demographic variables (see annex), including educational level, and understanding of FOP labels. It did, however, suggest that Asian respondents rarely used FOP labels because they were often cooking from scratch and the recipes determined what the meal contained, so they were less concerned with actual nutrient values of ingredients or final products. The same research found that, in the interests of children’s health and development, people with young families often looked for products that were ‘free from additives and preservatives’ and/or were lower in salt and sugar.

We propose to collect information on the following demographics as each may influence comprehension:

- Age
- Sex
- Household composition (numbers and status of adults and children, age of children)
- Ethnicity
- Terminal education age
- Social grade
- Postcode (to allow matching to geodemographic information)
- Colour blindness
The exact questions will be selected to ensure they will be useful for targeting any future marketing of FOP labelling to ensure the results can be used to communicate effectively with consumers. This is the reason for collecting social grade details (the mainstay of advertising) rather than alternatives such as NS-SEC (the social grading used by the Office for National Statistics). Where possible data will be collected in a flexible way (e.g. actual age rather than age bands, and building up household composition from different elements, rather than pre-set categories).

The survey design (see Chapter 7) means that the ability to analyse by some subgroups may be restricted. For example, the number of respondents in minority ethnic groups is likely to be small in a national population survey. For such groups further qualitative work would be advisable.

6.2 Shopping and dietary habits of household

Shopping habits could have an influence, particularly familiarity with different FOP label types. We will therefore ask where people usually shop, including details of whether this is in store or online as this may also affect exposure to labels.

In addition, dietary needs may impact on what information consumers look for, so we will collect details of the dietary habits on the people the respondent shops for. The focus will be on dietary habits for which FOP labels could be useful, but this will be asked within a broader context to avoid biasing responses.

6.3 Attitudes to labelling and nutrition

Much of the previous work carried out into labelling relied on measures such as preference, which is a subjective measure. There is no evidence as to how closely this is linked to comprehension. There is some discussion of this in Grunert and Wills (2007). Simple preference measures will be collected to explore this further. The hypothesis would be that preference is not an indicator of ease of comprehension.

Exposure to labels may have an impact on comprehension, so we will also collect information on whether people are aware of the different types of FOP labels, and to what extent they already use them.

Attitudes to healthy eating may also be relevant, as this may impact on whether people use FOP labels or not. On this basis a number of questions about attitudes should be included.

6.4 Numeracy and literacy levels

It is important to gather some simple data on level of numeracy. We plan to precede the tests with a basic literacy test (this approach was taken in a previous
survey of Health Literacy conducted by BMRB) to avoid respondent embarrassment if they are unable to read the labels. This will be a test of the ability of respondents to extract and replay information from the labels. This gives them the chance to familiarise themselves with the kind of information they contain before the tests themselves. If they are unable to identify and replay the simplest information they will bypass the tests and just answer the other questions.

The qualitative work revealed that numerical confidence was related to reactions to the %GDA signposting. There are a number of numeracy tests available, but it is most important to use tests that are tailored to the use of FOP labels. The key issues identified (in relation to numeracy) in the qualitative work were the ability to use %GDA and the ability to use portion size information. This means a need to be able to test simple calculations (e.g. if one portion contains x, and there are y portions, how much does the whole product contain). It also means a question is needed to identify the ability to use a simple percentage (along the lines of ‘if one portion contains x% of your GDA, how many could you eat in a day’). This second question is complicated by ensuring people understand that GDA is the maximum you should eat in a day, and the question assumes you have no other source of the given nutrient. This will be tested further in the cognitive testing stage.

6.5  Question wording

Where possible, existing surveys will be used as sources of tried and tested questions. All questions have been tested in a small scale in-home pilot study. Key questions were also included at the cognitive testing phase to ensure they collect the information needed.

6.6  Summary of other factors influencing comprehension

Information on the following will be collected, as each may influence comprehension:

Key demographics:
- Age
- Sex
- Household composition
- Ethnicity
- Terminal education age
- Social grade
- Postcode (to allow matching to geodemographic information)
- Colour blindness

In addition we will collect information on:
Shopping habits
Dietary needs
Attitudes to labelling and healthy eating
Basic numeracy and literacy (in relation to food labels)

Questionnaire wording will be developed using previous research and experience, and tested using piloting and (for some key questions) cognitive testing.
7 Data collection and sampling method

This final chapter contains details of how the data will be collected, using the design described in previous chapters.

7.1 Data collection method

The need to show respondents FOP labels during the interview, together with the need to time the tests and achieve a robust sample meant that data needs to be collected using a face to face, interviewer administered survey. Interviewing will also be in-home, rather than a hall-test, to provide an adequately robust nationally representative sample (see Section 7.2).

7.2 Sampling

The need for a robust sample means that the survey needs to be conducted using Random Probability sampling. The underlying principle of Random Probability sampling is that all respondents have a known and equal chance of selection and that, unlike quota sampling, there is no scope for interviewer choice and hence bias. This is the highest quality method of sampling and is used where accuracy and representativeness is of greatest importance. It is commonly used in a variety of surveys conducted on behalf of Government departments for publication and target setting. Random Probability sampling is the only form of sampling where true measures of sampling error can be computed allowing the use of statistical testing (vital for this survey). To be effective, random probability samples require high response rates. In this case we are confident that a response rate in the region of 65% will be achievable, as the subject of the study is likely to be a subject of some interest to shoppers.

