Exploring public discourses about emerging biotechnologies through automated coding of open-ended survey questions

Paul Stoneman, National Centre for Research Methods, University of Southampton
Nick Allum, Department of Sociology, University of Essex
Patrick Sturgis, National Centre for Research Methods, University of Southampton

WORKING PAPER PREPARED FOR THE WELLCOME TRUST – DRAFT APRIL 2011

ABSTRACT

The primary method by which social scientists describe public opinion about science and technology is to present frequencies in quantitative variables and to use multivariate statistical models to predict where different groups stand on issues of scientific controversy. Such an approach requires measures of individual preference which can be aligned numerically in an ordinal or, preferably, a continuous manner along an underlying evaluative dimension – generally the standard 5 or 7-point survey attitude item. The key concern motivating the present paper is whether, due to the low salience and 'difficult' nature of much biomedical science for members of the general public, it is sensible to require respondents to choose from amongst a small and predefined set of evaluative response categories. Here, we adopt a different methodological approach; the analysis of textual responses to 'open-ended' questions, in which respondents are asked to state, in their own words, what they understand by terms such as 'stem cell' and 'DNA'. To this textual data we apply the statistical procedures encoded in the Alceste software package to detect and classify underlying discourse and narrative structures. We then examine the extent to which the classifications, thus derived, can aid our understanding of how the public develop and use 'everyday' images of, and talk about, biomedicine to structure their evaluations of moral acceptability, risk, and harm of emerging technologies.

1. INTRODUCTION

It is, perhaps, something of a truism to say that quantitative researchers favour working with large-n datasets and testing statistical relationship between variables. Similarly, it seems almost tautological to define qualitative researchers as those who prefer the rich textual data that arises out of interviews with and observations of a small number of cases. Each tradition is, nonetheless, clearly grounded in preferences for different *kinds* of inference. The quantitative researcher strives for robust generalisation to the population as a whole and, in doing so, must make compromises over the amount (and nature) of the data that can be collected from each case. The qualitative researcher, by the same token, eschews the formalisations of population inference for the greater insight that can be gained from considering a smaller number of 'illustrative' cases, often in very fine detail. What, though, of the holy grail of combining the strengths of both approaches by 'mixing' them within a single research design? Although interest in the idea of mixed methodological research has certainly increased substantially over the past decade and more (Bryman 2007), it is also undoubtedly true that the goal of qualitative and quantitative *integration* is 'more honoured in the breach than in the observance' (Green et al 1989; Niglas 2004; Bryman 2006).

This characterisation is particularly apposite in the study of what has become known as the Public Understanding of Science (PUS), where the convention has been to favour either a quantitative or a qualitative approach, with an often unhealthy conflation of methodological orientation and theoretical perspective (Irwin and Wynne 1996; Sturgis and Allum 2004). The motivation underlying this paper, however, is that there is, potentially, much to be gained for our understanding of public responses to controversial and emerging areas of science and technology by applying the rigour and inferential power of quantitative analysis to rich and unstructured textual data collected from a large random sample of the general public. This is because, when a technology is unfamiliar and cognitively demanding to understand, it seems unlikely that survey designers will be capable of pre-determining the full range of responses that might be given by members of the public about it, when asked. Under such conditions, it is certainly possible that the standard closed-format survey question does not so much reveal pre-existing public opinion about the technology in question, as create it, a critique that has long dogged the social survey more generally (Moscovici 1963; Converse 1969).

In recognising that closed-ended questions probably constrain, to some degree at least, our understanding of public responses to new issues and technologies in biomedicine, we explore in this paper the extent to which quantitative analysis of verbatim responses to open-ended questions can provide a methodological solution to the problems associated with asking the public about new and emerging technologies in surveys. To be sure, the approach we apply here - so-called '*quantitizing*' of qualitative data (Sandelowski et al 2009) - is just one way of combining qualitative and quantitative methods, and one that might, indeed, stand accused of merely transforming the qualitative into the quantitative as opposed to being a genuinely *integrative* approach. Be that as it may, our goal here is not to resolve such definitional issues of method but to explore, pragmatically, whether this novel approach might be a useful tool for the analysis of public understandings of and reactions to new and emerging biomedical science. The remainder of the paper is set out as follows. First, we describe the social and political context in which the collection and analysis of public opinion about science and technology is situated, before reviewing some of the methodological challenges that arise when asking questions about low-salience and cognitively demanding societal issues. We then describe the data and key measures upon which our analysis is based and present our key results. We conclude with a discussion of the substantive implications of our findings and an evaluation of the methodology employed.

2 Science Policy, Practice and Public Opinion

A dilemma within science governance is how to deal with the democratic demands placed upon regulatory institutions when technologies are being rapidly developed and applied, and the public (as well as policy-makers) struggle to keep pace with their uses, implications, risks, and benefits (Ruckelshaus 1986; Kleinmann 2000; Salter and Jones 2006). Historically, in the UK, it has been left up to the discretion of Members of Parliament to call for a review of a policy which allows the use of demonstrably or potentially 'risky' technology or scientific practice. Science governance has, therefore, traditionally been a policy domain driven almost entirely by ‘elites’ and special advisors, with little opportunity for the preferences of the general public, such as they are, to feed into funding and regulatory decision-making. In response to this ‘democratic deficit’, whether real or imagined, a range of public engagement and consultational procedures and institutional frameworks have been established over the past ten to fifteen years, with the goal of enabling ordinary citizens to have more influence on the direction science policy and practice (Frewer and Salter 2007; Salter and Jones 2002).

Although not generally conceived of as being part of the public engagement apparatus, a crucial link between science, government and the public are social surveys because these are they key conduit through which public opinion and preferences regarding the speed and direction of science policy are communicated to institutional stakeholders. It is also, of course, through the findings of opinion polls and attitude surveys that the media often intervene to report on public fears or hostility to an area of scientific practice, often in the form of headline grabbing majority rejecting a particular technology or practice. This has been especially true in the UK since 1996 - a time identified as the start of agricultural biotechnology’s ‘watershed years’ (Gaskell & Bauer 2001) - when two issues in particular dominated public consciousness about science: Bovine Spongiform Encephalopathy (BSE) and Genetically Modified (GM) food. On 20 March 1996 the Daily Mirror ran the headline ‘Mad Cow Can Kill You’ on its front page. It outlined the then Health Secretary Stephen Dorrell’s plan to admit in the House of Commons that ten cases of a new strain of Creutzfeldt-Jakob disease (CJD) could be linked to the spread of (BSE) in cattle towards the end of the 1980s. Over the following weeks, the media increased the intensity of its reporting of the issue with dramatic front-page headlines such as ‘Could It Be Worse Than Aids?’ (Daily Mail, 22 March 1996) not uncommon.

The publication in October 2000 of the Phillips Report on BSE (colloquially referred to as ‘mad-cow disease’) marked a low point in UK science–public dialogue. One major problem highlighted in the report was the disjuncture between government communications and public opinion. Whilst the report concluded that the government did not actually lie to the public about the risks of BSE, it criticised officials for being preoccupied with preventing an alarmist public over-reaction by issuing a strong campaign of reassurance. This, however, ultimately had the highly unintended consequence of further undermining public confidence in government regulation of science when the government was forced, not long afterwards, to reveal that there was indeed a demonstrable link between CJD and human consumption of beef. This led the Phillips Report to conclude that “*confidence in public pronouncements about risk was a further casualty of BSE’* (Phillips Report, 2000, p.xviii).

The other great British science risk issue of the 1990s, GM foods, was another case of a government apparently mis-reading public opinion and exacerbating public hostility towards this key area of science as a result. The then Prime Minister, Tony Blair, was widely criticised in February 1999 for his attempts at reassuring citizens about the safety of GM food. One tabloid newspaper reported that Blair was ‘frustrated’ that the ‘potential benefits of GM food are being ignored in the escalating row’. The depressing implication was that very little had been learnt from the BSE debacle, in terms of the need for more than blanket reassurances when addressing public concerns. Once again, a complex area of scientific, social and ethical debate was being dealt with in an apparently arrogant and high-handed manner. At the same time, and as in the BSE episode, the public’s legitimate questions over risk and technological development were dismissed as irrational and uninformed.

In order to diffuse the tensions between government, the media and the public, a ‘GM Nation?’ debate was put in place, which was intended to bring together all of the relevant stakeholders to air their concerns in various forums. However, this failed to diffuse the problem:

> “*In 2004, an unusual convergence of views appeared as environmental and consumer groups and sections of the right wing press expressed outrage, following a government announcement that it planned to approve a GM maize developed by Bayer…Representatives of consumer and environmental groups were…scathing in their criticism of the government.*” (Gaskell 2004: 242)

This was, in part, fuelled by intense media coverage. ‘So we’re going to be forced fed GM’ claimed the Daily Mail, which continued by claiming that the government ‘treats with contempt the results of a huge survey organized by the government last year to gauge the public mood. It found that 90% of Britons oppose the growing of so-called ‘frankencrops’ and the sale of food derived from them’. Within the space of a few years, the UK had jumped from the BSE debacle to the GM debacle. Not surprisingly,

> “…*the House of Lords identified a ‘crisis of confidence’ and called for ‘more and better dialogue’. But as the public and media debate over genetically modified (GM) crops intensified, many scientists felt they were on the losing side of a battle for hearts and minds. There was a consolation prize: the GM saga made the scientific establishment sit up and think about the importance of dialogue on difficult issues. As a result, they have adopted new, and better, models of science communication. There is a growing confidence that lessons have been learned*.” (Wilsdon et al. 2005: 16)

These lessons learned have, it would seem, been of benefit in garnering support for, or at least in minimizing rejection of, stem cell research in the UK. On 16 March 2005, the UK Government announced the establishment of the UK stem cell Initiative. It promised that government would work with public and private sector stakeholders to formulate a ten-year strategy for UK stem cell research, which included a long term funding plan. The huge public investment in this area, along with better communication channels from central government, have placed UK stem cell research at the forefront of this field of biomedicine (Pattison 2005). Despite this, although broadly positive, within the UK itself, opinions are still divided on the exact nature, role and risks of stem cell cloning (BMRB Report: 45106748 Stem cells public dialogue v). It has been the focus of much media attention and political controversy, with the media and politicians alike quoting statistics from social surveys to support their own political position (notably during the 2010 stem cell debates in the USA).

Such dissemination of public opinion through the national media is probably the most effective way for the public to have an influence on science governance. When a negative consensus emerges around a technology or practice, policy-makers are, history shows, obliged to take the findings of public opinion surveys seriously. Indeed, it is not uncommon for survey findings to provide the political impetus for a wholesale policy review. The way in which public opinion about science is measured in surveys is, therefore, of critical importance.

2.2 Closed-ended Questions and Non-attitudes

In the examples described above, survey measures of public opinion exerted substantial influence on science policy via data generated through standard ‘closed-ended’ questions. That is to say, respondents to these surveys were asked to express their position with regard to say, GM crops, by selecting one of a limited number of pre-specified evaluative descriptors such as ‘strongly agree’ that ‘On balance, the advantages of genetically modified foods outweigh any dangers’[1]. The limitations of survey questions which present a fixed set of response alternatives for uncovering the complexities and dynamism of public opinion are well known (Blumer 1969; cites). Most importantly for our purposes here, it is quite possible that by constraining the range of opinions available to be expressed to a predefined set of answers that are themselves selected by the researcher (or funder), closed-format questions result in a view of public opinion which is, essentially, ‘rigged’ in advance to reflect the (often implicit) assumptions of those who commissioned and designed the survey.

In addition to the potential for closed-ended questions to shape or steer responses in a particular direction, a further consideration is that, when people have little or no understanding of the science they are being asked to evaluate, many will select one of the fixed alternatives offered to them rather than admit ignorance, simply because the formalities and conventions of the survey interview stipulate that providing answers is ‘what you are supposed to do’ (Converse 1964; Bishop et al 1980). These types of ‘non-attitudes’, or ‘pseudo-opinions’ can represent a large proportion of all responses on questions about issues of scientific complexity and are, unsurprisingly, more prevalent amongst the less scientifically knowledgeable members of the public (Sturgis and Smith 2010). For these reasons, it seems sensible to ask whether close-ended questions are really the most appropriate way of measuring public opinion about complex areas of science and technology, particularly if the results are intended to feed into policy and regulatory decision-making.

2.3 The Potential Value of Open-ended Questions

The obvious alternative to presenting a set of fixed response alternatives is to ask respondents to report their thoughts and perspectives on a particular issue in their own words and for interviewers to record these responses ‘verbatim’. The potential advantage of this type of ‘open-ended’ question is that it allows the respondent to use their own frame of reference in determining a response, even if this might seem inappropriate or ‘irrational’ to the survey designer or analyst. Thus, this approach should result in a fuller and more heterogeneous set of perspectives than the standard closed-ended question. Additionally, of course, the amount of information about an individual’s position on an issue that can be derived from an open question is considerably greater than that which is afforded by a closed-ended alternative. And, indeed, many large scale academic surveys have employed open-ended questions, particularly in the pioneering period of survey research in the United States, during the post-war era. Most notable, perhaps, in this regard is the American

[1] This question was fielded in the 2003 British Social Attitudes Survey. Respondents could either strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree.

National Election Survey series which has fielded open-ended questions on American political issues since the 1940s and which formed the basis of the important idea that the American public can be stratified into different 'levels' of ideological sophistication, based on their responses to these questions (Converse date).