7.2.1 Who to interview

This will be a UK survey of those with the main responsibility for shopping in the household (defined as an adult aged 16 responsible for at least half of the food shopping). Whilst this could be carried out as a survey of all adults in the UK, the main use of labels is likely to be amongst those who shop for the household food, and particularly those responsible for preparing meals. Shoppers are also likely to be the target of any future communications about food labelling. A random probability sample with a high response rate will deliver a representative sample of main shoppers.

7.2.2 Sample size and coverage

We will interview 3000 shoppers in the UK, with this split into four equal samples of around 750 shoppers, each representative of the UK population of shoppers. This gives a confidence interval of around +/-1.8% for the whole sample, and around +/-3.6% for the quarter samples at the 95% confidence level.
The confidence intervals discussed above do, of course, assume a perfect random sample with 100% response. In reality, there will be some clustering in the sampling, and some potential non-response bias (likely to require minor corrective weighting). We will calculate the size of this design effect, and this is likely to marginally increase the confidence intervals.

These sample sizes are likely to deliver around the following numbers of people in different sectors of the population:

| TOTAL national | 750 | 3000 |
| SEX | | |
| Men | 227 | 909 |
| Women | 523 | 2091 |
| Social Grade | | |
| AB | 192 | 768 |
| C1 | 218 | 873 |
| C2 | 146 | 582 |
| DE | 195 | 780 |
| Age | | |
| 16-24 | 54 | 216 |
| 25-34 | 125 | 498 |
| 35-44 | 154 | 615 |
| 45-54 | 127 | 507 |
| 55-64 | 117 | 468 |
| 65+ | 174 | 696 |
| Country | | |
| England | 660 | 2640 |
| Scotland | 63 | 252 |
| Wales | 38 | 150 |
| Northern Ireland | 21 | 84 |
| Ethnicity | | |
| White | 691 | 2763 |
| Asian | 30 | 120 |
| Black | 15 | 60 |
| Chinese | 3 | 12 |
| Mixed | 9 | 36 |
| Other | 3 | 12 |
| Education | | |
| Did not finish school | 53 | 210 |
| Up to 16 | 278 | 1110 |
| Up to 18 | 90 | 360 |
| Further qualification | 158 | 630 |
| Degree or beyond | 143 | 570 |

This will not allow for robust analysis by country (other than England) or for minority subgroups of the population. It is more important to assess the different FOP label elements and aspects of use at a UK level. The best way to achieve this
is with a purely representative UK sample with no boosts or skews. Some key minority groups could be further explored qualitatively. Being able to analyse separately by all the demographic variables or by region is of lesser importance since the FSA has no plans to suggest tailored schemes for specific groups. Furthermore, the qualitative work did not suggest there would be key differences by region or ethnicity.

### 7.2.3 Sample selection

The sample of addresses will be selected in the following way:

1. Identify all Postcode Sector Units (PSU) Clumps in UK using the 2008 Postal Address File (PAF) (September 2007 release). Clumps are single postcode sectors or combinations of these to ensure a minimum PSU size of 500 addresses.

2. Sort PSUs by the 9 Government Office Regions (GOR) plus Wales, Scotland and Northern Ireland, and within the 3 smaller countries, by health region, then sort by population density. Create 3 equally sized bands within each GOR/country.

3. Sort within GOR by NS-SEC (social class)

4. Select 248 PSUs from the resulting list ensuring each of the region by PAF strata is represented

5. Select 20 addresses from each PSU using a random start and a fixed interval to produce 4960 addresses.

Each address would be allocated to one of the four testing groups before being issued to fieldwork. As described in Chapter 5, within the four groups, each respondent will be randomly allocated to one of 8 or 10 groups (8 for test 1, and 10 for tests 2 and 3). All random allocation will use a random start and fixed interval (1 in n) approach. This will ensure each of the 38 sample groups are equivalent.

We will identify and attempt to interview the main shopper at each of these addresses. Where there is equal responsibility for shopping between two or more adults, interviewers would make a random selection of whom to interview.

### 7.2.4 Maximising response rates

In order to minimise bias in the sample, it is important to achieve a high response rate. This will be achieved using the following techniques:
• All interviewers and supervisors involved in the survey will be fully briefed on the importance of achieving a high response rates, and strategies for doing so, prior to the start of the main stage fieldwork.

• At least five attempts will be made to make contact at each address. The attempts will be made on different days of the week and at different times of day, with a spread over the course of fieldwork, and appointments scheduled in line with respondents’ requests.

• A well designed advance letter on FSA letterhead paper will be sent to respondents in advance of the survey.

• In addition, details of the survey will be included on BMRB’s web page so that respondents can access this for reassurance about the bona fide nature of the study.

• Respondents will be offered freephone numbers that they can call to check the genuine nature of the survey: numbers for the research team, the research and field management team, and the Market Research Society helpline are available. We would also recommend a contact at FSA be named in the letter.

• If necessary, in order to achieve the target response rate those who are initially reluctant to take part in the survey (e.g. because they are too busy at the time), a supervisor or senior interviewer to make the final calls.

• As a matter of course, we monitor progress on surveys very closely as an essential prerequisite to ensuring that the survey remains ‘on track’ in terms of achievement versus response rate and productivity targets. This enables us to take swift remedial action in the event that problems do occur.

7.3 The questionnaire and questionnaire development

The testing part of the questionnaire was discussed in detail in Chapter 5 and the other issues to cover in Chapter 6. The structure of the questionnaire is as follows:

1. Literacy tests (to determine whether to administer tests)
2. Introduction to tests and self completion method
3. Tests (self completion)
4. Numeracy tests
5. Attitudes to labelling
6. Attitudes to healthy eating
7. Dietary needs and shopping habits
8. Demographics

The tests are to be included as near to the start of the questionnaire as possible to avoid any influence on responses from questions (e.g. on attitudes). The self completion introduction gives respondents the chance to practice a couple of questions.
7.3.1 Field piloting

In July a small pilot was carried out using a draft questionnaire to test out timings and initial reactions to the tests and other questions. Members of the research team briefed a team of four interviewers and accompanied them on their assignment to listen in to the interviews, probe respondents over any confusion, and identify any problems. The pilot took place over two days.

We interviewed 25 people from a range of ages, both men and women, and across different ethnic groups in the Ealing/West Ealing/Southall area. Only one person failed the numeracy test and did not do the full test set. All agree to complete the tests using self-completion, even those with little initial confidence. The total length varied from 25 minutes to 45 minutes at pilot. An average of 30 minutes should be achievable at the main stage on this basis.

The tests took between 5 and 20 minutes. All completed their tests (one respondent only just finished in 20 minutes). The majority took between 5 and 12 minutes, but we do not recommend adding any further tests as there were some signs of fatigue for some respondents (e.g. saying ‘are there many more of these’ part way through the tests).

The pilot revealed the need to amend a few questions for clarity, to further test the %GDA numeracy question, and improve multiple label formats.

7.3.2 Cognitive testing

The discussion in Chapter 4 revealed the need to evaluate the proposed test questions to ensure they were true tests of our hypotheses. Cognitive testing is an approach that is regularly used in research to ensure that the questionnaire design addresses research aims.

The aim of this work is to determine exactly how respondents interpret and respond to each of the planned tests and a few other questions, in order to judge what question wording will best provide the required measures of comprehension.

‘Talk aloud’ techniques were used, with researchers asking respondents to read the question aloud, commenting on what they understand it to mean, and then asked to talk through the decision process they use to come up with their final answer. The researchers probed further to get behind any superficial responses. Respondents were also asked to describe the task in their own words, and asked for their understanding of words such as ‘healthiness’, and their interpretation of the response scales.

This stage was conducted in a central location in West London. Respondents were be recruited on street to ensure coverage of a range of demographics (age, sex, lifestage, ethnicity etc). They then took part in a 20-40 minute self-completion
interview using the techniques described above. The researchers noted down their comments as the interview progresses and probe as necessary. Interviews were recorded to allow the researcher to check their notes. We interviewed 100 respondents over two days. There were be four different versions of the questionnaire to ensure the full range of tests and labels were covered, without over-burdening any individual respondent.

The researchers collated their notes and met to discuss their findings, and to suggest possible improvements for each question. This stage also produced information on the range of ways in which different consumers address each task, and insight into attitudes towards the different label types which can be used to help understanding the findings at the main stage.

The results of this stage are included in an annex to this report.

7.4 Fieldwork

Face to face interviewing is to be conducted by BMRB using CAPI (Computer Assisted Personal Interviewing). This method, which has been standard within BMRB for over 10 years, provides advantages of data accuracy, the ability to administer complex questionnaires and fast feedback of topline results. The use of multimedia CAPI (enabling the use of prompt material on screen) and Quantime software makes self-completion CAPI attractive and user friendly.

Fieldwork will be carried out by Kantar Operations, BMRB’s parent company operations resource. Fieldwork in Northern Ireland will be carried out by a sister company, Millward Brown Ulster. All fieldforces are quality accredited.

Interviewers will be briefed via a recorded video briefing. All of the interviewers will be experienced in random probability methods so will be familiar with the approach. The tests will be self explanatory, and run automatically by the CAPI machines, and so do not require personal briefing. The timing is also run by the CAPI machines and does not require any input from the interviewers.

7.5 Data processing

Once collected, all data will be cleaned, and coded. An advantage of CAPI is that the accuracy of the data is greatly increased, and thus the need for editing is reduced. The computer programme will not allow interviewers to enter multiple responses at single coded questions, routing is controlled in the script, and logic checks can be built into the questionnaire. However, the data processing team

---

will be fully briefed to identify any inconsistencies in the data, and editing would be carried out where necessary.

Responses to open questions will be coded. First, responses from a sample of questionnaires will be listed (at least 100 answers for each question). These will be inspected and code frames drawn up to allow the responses to be coded.