Despite their appeal as a means of avoiding the shortcomings of fixed response alternative questions, the open-ended question is not widely used in survey research today. One reason for this is that concerns have long been raised about whether open-ended questions favour the articulate and well-educated, who are likely to provide longer and potentially richer responses and, thereby, exert a disproportional influence on public policy (see Sturgis and Allum 2006; though see Greer (1988) for a counter-position). On a more practical level, open-ended questions are not used frequently due to the high cost of fielding these questions, because they take longer for interviewers to administer and require additional resources to transcribe and to code into a frame. And, even when open-ended questions are included in a survey, it is rare for analysts to use them in a way that exploits the additional information provided by the full-text strings. Rather they are generally employed by analysts in a quantitative manner, as if the question had been asked as a closed-ended question in the first place, raising further questions about the returns to the additional cost of including them. It is clear that, partly, this tendency is a result of 'habitual practices'; quantitative researchers using the procedures with which they are comfortable and familiar. It is also, however, due to the fact that there are few well-established procedures for using verbatim responses in any other way, beyond their occasional use as quotations to offer a 'flavour' of the kinds of responses that were given, alongside a conventional quantitative analysis. What is needed, then, is a more systematic and robust methodology for the analysis of open-ended survey questions, that is capable of utilising the rich semantic information of the verbatim responses, but in a way that retains the possibility of reliable population inference. It is to this objective that we aim to contribute in the following sections of the paper.

3. DATA AND MEASURES

We base our analyses on data from the 2010 Wellcome Trust Monitor of public knowledge, interest and engagement in biomedical science. The Monitor is a stratified, clustered probability sample, with the Postcode Address File (PAF) used as the sampling frame of households. One adult member, aged 18 or above, of each responding household was randomly selected for interview using the Kish grid procedure (Kish 1962). The survey achieved a response rate of 59%, yielding 1,179 adults as our analytical sample size (see cite for full technical details of the survey). The strength of the Monitor for our purposes here is that, while predominantly employing standard closed-ended questions, it also contains several open-ended questions, in which respondents were asked to say, in their own words, what came to mind when they heard a particular scientific term or phrase. The interviewer transcribed the verbatim reply given by each respondent and it is these free-form text responses on which we base our analysis. We focus on two questions, one about stem cells and one about DNA. In order to avoid asking people to talk openly about something they have never heard of, respondents were first asked to rate their level of awareness of each term[2]. Only respondents who indicated that they thought they had some understanding of DNA or stem cells were then administered the following questions:

1) What do you understand by the term "DNA"?

2) What do you understand by the term "stem cell"?

Interviewers were instructed to record, verbatim, the answers respondents provided to these questions.

4. ANALYTICAL APPROACH

Our analysis uses a set of statistical procedures encoded in the software package, Alceste (Reinert, 1998). In conceptual terms, an Alceste analysis seeks to inductively reveal common narratives or discourses within a body of textual data. In the context of biomedical applications, for example, we can think of there being different ways in which individuals will conceptualise, understand and evaluate stem cell science and that this cognitive and affective variation will be reflected in the ways in which people speak about the science in question. Thus, the procedure can be used as a means of uncovering the latent social-cognitive basis of the verbatim responses in a way that enables the structure to emerge from people's actual talk, rather than being imposed, *a priori*, by the researcher. Alceste does not use extensive dictionaries of semantic categories for its primary analysis but, rather, it relies on the distribution pattern of words within a body of text, often referred to as a 'corpus'. Reinert's theoretical rationale for the Alceste program comes, at least in part, from the idea of 'free association' and dream analysis in the Freudian and Jungian psychoanalytic traditions. In Freudian psychoanalysis, a subject who is asked to speak freely about anything which enters their head may reveal the presence of unconscious thoughts and intentions that are submerged during normal conversation. Indications of meaning can be inferred by the researcher via the examination of the characteristic classes of words, which form the primary output of the program.

---

[2] The question wording was, *"I'd now like to ask you about your understanding of different scientific terms that are used in news stories dealing with medical research. First, when you hear the term DNA, how would you rate your understanding of what the term means: 1. I have a very good understanding of what DNA means 2. I have a good understanding 3. I have some understanding 4. I have heard the term "DNA" but have little understanding of what it means 5. I have not heard the term DNA".* Only respondents selecting response alternatives 1-3 were administered the open-ended follow-up question.

The Alceste program carries out a sequence of textual processing and statistical analysis procedures in order to produce its final output. The first of these is *Recognition of dictionary words in the corpus.* During this stage, a dictionary of common functional words such as prepositions, definite and indefinite articles, pronouns and so on are identified, so that these can be disregarded during the analysis. The automatically generated dictionary can be modified manually by the analyst, if necessary. The next stage is referred to as *Lemmatisation of functionally equivalent nouns and verbs* and is the process whereby verbs and nouns are reduced to their shortest stem. This means that, for instance, 'gene', 'genes', and 'genetics' would all be lemmatised to 'gene+', where the '+' indicates that more than one suffix has been detected for the common term 'gene'. The purpose of this procedure is to render equivalent different versions of functionally equivalent words, so that they are not regarded as separate entities for the purpose of analysis.

The third stage is referred to as *Parsing of the text into 'context units'.* With a large corpus of semi-structured text, such as focus group interview transcripts, Alceste has a procedure for parsing the text into smaller analysable chunks, which are referred to as elementary context units (ECUs). In the present case, because we have comparatively short text strings, we set the ECUs to be coterminous with the complete response for each respondent. Once these data preparation and management procedures have been implemented, a *Hierarchical cluster analysis* of words by ECU is carried out to produce word 'classes': A large contingency table is generated from a matrix containing the ECUs in the rows and the words (presence or absence of all substantive words in the corpus) as the column variables. The program begins with all of the ECUs as one class and then iteratively splits the corpus into two maximally distinct sub-classes according to a Chi-square criterion (????). In other words, the ECUs are first split into two groups and the group indicator is cross-tabulated with the words. The squared difference between the observed and the expected word counts is then evaluated. This process is repeated until the two maximally different classes are found in terms of the distribution of words.

In a final stage, the set of (typically, three to seven) classes is cross-tabulated with the words in the corpus and subjected to a correspondence analysis. This is a geometric technique for visualising the variation in a contingency table in a low-dimensional space (Greenacre, 2007) and can be thought of as analogous to a principal components analysis for categorical variables. The output from this analysis can be used to identify the proximity of words and classes to each other along the key dimensions of variation.

5. RESULTS

Before presenting the results of the Alceste analyses, it is useful to provide some descriptive information about the nature of the responses provided to the open-coded questions asked to think about 'stem cells' and 'DNA'.

The basic information for the responses to these questions is presented in Table 1. The first thing to note in table 1 is the high proportion of individuals who did not provide any verbatim material at all, around 1 in 10 for the DNA question and more than 1 in 3 for the stem cell question. These figures include those respondents who said they had not heard of DNA/stem cells, so were not asked the open-ended question along with the small number of respondents who were asked the question but did not provide any verbatim response (17 for DNA and 2 for stem cells). Clearly, then, large minorities of the public have heard of neither technology, although awareness is substantially greater for DNA than for stem cells, a fact which no doubt reflects the more general nature of DNA as a scientific concept, and its longer heritage within the English language. The greater familiarity with DNA is also reflected in the fact that more words (on average) are used in response to the DNA (11) than to the stem cells (8.3) question.

Table 1: Descriptive Information of Verbatim Responses to DNA and Stem Cell Questions

	DNA	*Stem Cell*
Response	88.7	62.1
Non-response	11.3	37.9
Average word count	11	8.3

The most frequently used terms to describe DNA were: ‘gene’, ‘make up’, ‘person’, ‘individual’, ‘cell’ and ‘unique’. This suggests that the general narrative around DNA is that it is related to genetics and it is part of what makes individuals different from one another. Interestingly, there are also some not so intuitive terms such as ‘fingerprint’ and ‘blood’ which would seem to be related to respondents linking DNA with criminal investigations. These potentially different ways of thinking about DNA can be seen in the examples below which are from the verbatim responses (the respondent's unique serial number is in brackets):

(111171) “A unique human fingerprint”
(112051) “Body’s blueprint; specific to you as an individual”
(133201) “Double helix – the blue print of life”
(141111) “It’s the code that is passed on to your personality”

These four examples demonstrate the idea of DNA being the basic code upon which our individuality is built, making us ‘who we are’. In this sense, DNA is defined in terms of *what it does*. But there are also many ‘non-functional’ accounts of DNA offered. Compare the first and second pair of responses below:

(111231) “Used to determine the true father of a child”

(138011) “What I have seen on police things – they take swabs out of your mouth if you’ve been to a certain place”
(123121) “Deoxyribonucleic acid is the formation/structure of the cell nucleus to determine the character of any living organism”
(127191) “Coil of atoms”

The first pair focus on how our knowledge about DNA is *applied* to everyday life, with genetic parentage and criminal investigations as examples. The second pair show offer *ontological* definitions of DNA, which in strict technical terms is the most accurate type of response. Nonetheless, even when offering this type of response, we can see from the two examples above that the first one is full and completely accurate while the second, whilst trying to define what DNA actually is (as opposed to what it does or how it is applied) relies on a visual representation and not a technical definition.

Similar differences can be observed in the responses to the stem cells question. The most frequently used terms to describe respondents’ understandings of this term are physical references such ‘body’ and ‘umbilical cord’, as well as irregular verbs such as ‘know’, ‘take’, ‘grow’, ‘develop’ and ‘use’. So it seems there isn’t a strong overall narrative but rather an array of descriptive images and a selection of verbs intended to convey the function of stem cells. For example, the two verbatim responses below show a generic response which touches upon the *application* of stem cells in medical research. The respondents know that stem cells are an important part of medical research, but exactly what they are and how they help is not specified.

(139181) “The ability to cure, help with genetic diseases”
(153141) “Use it to advance medical research”

These responses can be contrasted with the following examples:

(155011) “Cells taken from the umbilical cord”
(147061) “Genetic material taken from an embryo”
(146101) “Cells taken from the brain stem”

In these responses, what the stem cells are, what they do and how they are applied is ignored. The main concern is seemingly where they originate. Of course, stem cells can be taken from a variety of sources (embryonic / fetal / umbilical / placenta, for example) so the respondents are clearly thinking of one or two examples they can recall and treating them as representative of how all stem cell research is undertaken. Finally, there are a set of answers that focus on the application of stem cells. For example:

(181051) “It’s where they take bits of you and grow it”
(153151) “Bits inside the brain to help with cloning”

With these types of responses we can see that respondents are thinking about therapeutic cloning. They demonstrate how knowledge about stem cells is applied to medical research, but fall short of saying what these types of cells actually are. Compare with:

(182211) “Non-differentiated cells which have the ability to turn into any type of cell”
(1821131) “A basic cell that can be programmed by the DNA to perform various roles in the body”

These final types of responses go one step further. Whilst still talking about the application of stem cell research to medical practices, they also talk about the ontological status of stem cells, which is of them being “non-differentiated cells” or “basic cells”.

Of course, there is a severe methodological weakness with relying on subjective scanning of the verbatim responses and the selection of ‘illustrative’ quotes. First, it uses information from only a tiny fraction of the total sample and, second, there is no way of telling how ‘representative’ these selections really are. We therefore now move on to the results of the Alceste analysis which will enable us to over come these limitations.

5.1 Alceste Analysis of the Stem Cells Open-ended Question

We begin with the analysis of the stem cells data. Table 3 presents the headline results of the Alceste cluster analysis, which shows that there are five distinct classes, representing narrative structures, within the corpus of statements. The first and largest class contains those respondents who offered no response at all (37.9%). The next class is those individuals who could not be classified, due to offering a very generic or very idiosyncratic verbatim response (15.7%). The remainder are those who were successfully classified, with three classes appearing all of different sizes: class 1, 14.7%; class 2, 8.3% and class 3, 23.4%. The number of words contained within each class is very similar, with class 1 containing 38 terms, class 2 containing 39, and class 3 containing 32.

Table 3: Stem Cell Class membership

	No. of statements	Percent	No. of unique words within class
Non-response	447	37.9	
Unclassified	185	15.7	
Class 1	173	14.7	38
Class 2	98	8.3	39
Class 3	276	23.4	32
Total	1179	100	

Table 4 shows the most distinctive words in each of the three classes, while table Table 5 shows the ten most distinctive statements for each class. Looking first at Class 1, this seems to represent a narrative about how one can obtain stem cells and what they are used for. The principal representations are of the umbilical cord and blood as the source of cells. However, there is some vagueness and apparent confusion too – haemophilia is mentioned, as well as spinal cord. The words 'cure' and 'help' feature strongly, showing that people are aware of the potential of stem cell research for improving health. Umbilical stem cells are only one potential source, but nonetheless, this is a widely held view, which presumably emanates from media representations.

Table 4: Alceste classes: stem cells

Class 1			Class 2			Class 3		
Word	Freq	%	Word	Freq	%	Word	Freq	%
cord	65	89	type	33	88	part	100	86
umbilical	53	91	into	71	55	body	120	78
cure	37	86	become	16	88	grow	79	78
bone	21	95	any	37	57	make up	18	94
disease	27	85	cell	309	27	dna	22	86
help	28	82	start	8	100	using	14	93
blood	25	84	specialist	8	100	tissue	41	73
marrow	14	100	liver	7	100	brain	26	77
babies	18	83	kind	9	89	replace	18	83
illness	10	100	basic	24	58	growth	11	91
leukaemia	10	100	differentiated	6	100	base	10	90

Class 2 comprises statements that predominantly describe, in what appears to be a quite scientifically-informed way, what stem cells actually are, rather than what they could be used for in relation to health care. ‘Cell’ is the most frequent word and respondents in this class appear to understand the nature of stem cells and their potential to develop into different types of cell. One would expect respondents in the class to be more knowledgeable and engaged with biomedical research.