Design weights will be applied to account for differential probabilities of selection inherent in the sample design (crucially, individuals in multiple-shopper households will be under-represented), and non-response weights applied if needed to correct for any substantive differential non-response in the final sample, in order to ensure that the final data fully represents the UK population of shoppers.

7.6 Ethical considerations

The design must be informed by ethical considerations. Among the core ethical tenets of the MRS Code of Conduct are those of respondent confidentiality and the principle of informed consent. BMRB always ensure that it is clear to respondents that all of their responses are held in confidence. No individual responses will be passed on to the Food Standards Agency that could be linked to named individuals. In terms of consent, we make it clear to respondents that participation in the survey is voluntary and all respondents have the opportunity to opt out of the research when contacted by an interviewer or researcher. Only those agreeing to further contact are approached for any follow up research.

With a survey of this nature there is a potential worry of respondent burden. We plan to restrict the time any individual is asked to give to the project to overcome this, with a maximum time set for the tests to ensure that the interview is restricted to around 30 minutes on average.

It is also possible that the tests presented to respondents in the survey might cause them stress and anxiety. We will minimise this by not imposing a time limit for any individual test presentation, and restricting the total amount of time they spend on the tests. The tests will be administered using self completion methods to avoid any concern about getting the answers wrong, and the introduction to the tests will also be worded carefully to reduce stress. Furthermore, the early literacy test will ensure that anyone with inadequate reading skills will not be given the tests.

7.7 Summary of data collection and sampling method

We will collect data using a random probability sampled, in-home face-to-face method, using multimedia Computer Assisted Personal Interviewing to allow labels to be shown on screen and the tests to be self-completed. We anticipate a
response rate of around 65%. We will interview 3000 UK main shoppers, providing a robust basis for statistically testing respondents’ comprehension of the different labelling approaches.

The method and questionnaire will be tested in a small scale field pilot of 25 respondents, and a cognitive interviewing stage (around 100 respondents). The interview will last around 30 minutes on average, including a maximum of 20 minutes on the tests.

A programme of statistical modelling will then allow us to understand how the different label elements influence comprehension for different groups of people.
8 Appendix

8.1 Appendix A. Evaluation of tests used in previous studies

Table A1. Identified tests of objective understanding that replay part of the label information.

<table>
<thead>
<tr>
<th>Test Description</th>
<th>Source</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 How much protein/vitamin A/total carbohydrate/saturated fat is in one serving of this food?</td>
<td>National Institute of Nutrition (1999); Byrd-Bredbenner (2000); Alfieri, L. &amp; Byrd-Bredbenner (2001)</td>
<td>+ Captures all nutrients featured on the label in a single score.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Signposting not likely to have an impact because levels of nutrients are only being asked to be replayed.</td>
</tr>
<tr>
<td>1.2 Following prior research (Dickson &amp; Sawyer 1990) the authors derived the respondents attribute (i.e. nutrient) specific relative recall error (RRER) for sodium, potassium, protein, calories, cholesterol, fat and fibre as the absolute value of [(actual attribute value (i.e. nutrient level) - recalled attribute value) x 100/(actual attribute value)]. Respondent specific recall error-index (REI) was also calculated, this captures the overall difference between recall errors for key negative and positive attributes as follows: [(RRER calories + RRERfat) - (RRERprotein + RRERfiber)/(RRER calories + RRERfat + RRERprotein + RRERfiber)]. Scores range between -1 and +1, where a value of -1 (+1) indicates that positive (negative) attributes mainly contribute to the consumer’s total recall error.</td>
<td>Balasubramanian &amp; Cole (2002)</td>
<td>+ Captures all nutrients featured on the label in a single score.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Signposting not likely to have an impact because levels of nutrients are only being asked to be replayed.</td>
</tr>
<tr>
<td>1.3 How many grams of saturates/saturated fat [sugar] are there in one serving of this product?</td>
<td>Institute of Grocery Distribution (2005)</td>
<td>- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Signposting not likely to have an impact because levels of nutrients are only being asked to be replayed.</td>
</tr>
<tr>
<td></td>
<td>Scale used: 0.3g; 0.8g; 1.1g; 1.5g; 2.0g; 2.8g; 3.0g; 4.6g; 5.3g; 7.0g; 8.6g; 9.5g; 10.0g; 16.1g; 30.1g; 33.3g; other amount</td>
<td></td>
</tr>
</tbody>
</table>
1.4 I’d like you to imagine you were going to eat one serving of this product.....Given the information you have in front of you, I would like you to think how much of your GDA for saturates/saturated fat [sugar] you think you would get from one serving of the product.

*Scale used:* line with labelled endpoints, ‘none of your GDA’ to ‘all of your GDA’

<table>
<thead>
<tr>
<th>Institute of Grocery Distribution (2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.</td>
</tr>
<tr>
<td>- Text and TL signposting do not provide the information to answer correctly, i.e. favours GDA signposting, or rather is asking for GDA information to be replayed rather than the nutrient level.</td>
</tr>
<tr>
<td>- Signposting not likely to have an impact because levels of nutrients are only being asked to be replayed.</td>
</tr>
</tbody>
</table>

---

### Table A2. Identified tests of objective understanding that involve carrying out computations using numbers appearing on the label.