Class 3 is somewhat harder to define than the other two classes. People making statements of this kind tend to refer to body parts, tissues and to the idea that tissue can be reproduced or regenerated. The brain figures quite prominently in this narrative structure, but there is little mention of any of specific stem-cell related processes.

Figure 1 shows the correspondence analysis plot. This displays the variation in a contingency table that consists of classes in the columns and words in the rows. The further from the centre of the chart the words are, the more they can be thought of as being distinctive to a particular class. The closer a word is to a class, the stronger is the association with that class. Words that are close together are more likely to co-occur in the same statement. The plot shows that the classes are all very distinct from each other, with little overlap in the probability of words from one class appearing also in other classes. This suggests that there really are quite distinctive groups in the population with regard to how people understand and relate to the concept of stem cells. A small number of words are less class-specific. ‘Develop’ and ‘cell’ and ‘human’ are somewhere between classes 2 and 3, meaning they are more likely to appear in both of them. Most words are clearly associated with a particular class, and reinforce the results shown in the previous two tables.

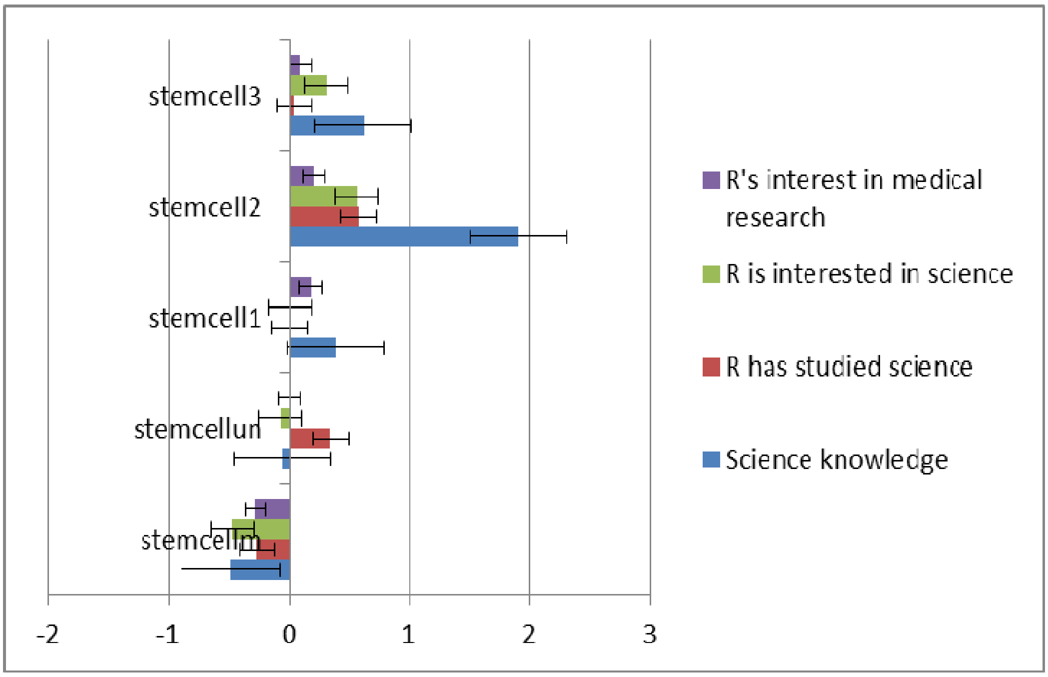
Figure 1: Alceste correspondence analysis: stem cells

```
   +-----|---------|---------|---------+---------|---------|---------|-----+
   |                                   |              find.                |
   |               block               |                  research         |
   |           kind specialis          |                                   |
   |build start   change become        |                                   |
   |   type into particular            |                        treatment  |
   |      turn ability                 |                                   |
   |   basic embryo                    |                                   |
Class 2
   |                                   |                           baby    |
   |     specific                      |                          bone know
   |      life create                  |                    marrow cure born
   |                                   |                 cord umbilical illness
   |   different                       |                disease babies  leukaemia
   |                                   |                 come placenta   child
   |                                   |                        certain help
   | develop                           |                            use people
Class 1
   | cell                              |                            lot health
   |                                   |                                inject
   |                                   |                                   |
   +human----------------------------- -------------------------really out
   |                                   |                                new
   |                                   |                                spinal
   |                                   |                                   |
   | form                              |                                   |
   |                                   |                                   |
   |                                   |                                   |
   |                                   |                                   |
   |                                   |                              gene
   |                                   |                                   |
   |     organ                         |                                   |
   |      foetuses                     |                             take  |
Class 3
   |                                   |                           give    |
   |                                   |                     think thing   |
   |                                   |                        genetic    |
   |                                   |                                   |
   |          tissue various           |                                   |
   |               go base             |                                   |
   |       grow can_t   ear  produce brain                                 |
   |             stem  part repair body damage     stem_cell               |
   |             DNA nothing          using   growth regenerate person     |
   |                         sure     |individual extract                  |
   +-----|---------|---------|---------+---------|---------|---------|-----+


X axis   58% of the inertia
Y axis  41% of the inertia
```

The analysis of the verbatim statements has produced an inductively-generated, yet systematic way of classifying respondents using their own verbatim responses to an open-ended question about stem cells. By importing this classification back into the main survey dataset, we can begin to relate these groups of respondents to other relevant variables. In principle there are many analyses that could be carried out. We choose here to focus on engagement variables: interest in medical research, interest in science, whether a person has studied a science subject at university level and score on a science knowledge quiz. Figure 2 shows deviations from the means of these four indicators for respondents in each of the three substantive Alceste classes, as well as the non-responders and unclassified groups: 95% confidence intervals are superimposed on the bars.

Figure 2: Mean differences in interest and knowledge from sample mean



Respondents in classes 2 and 3 have higher than average interest in and knowledge of science. This is particularly the case for class 2, where respondents are also significantly more likely to have studied science and express an interest in medical research. Class 1 does not seem to be particularly diagnostic of more engaged respondents, although respondents in this class show slightly higher than mean levels of interest in medical research. The class of responders who have no familiarity with stem cells, or who say nothing when asked what they understand by the term are, perhaps unsurprisingly, significantly below the mean on all the engagement variables.

5.2 Alceste Analysis of the DNA Verbatim Responses

We now turn to the analysis of the verbatim responses to the question about DNA. As Table 6 shows, the cluster analysis produced seven classes. The non-response category (11.3%) is comprised almost entirely of those who said they were not familiar with the term DNA and so were not asked the verbatim question. There is a second non-substantive group of respondents, who provided some verbatim response to the open-ended question but whom it was not possible to allocate to one of the narrative classes. This group comprises almost a quarter of all respondents. For those who were allocated to one of the 5 substantive classes, class one is the biggest (23.7%) followed by class 2 (15.9%), class 3 (10.6%), class 4 (8.1%) and finally class 5 (7.8%). There is a good deal of variation in the number of unique words found within each class, particularly compared with the solution for the stem-cell analysis. Class 2 is the most numerous with 69 unique words, while classes 3 and 5 have 44 and 38 terms respectively. Classes 1 and 4 are very homogenous with only 20 and 19 unique words in them respectively.

Table 6: DNA class membership

	Frequency	Percent	No. of words within class
Non-response	133	11.3	
Unclassified	266	22.6	
Class 1	280	23.7	20
Class 2	187	15.9	69
Class 3	125	10.6	44
Class 4	96	8.1	19
Class 5	92	7.8	38
Total	1179	100	

Table 7 shows the most common words defining the classes, while Table 8 shows the ten most common statements for each class. These can be used to aid interpretation of what the narrative structures and discourses underlying the class formations are. Looking at Table 7, Class 1 is defined by a discourse about genetic make-up. People understand that DNA defines individuality in some way, related to genes, that is, they focus on the direct effect of DNA on people. Class 2 on the other hand appears to be focused on the function and uses of DNA testing. Proving paternity, solving crimes by taking samples of blood or other tissues.

Table 7: Alceste classes: DNA

Word	Freq	%	Word	Freq	%	Word	Freq	%	Word	Freq	%	Word	Freq	%
gene	222	71	find out	27	100	acid	50	96	personal	20	85	everyone	60	63
make up	206	71	Parent	36	86	life	46	93	human	40	55	hair	27	85
individual	107	68	Test	35	86	building block	48	90	living	28	57	DNA	123	37
make	60	75	Child	27	93	helix	22	95	identify	32	53	leave	8	100
map	10	100	Prove	20	95	organism	14	100	basic	9	89	saliva	26	58
body	112	52	Sample	33	79	deoxyribonucle	15	93	signature	6	100	finger	9	89
cell	76	53	identify	49	67	protein	13	92	footprint	11	73	own	44	44
pattern	13	77	take	33	76	dioxin	4	100	molecular	7	86	different	53	53
			murder	14	100	nucleus	4	100	link	4	100	skin	8	8
			blood	54	61	instruct	4	100	imprint	6	83	their	31	31
			father	24	79				unique	81	27	fingerprint	66	33

Class 3 consists of what appears in many cases to be a bona fide description of DNA – the double helix and other more technical terms are used in this discourse. Class 4 contains statements that again describe DNA, but much more vaguely. These responses are couched in terms of a blueprint or a signature, but no biological or technical knowledge seems to be present. Class 5 is close in meaning to Class 2. DNA is described in terms of how it can be used to identify people and how it can be thought of as analogous to a 'fingerprint'.

Turning to the correspondence plot in Figure 3, we can see that all of the descriptive discourse is on the right hand side of the chart, while the more function-based classes appear on the left. Thus, we can think of members of the public being positioned on a dimension from description to functionality in relation to DNA. The accurate description in Class 3 is the most distinctive class, in terms of its distance to the middle of both horizontal and vertical axes, and in terms of the distance between it and the other Classes. Respondents in Class 3 appear, then, to be a distinct group.

Table 8: Characteristic statements: DNA

1	1. make_up of a persons cells which makes each person an individual 2. the make_up of the genes in the individual person or plant the unique pattern of the individuals genes 3. your genes and cells your body is make_up of 4. the genes which make_up of the individual or person 5. genes which make_up your own cells in your own body 6. our genes each individual make_ups 7. its your make_up and genes of the individual person 8. the it_s like the map of the what make_up person_s the being of a person_s body 9. gene cells classification for each person 10. what your body is make_up of individual make_up of genes
2	1. it is a way of find_out the book group whether the blood group is passed from parent to child it is used to test the paternity of a child 2. blood samples to find_out who is who i wd say for example in a murder case you can find_out and how you find_out if you are related, parents etc 3. its taking samples of blood or tissue to identify murder victims or fathers of children 4. it means that you can tell when people are related a test find_out how people are and who their biological father is and who a murderer is it is an examination of bodily fluids and tissues 5. blood tests that you take to identify you and to prove that you are the parent of a child 6. taking specimens to find_out if someone is a murderer or to prove who is the father of a child 7. police use it as evidence used to determine paternity take a swab that can be tested 8. i know the letters but not what they mean if it matches or not you can find_out your real mother it solves crime dna is in every one blood hair skin test to seeif a man is father of child 9. to find_out if a child belongs to father same blood groups as parents 10. if anybody does a murder it can help identify them and find_out which parent is correct etc

3	1. its a spiral of types of molecules containing the building_blocks of life 2. double helix of amino acids building_block if life 3. deoxyribose nucleic acid building_blocks of life 4. the double helix structure carrying all the gene information colour of eyes etc 5. deoxyrybo nucleic acid the building_blocks of all life 6. the building_blocks of life its an acid but i cant remember what the dkny stands for 7. it is double helix building_blocks of life, but that_s amino acids 8. deoxyribose nucleic acid spiral helix acgt defines characteristics of living things 9. dioxyribo nucleic acid it is in strand form and you get the double helix and they sort it out it is what makes you who you are 10. deoxyribonucleic acid material coding that determines physical characteristics
4	1. it is the basic molecular structure of living things a unique footprint 2. molecular structure of humans personal barcode 3. its the basic blueprint for al living things s 4. unique personal signature 5. personal identity links 6. long link of parts that make_up the human blueprint 7. signature foe a person unique to all living things 8. its the code blueprints what any living thing is defined brice down into different sections of the chain 9. the complex chemical structure that determines our identity 10. what our basic make_up is our unique chemical make_up
5	1. cant explain everyone_s dna is different 2. its like your id code and everyone has their own dna and dictates right handed left handed black or brown hair big or small etc 3. everyone has different dna it_s like a fingerprint 4. not everyone has the same dna don_t know any more 5. dna is a finger print everyone has there own dna 6. this is a specific if you any saliva or hair there is a one in many billions of chances that it would be the same 7. only you have the dna which is the same as you its like a finger print 8. everyone has individual dna blood fingerprints saliva etc doesn_t know exactly what it is 9. you can get dna from to your mouth fingerprints and every part of your body everyone_s dna is different 10. fingerprints other tissues hair particles nails etc everyone has their own dna no dna is the same

If we now look at the analysis of the engagement variables in Figure 4 this hypothesis is borne out. Class 3 respondents are more knowledgeable and engaged than any other group. However, it is also true that all those classes towards the 'description' end of the horizontal dimension in Figure 3 tend to be more engaged than those on the function side. Hence we might conclude that DNA as it applies to everyday life is a representation salient to many of the public, even where they are not especially engaged with biomedical science. However, for those that are engaged, the actual nature of DNA and its double helix is salient. In terms of communication strategies, it may well be more effective to focus on the uses of DNA rather than 'what it is' for some groups of the public.