<table>
<thead>
<tr>
<th>Test</th>
<th>Source</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>First assessed was whether the respondent understood that these nutrient values were provided on a per serving basis. Then the respondent was asked for the quantity of each nutrient to be found in the entire box.</td>
<td>Jacoby, Chestney &amp; Silberman (1977)</td>
</tr>
<tr>
<td></td>
<td>- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Signposting not likely to have an impact because respondent only required to carry out a computation.</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>How many servings of this food would you need to get all the carbohydrates you need for a day? Scoring: Answers were classified as correct if within plus or minus 2 servings of the right answer.</td>
<td>Levy, Fein &amp; Schucker (1996); Levy &amp;Fein (1998)</td>
</tr>
<tr>
<td></td>
<td>+ Assesses the product in the context of a meal choice or daily intake.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Text and TL signposting do not provide the information to answer correctly, i.e. favours GDA signposting.</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Source</td>
<td>Pros (+) &amp; Cons (-)</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>---------------------</td>
</tr>
</tbody>
</table>
- Captures all nutrients featured on the label in a single score.  
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.  
- Text and TL signposting do not provide the information to answer correctly, i.e. favours GDA signposting. |
- Text and TL signposting do not provide the information to answer correctly, i.e. favours GDA signposting, though one could argue that with regard to the nutrients we are interested in, decreasing levels will always enhance healthiness.  
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire. |
| 2.5  | Byrd-Bredbenner (2000); Alfieri & Byrd-Bredbenner (2001) | + Assesses the product in the context of a meal choice or daily intake.  
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire. |

**2.3** Subjects were asked if they were to consume only three servings of the product in a day, for which of five different nutrients [fat, cholesterol, sodium, saturated fat, and calories from fat content] (in Burton, Biswas & Netemeyer (1994) it was eight nutrients, i.e. the addition of fibre, calcium, and carbohydrates) would they consume more than the recommended amount.

*Scoring:* On the basis of government guidelines for consumption amounts, a response was considered incorrect if the nutrient was (1) not included but should have been or (2) when it should not have been.

For the low nutrition value condition, the product stimulus contained more than the recommended amount for four of the five or eight nutrients; for the high nutrition value condition, the stimulus did not contain more than the recommended amount for any of the five nutrients.

Accuracy scores were computed by summing the number of correct responses across the five or eight negative nutrients and then dividing by five and multiplying by 100. These measures thus represent percent correct scores for which 100 represents a perfect score.

**2.4** If you ate 3 servings of this food in the day, what nutrients would you try to get more of in the other foods that you eat that day, and what nutrients should you try and cut back on in the other foods that you eat in that day?

*Scoring:* Only macronutrient mentions were analyzed. Low and high levels were defined as 5% and 20% of the DRVs. Each product had between three and five macronutrients that were either low or high. Mentions of nutrients that had medium levels in the product were not scored as correct or incorrect and did not enter into the analysis.

**2.5** Five items that assessed the participants' ability to perform simple computations related to diet planning, e.g.:

- If you ate two servings of this food, how much fibre would you get?

*Scoring:* A score for this scale was calculated by summing the number of correct responses. Scores could range from 0 to 5, with higher scores indicating a greater ability to manipulate label information.
<table>
<thead>
<tr>
<th>Test</th>
<th>Source</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
</table>
| 2.6  | Five items that assessed the participants’ ability to perform simple computations related to diet planning, e.g.:
  
  How many servings of this food would you need to eat to get all the calcium you need in a day?
  
  Scoring: A score for this scale was calculated by summing the number of correct responses. Scores could range from 0 to 5, with higher scores indicating a greater ability to manipulate label information.
  
  Byrd-Bredbenner (2000); Alfieri & Byrd-Bredbenner (2001)
 | + Assesses the product in the context of a meal choice or daily intake.  
  - Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.  
  - Text and TL signposting do not provide the information to answer correctly, i.e. favours GDA signposting. |
| 2.7  | According to the information you have been shown what is the Guideline Daily Amount (GDA) for saturates/saturated fat [sugar] in grams for you?
  
  Scale used for saturates/saturated fat [sugar]: 2.5g; 6g; 20g; 24g; 30g; 70g; 75g; 90g; 95g; 120g; 230g; 300g; other amount
  
  Institute of Grocery Distribution (2005)
 | - Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.  
  - Text and TL signposting do not provide the information to answer correctly, i.e. favours GDA signposting, or rather is asking for GDA information to be replayed rather than the nutrient level. |
### Table A3. Identified tests of objective understanding that evaluate a given product based on a given nutrient.

<table>
<thead>
<tr>
<th>Test</th>
<th>Source</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3.1 Healthiness with regard to fibre/calories/sodium</strong>&lt;br&gt;Scale used: 1=very low; 9=very high</td>
<td>Barone, Rose, Manning &amp; Miniard (1996)</td>
<td>- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.&lt;br&gt; - Favours text signposting because of use of words high and low, in the case of text signposts the test can be considered replay.&lt;br&gt; - Need a predefined objective criterion, including anchoring of scale points.</td>
</tr>
<tr>
<td><strong>3.2 Communication of relative amount of nutrient in one serving e.g. ‘A lot of fat’; ‘A lot of calcium’</strong></td>
<td>National Institute of Nutrition (1999)</td>
<td>+ Does not favour one form of signposting.&lt;br&gt; - Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.&lt;br&gt; - Need a predefined objective criterion, including anchoring of scale points.</td>
</tr>
<tr>
<td><strong>3.3 On this label you will see that it indicates that the food product contains 10g of sugar per 100g, do you think this is a lot or a little sugar?</strong>&lt;br&gt;This label on a food product says it contains 20g of fat per 100g; do you think this is a lot or a little fat?&lt;br&gt;\textit{Scale used: A lot; A little}</td>
<td>Food Standards Agency (2001, 2002, 2003, 2004, 2005a, 2006)</td>
<td>+ Does not favour one form of signposting.&lt;br&gt; - Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.&lt;br&gt; - Dichotomous scale and thus prone to lack of response variability, i.e. not all products that contain ‘a little’ contain ‘a lot’…&lt;br&gt; - Need a predefined objective criterion, including anchoring of scale points.</td>
</tr>
<tr>
<td>Test</td>
<td>Source</td>
<td>Pros (+) &amp; Cons (-)</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>---------------------</td>
</tr>
</tbody>
</table>
| 3.4  | Li, Miniard & Barone (2002) | + Does not favour one form of signposting.  
- Captures all nutrients featured on the label in a single score.  
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.  
- Need a predefined objective criterion, though one could argue that with regard to the nutrients we are interested in, decreasing levels will always enhance healthiness. |
| 3.5  | Li, Miniard & Barone (2002) | + Does not favour one form of signposting.  
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.  
- Need a predefined objective criterion, including anchoring of scale points. |
| 3.6  | Li, Miniard & Barone (2002) | + Does not favour one form of signposting.  
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.  
- Need a predefined objective criterion, including anchoring of scale points. |
| 3.7  | Food Standards Agency (2005b) | - Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.  
- Favours text signposting because of use of words high and low, in the case of text signposts the test can be considered replay.  
- Need a predefined objective criterion, including anchoring of scale points. |
### Table A4 Identified tests of objective understanding that evaluate the overall healthiness of a product.

<table>
<thead>
<tr>
<th>Test</th>
<th>Source</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8</td>
<td>Institute of Grocery Distribution (2005) + Assesses the product in the context of a meal choice or daily intake. - Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire. - Text and TL signposting do not provide the information to answer correctly, i.e. favours GDA signposting. - Need a predefined objective criterion, including anchoring of scale points.</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Black &amp; Rayner (1992) + Assesses the product in the context of a meal choice or daily intake. + Does not favour one form of signposting. + No need to ask about separate nutrients, i.e. fewer questions need to be asked. - Need a predefined objective criterion, including how nutrients would be weighted.</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Burton &amp; Biswas (1993) + Does not favour one form of signposting. + No need to ask about separate nutrients, i.e. fewer questions need to be asked. - Need a predefined objective criterion, including how nutrients would be weighted.</td>
<td></td>
</tr>
</tbody>
</table>

**Test 3.8**

And thinking about your Guideline Daily Amount (GDA) for saturates/saturated fat [sugar], if you were eating one serving of this product as a main meal, do you think this amount of saturates/saturated fat [sugar] is...?

Thinking about your Guideline Daily Amount (GDA) for saturates/saturated fat, if you were eating one serving of this product as a snack eaten between meals, do you think this amount of saturates/saturated fat [sugar] is...?

**Scale used:** a lot; quite a lot; a reasonable amount; not very much; hardly any

**Table A4**

<table>
<thead>
<tr>
<th>Test</th>
<th>Source</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Black &amp; Rayner (1992) + Assesses the product in the context of a meal choice or daily intake. + Does not favour one form of signposting. + No need to ask about separate nutrients, i.e. fewer questions need to be asked. - Need a predefined objective criterion, including how nutrients would be weighted.</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Burton &amp; Biswas (1993) + Does not favour one form of signposting. + No need to ask about separate nutrients, i.e. fewer questions need to be asked. - Need a predefined objective criterion, including how nutrients would be weighted.</td>
<td></td>
</tr>
</tbody>
</table>
4.3 Overall healthiness
*Scale used*: 1=very low to 9=very high

Barone, Rose, Manning & Miniard (1996); Li, Miniard & Barone (2002)

+ Does not favour one form of signposting.
+ No need to ask about separate nutrients, i.e. fewer questions need to be asked.
- Need a predefined objective criterion, including how nutrients would be weighted.

4.4 Participants rate the healthfulness of a food
*Scale used*: 1=not healthful at all to 10=very healthful


+ Does not favour one form of signposting.
+ No need to ask about separate nutrients, i.e. fewer questions need to be asked.
- Need a predefined objective criterion, including how nutrients would be weighted.