Figure 3: Alceste correspondence analysis: DNA

```
+-----|---------|---------|--------- ---------|---------|---------|-----+
|                                  |      out                          |
|                                  |                information        |
|                                  |          understand    organism |
|                                  |          blue spiral    helix acid


|         up                       |           deoxyribonuc nucleic  |
|      used mother                 |             colour life  protein|
|         down pass                |                                 |
|father match way                  |              stand       coding |
crime       police see know little  |                                 |
| use       take come catch         |  print                          |
```

Class 5

Class 2

Class 1

Class 4

Class 3

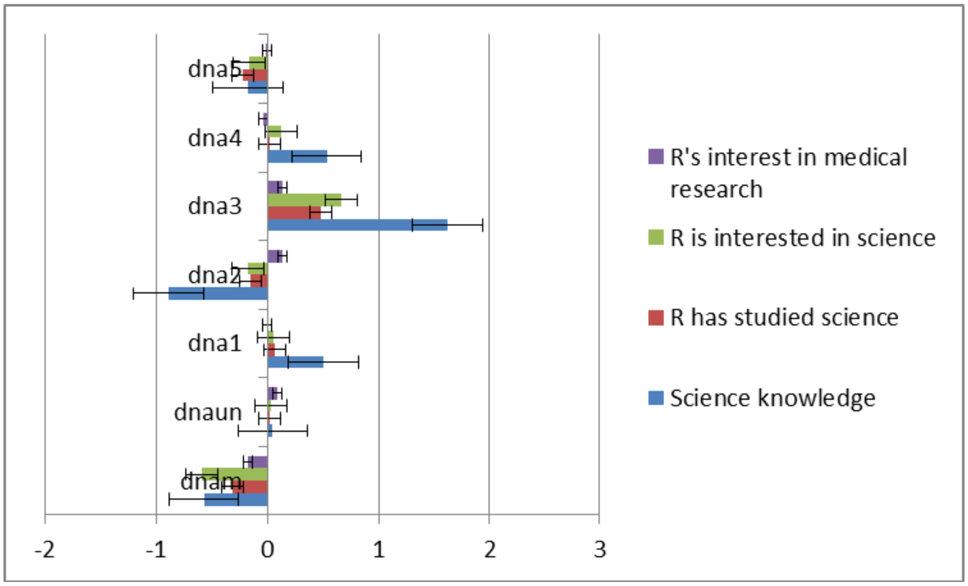
```
|dad to     sample swab criminal    |                                 |
else  same    blood mouth identify  |                                 |


|tell    DNA trace identification   |                                 |
|   people   saliva                 |                           strand
|   someone can_t identifying  that_s                        develop |
|           show                    |        determine give           |
|    look    tissue    skin         |             long living    character
 --------------------hair--leave finger----------------blueprint- ---form
|   find think              everyone                            molecule
|                                  thing            define              |


|                                  |           chemical chain        |
|                                  personal molecular                 |
|                     certain part    identity      human              |
|                                  |    word                          |
|                             different  basic                         |
|                                  |           code structure   cell
|                                  |   footprint basi                 |
|                              into|                                  |


|                                  |                                  |
|                               fingerprint  go   plant               |
|                                  |          animal                   |
|                                  |             gene                 |
|                                  |        map                       |
|                                  |             make                 |
|                        each inherit  person bodies make_up          |
|                                  body unique   individual pattern    |
+-----|---------|---------|--------- ---------|---------|---------|-----+
X axis   40 % of the inertia
Y axis  28 % of the inertia
```

Figure 4: Mean differences in interest and knowledge from sample mean



5.3 Do the Narrative Classes Have Explanatory Power?

Thus far we have shown that the verbatim responses can be partitioned into substantively interpretable groupings that vary with respect to how people talk about these two aspects of biomedical science when asked to do so in their own terms. Additionally, we have found that the derived narrative classes differ in their levels of knowledge about and, to a lesser extent, interest in science, for both stem cells and DNA. An additional question of interest is whether the narrative classes have any explanatory potential for understanding attitudes and orientations toward other aspects of biomedical science, over and above a set of standard predictors. The Alceste generated classes should represent *different ways of thinking* about stem cells or DNA, and it is illuminating to explore whether this dimension of variation allows us to gain greater analytical leverage.

To evaluate the analytical potential of the derived classes, we use them as predictors in regression models alongside a set of standard explanatory variables. We do this for two different outcome variables, one measuring optimism about the potential of genetic science to produce medical advances. The wording and response alternatives for this item are:

(1) “How optimistic are you about the possibility of medical advances as a result of genetic research?”

1 = very optimistic; 2 = somewhat optimistic; 3 = not too optimistic; 4 = not at all optimistic

And a second variable which measures self-rated understanding of the ethical issues raised by genetic research, which has the following wording and response alternatives:

(2) “To what extent do you agree with the following statement: ‘I feel I have a good understanding of the ethical issues raised by genetic research’?”

1 = strongly agree; 2 = agree; 3 = neither agree/disagree; 4 = disagree; 5 = strongly disagree

Frequency distributions for both variables are presented in Appendix B. The ‘optimism’ variable was recode into a binary variable where: 0 = somewhat optimistic; not too optimistic; and not at all optimistic; 1 = very optimistic. The ‘ethics’ variable was likewise coded into a binary variable where: 0 = neither agree/disagree; disagree; and strongly disagree, while 1 = agree and strongly agree. We then use binary logistic regression to predict the probability that an individual is 1. ‘very optimistic’ about the benefits of medical research and 2. possesses a ‘good understanding’ of ethical issues surrounding genetic research.

Table 9 presents the results for the ‘optimism’ item. Whilst interest in science is not significant, knowledge is a strong predictor. Surprisingly, being a member of stem cell class two is a equally strong predictor. The same is also true for the DNA classes, although this time non-response becomes a strong negative predictor of optimism.

Table 9: Binary logistic regression: class membership on optimism about medical research						
	Stem cell classes			*DNA classes*		
	Exp(B)	Sig	S.E.	Exp(B)	Sig	S.E.
Constant	.141	.000	.310	.165	.000	.271
Age 22-34	1.022	.887	.128	.997	.482	.340
Age 35-44	1.018	.889	.130	.954	.383	.316
Age 45-59	.729*	.014	.136	.732	.085	.313
Age 60 +	1.162	.269	.146	1.290	.939	.318
Sex (Female = 0)	1.545**	.003	.159	1.533**	.003	.144
No qualification (GCSE/A level = 0)	1.218	.216	.206	1.226	.798	.209
Degree and above (as above))	.923	.698	.185	.930	.359	.185
Interest in science (no = 0)	1.060	.753	.243	1.154	.192	.162
Knowledge score medium (low = 0)	1.397	.169	.276	1.406	.155	.242
Knowledge score high (as above)	2.491***	.001	.229	2.701***	.000	.271
Stemcell non-response	.752	.213	.256			
Stemcell class one	1.350	.240	.277			
Stemcell class two (knowledgeable def)	2.409**	.002	.219			
Stemcell class three	1.481	.072	.128			
DNA non-response				.380**	.009	.372
DNA class one				1.017	.947	.198
DNA class two				.660	.120	.267
DNA class three (knowledgeable def)				1.642*	.040	.234
DNA class four				1.077	.748	.264
DNA class five				.798	.482	.296
Cox & Snell R sq	.086			.083		
N	1179			1179		

Stemcell class two is a very strong predictor of having a good understanding of ethical issues. Its effect is greater than having high knowledge. Stem Cell non-response has a large negative effect. DNA classes demonstrate similar effects though not as strong as stem cell.