4.5 Based on the picture and description you’ve just read, how would you rate the following product in terms of its overall healthiness?
*Scale used*: 1=extremely healthy to 7=not at all healthy


+ Does not favour one form of signposting.
+ No need to ask about separate nutrients, i.e. fewer questions need to be asked.
- Need a predefined objective criterion, including how nutrients would be weighted.

---

**Table A5. Identified tests of objective understanding that compare multiple products on given nutrient(s).**

<table>
<thead>
<tr>
<th>Test</th>
<th>Source</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
</table>
| 5.1  | Participants were shown nutrition labels in the same format for a pair of products from the same category and was asked to identify all nutrient differences between them.  
  ACCURACY, the number of nutrient differences between product pairs that were correctly identified  
  TIME, the seconds needed to perform the comparison task, was assumed to measure the information processing effort required by the format (Russo, 1987).  | Levy, Fein & Schucker (1996); Levy & Fein (1998)  
+ Does not favour one form of signposting.  
+ Captures all nutrients featured on the label in a single score.  
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire. |
| 5.2  | Using the information on the packs, which, if any, of these products would you say has a higher .......... (insert Nutrient) content or would you say they are both the same?  
  *Scale used*: x has higher content; y has higher content; Both the same; Can’t tell; Don’t know  
  *Scale used*: Immediately (within 2 seconds); Within 5 seconds; Within 10 seconds; Within 20 seconds; Longer than 20 seconds)  | Food Standards Agency (2005b)  
+ Does not favour one form of signposting.  
+ A correct answer can be defined.  
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire. |
5.3 The first 12 items of the survey were open-ended and asked subjects to interpret the food labels, such as determining carbohydrate or caloric content of an amount of food consumed. 

**Scoring:** Incorrect responses were coded by two reviewers into three possible categories: (1) did not apply serving size/servings per container appropriately, (2) confused by extraneous or complex information, and (3) calculation or other errors.


+ Does not favour one form of signposting.
+ A correct answer can be defined.
+ Captures all nutrients featured on the label in a single score.
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.

5.4 When comparing a healthier and less healthy version of products - ABILITY TO CORRECTLY IDENTIFY THE LEVELS OF KEY NUTRIENTS (%) - All 4 nutrient levels; 3 nutrient levels; 2 nutrient levels; 1 nutrient level; 0 nutrient levels

**SPEED OF USE (%) who said at a glance/quickly but not at a glance**


+ Does not favour one form of signposting.
+ A correct answer can be defined.
+ Captures all nutrients featured on the label in a single score.
- Need to ask about single nutrients, if all nutrients need to be covered this increases the number of questions to be included in the questionnaire.

---

**Table A6. Identified tests of objective understanding that compare multiple products in terms of overall healthiness.**

<table>
<thead>
<tr>
<th>Test</th>
<th>Source</th>
<th>Pros (+) &amp; Cons (-)</th>
</tr>
</thead>
</table>
| 6.1  | Which cereal would be a better choice, assuming you follow the dietician's instructions?  
*Note:* a definition comprising of multiple nutrients of varying stated importance is provided | Cole & Gaeth (1990)  
+ Does not favour one form of signposting.  
+ No need to ask about separate nutrients, i.e. fewer questions need to be asked.  
- Need a predefined objective criterion, i.e. how nutrients would be weighted. |
| 6.2  | Which of these foods do you think would be a wiser choice for a healthy diet?  
*Scoring:* Participants initial response and time took to make it was recorded and any reasons given. | Black & Rayner (1992)  
+ Does not favour one form of signposting.  
+ No need to ask about separate nutrients, i.e. fewer questions need to be asked.  
- Need a predefined objective criterion, i.e. how nutrients would be weighted. |
8.2 Appendix B: Final product and nutrient allocation for tests

The products are labelled using the following key.

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product grouping P1</td>
<td>Ready Meal</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Soup</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Sandwich</td>
<td>C</td>
</tr>
<tr>
<td>Product grouping P2</td>
<td>Fruit yoghurt</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Breakfast cereal</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>Crisps</td>
<td>F</td>
</tr>
</tbody>
</table>

Each address will have been allocated in advance to one of groups 1 and 4. Within these 4 groups they will also be further allocated to one of 10 groups (8 for test 1) at random. This gives a total of 38 groups. The product are labelled A to F as per the key above. Label A1 will be the same label presented across all tests where A1 appears in each grid. Ditto label B1 etc.

Total – 256 test presentations to set up for test 1

<table>
<thead>
<tr>
<th>Respondent group</th>
<th>Test 1</th>
<th>P1</th>
<th>Test 2</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL Text</td>
<td>GDA</td>
<td>A1</td>
<td>B3</td>
<td>A3</td>
</tr>
<tr>
<td>no GDA L1</td>
<td>B2</td>
<td>A1</td>
<td>B3</td>
<td>A3</td>
</tr>
<tr>
<td>no text</td>
<td>GDA</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
</tr>
<tr>
<td>no GDA L4</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
</tr>
<tr>
<td>Text</td>
<td>GDA</td>
<td>L5</td>
<td>B2</td>
<td>A2</td>
</tr>
<tr>
<td>no GDA L6</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
</tr>
<tr>
<td>no text</td>
<td>GDA</td>
<td>L7</td>
<td>A3</td>
<td>C2</td>
</tr>
<tr>
<td>no GDA L8</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
</tr>
</tbody>
</table>
Total – 200 test presentations to set up for test 2 with energy

<table>
<thead>
<tr>
<th>Respondent group</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 with energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDA L1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
</tr>
<tr>
<td>no GDA L2</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
</tr>
<tr>
<td>no GDA L3</td>
<td>C1</td>
<td>B1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>F1</td>
</tr>
<tr>
<td>no text</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
</tr>
<tr>
<td>GDA L5</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
</tr>
<tr>
<td>no GDA L6</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
</tr>
<tr>
<td>no GDA L7</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
</tr>
<tr>
<td>no GDA L8</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
</tr>
<tr>
<td>(Tesco) No TL</td>
<td>no text</td>
<td>GDA L9</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
</tr>
<tr>
<td>(Sainsbury) TL</td>
<td>no text</td>
<td>no GDA L10</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
</tr>
</tbody>
</table>

Total – 200 test presentations to set up for test 2 without energy

<table>
<thead>
<tr>
<th>Respondent group</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 without energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDA L1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
</tr>
<tr>
<td>no GDA L2</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
</tr>
<tr>
<td>no GDA L3</td>
<td>C1</td>
<td>B1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>F1</td>
</tr>
<tr>
<td>no text</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
</tr>
<tr>
<td>GDA L5</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
</tr>
<tr>
<td>no GDA L6</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
</tr>
<tr>
<td>no GDA L7</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
</tr>
<tr>
<td>no GDA L8</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
<td>A1</td>
<td>A4</td>
<td>C3</td>
</tr>
<tr>
<td>(Tesco) No TL</td>
<td>no text</td>
<td>GDA L9</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
<td>B1</td>
</tr>
<tr>
<td>(Sainsbury) TL</td>
<td>no text</td>
<td>no GDA L10</td>
<td>A4</td>
<td>C3</td>
<td>B3</td>
<td>A3</td>
<td>C2</td>
<td>B2</td>
<td>A2</td>
<td>C1</td>
</tr>
</tbody>
</table>
Total – 200 test presentations to set up for test 3

| Respondent group | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
|------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| P1               |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| P2               |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

| Text | GDA L1 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | no GDA L2 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | no text L3 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | no GDA L4 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | Text GDA L5 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | no GDA L6 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | no text GDA L7 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | no GDA L8 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | (Tesco) No TL GDA L9 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|      | (Sainsbury) TL no text no GDA L10 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

Note – each cell at test 3 represents a pair of labels within the same product. It shows the label of the first of the pair only. The second of the pair will be a product of the same type (e.g. A with A). Where possible these second labels will be selected from within the same products (e.g. A1 to A4) but, if necessary, further products will be selected.
8.3 Appendix C - criteria for signposting

The criteria used to signpost (TL) calories (kcal) are given below.

<table>
<thead>
<tr>
<th></th>
<th>Amount of calories (kcal) per portion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Green</td>
</tr>
<tr>
<td>Ready meals</td>
<td>0 to 450</td>
</tr>
<tr>
<td>Sandwiches</td>
<td>0 to 450</td>
</tr>
<tr>
<td>Breakfast cereals</td>
<td>0 to 450</td>
</tr>
<tr>
<td>Soups</td>
<td>0 to 200</td>
</tr>
<tr>
<td>Yogurts</td>
<td>0 to 200</td>
</tr>
<tr>
<td>Crisps</td>
<td>0 to 100</td>
</tr>
</tbody>
</table>

The FSA criteria for signposting (TL) nutrients are given below:

<table>
<thead>
<tr>
<th></th>
<th>Amount per 100g</th>
<th>Per portion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Green</td>
<td>Amber</td>
</tr>
<tr>
<td>Fat</td>
<td>≤ 3.0g</td>
<td>&gt; 3.0g &amp; ≤ 20.0g</td>
</tr>
<tr>
<td>Saturated fat</td>
<td>≤ 1.5g</td>
<td>&gt; 1.5g &amp; ≤ 5.0g</td>
</tr>
<tr>
<td>Sugar</td>
<td>≤ 5.0g</td>
<td>&gt; 5.0g &amp; ≤ 12.5g</td>
</tr>
<tr>
<td>Salt</td>
<td>≤ 0.3g</td>
<td>&gt; 0.3g &amp; ≤ 1.5g</td>
</tr>
</tbody>
</table>

If amount per portion is exceeded, label will always be red regardless of amount per 100g.
9 References


10 Annex 1 - Initial results of phase 1 qualitative work

The annexed report contains a selective analysis of the phase 1 qualitative data to inform the development of the quantitative stage. It does not include full analysis of the data and this will be included in the full report of the results of both qualitative and quantitative work.
11  Annex 2 - Results of cognitive testing stage

The second annexed report contains the results of the cognitive testing phase described in Chapter 7.