Table 10: Binary logistic regression: class membership on understanding of ethical issues

	Stem cell classes			*DNA classes*		
	Exp(B)	Sig	S.E.	Exp(B)	Sig	S.E.
Constant	.241	.000	.286	.237	.000	.254
Age 22-34 (age 16-21 = ref)	.971	.861	.165	.914	.594	.168
Age 35-44	1.036	.805	.141	.892	.425	.143
Age 45-59	1.055	.692	.134	1.103	.480	.138
Age 60 +	1.272	.098	.145	1.580**	.002	.150
Sex (Female = 0)	1.119	.412	.137	1.152	.286	.133
No qualification (GCSE/A level =	1.330	.064	.154	1.504**	.007	.151
Degree and above (as above)	.959	.826	.192	.865	.443	.189
Interest in science (no = ref)	1.713**	.003	.179	1.928***	.000	.174
Knowledge score medium (low =	1.592*	.038	.224	1.891	.003	.217
Knowledge score high (as above)	2.634***	.000	.259	3.443***	.000	.250
Stemcell non-response	.432***	.000	.208			
Stemcell class one	1.846**	.007	.226			
Stemcell class two (knowledgeable	3.049***	.000	.281			
Stemcell class three	2.327***	.000	.196			
DNA non-response				.149***	.000	.410
DNA class one				1.281	.165	.178
DNA class two				.818	.366	.222
DNA class three (knowledgeable				1.769*	.014	.232
DNA class four				.981	.939	.245
DNA class five				.474**	.008	.282
Cox & Snell R sq	.197			.163		
N	1179			1179		

6. DISCUSSION

The premise of and motivation for this paper is that standard closed-format survey questions are, for a variety of reasons, potentially problematic in the way that they seek to gauge public opinion about relatively low-salience and technically ‘difficult’ areas of biomedical science (Kotchetkova et al. 2007: 14). In particular, the use of a small number of pre-determined evaluative response alternatives as the universe of possible responses to complex areas of science is unlikely to represent the full range of public reactions to and ideas about science but to reflect, rather, the orientation of the survey designer and commissioner. On the other hand, qualitative approaches are also considered problematic but for different reasons, because they do not allow reliable generalisation to the population as a whole. In recognising the weaknesses of both approaches, in this paper we have explored a set of procedures which are designed to combine the inferential strengths of quantitative analysis with the richness of 'free-form' textual data.

The results from the Alceste analyses have shown that verbatim responses prompted by the terms 'DNA' and 'stem cell' can be classified into groups which differ in their ways of thinking about these terms. That is to say, we are able to uncover latent narrative structures and discourses that are evident in the data, rather than imposed *a priori* by the formulation of fixed response alternatives. While some classes are somewhat ambiguous and difficult to interpret substantively, in both the DNA and stem cell examples, two classes emerged which performed better than 'interest in science' and always as good as standard knowledge variables in multivariate analyses – the 'non-response' class and the 'knowledgeable' class. It is worth recalling from the binary logistic regressions that these two classes were still significant predictors of the dependent variables even once age, gender, education, interest in science and knowledge were controlled for. So while at this stage the actual interpretative meanings of each class are unclear, there is clearly some evidence to support the use of open-ended questions and the verbatim responses they generate as a way of better identifying an engaged and knowledgeable class of individuals.

Although these procedures appear to offer some useful additional insights relative to a standard quantitative analysis, a number of limitations both with the data and with the Alceste approach were evident. The first of these is that, while the verbatim responses can in some senses be thought of as ‘qualitative’ data, many of the answers respondents provided were very short, containing just a few words in many instances. Usually, CAQDAS software is used to code large segments from paragraphs to speeches - if the individual is the unit of analysis - to manifestos or novels, if a particular type of document is the unit of analysis (see for example Bara et al. 2007 and Schonhardt-Bailey 2005). As we were working with on average 12 words for each unit of analysis, our search to uncover any latent heterogeneity was restricted. This is not, of course, an inherent limitation of the method but points toward the need to implement ways of collecting verbatim responses which maximises the semantic content of the data obtained. One interesting direction that might be usefully pursued in this regard would be to use audio-recording of the verbatim responses, which could later be transcribed (possibly via automated procedures), rather than using interviewers to type the oral responses as they are spoken.

A second limitation relates to the ontological status and meaning of the derived classes. As already mentioned, some of the classes were quite difficult to interpret substantively, given their loose structure of associated words. This difficulty of interpretation of the substantive meaning of latent structures which are produced via inductive statistical techniques is not unique to Alceste. Nonetheless, it is clear that the somewhat ambiguous nature of the derived classes, despite their evident analytical power, must be considered a limitation of this approach, particularly if the intention is to feed into policy and other areas of decision-making.

Finally, an interest for any quantitative researcher is identifying which side of a regression equation a variable should be on. Put differently, we wish to know whether a variable should be considered as a cause or as an effect when placed alongside other variables of interest. It is not clear, however, whether it is indeed possible at all to think of the Alceste classes in such causal terms. Perhaps the verbatim responses and the general classification that can be applied are all we need to know about how individuals generally relate towards X; perhaps these classes themselves signal different underlying orientations and representations of X and for each individual these different ways of thinking about X also interact with attitudinal variables which in turn stimulate engagement. As such, the role of these classes in the chain of causality, if there is one, is very much unknown.

Despite these limitations, the analyses we have conducted point the way to a new and potentially useful approach to both collecting and analysing survey data on public opinion toward biomedical science and technology. It is an approach which offers both the capability to provide insight of a different nature than is possible using standard fixed-response survey questions and which can make use of an often-neglected aspect of survey data, the open-ended question.

BIBLIOGRAPHY

Bauer M, J Durant and G Evans (1994) European public perceptions of science, *International Journal of Public Opinion Research*, 6, 2, 163-186

Bhattachary, Darren (2008) Stem Cell Dialogue (London: British Market Research Bureau). www.bbsrc.ac.uk/nmsruntime/saveasdialog.aspx?lID=2937&sID

Bishop, G. D., David L. Hamilton, and John B. (1980) Attitudes and nonattitudes in the belief systems of mass publics. *The Journal of Social Psychology*, Vol 110(1), Feb, 53-64

Converse, Philip (1964) `The Nature of Belief Systems in Mass Publics', in David Apter (ed.) Ideology and Discontent, pp. 206-261. New York: Free Press

De Vaus, D. (2002). Surveys in Social Research (5th ed.). London: Routledge.

Frewer, L.J. and Salter, B (2007). Societal trust in risk analysis: implications for the interface of risk assessment and risk management. In Siegrist, M. (editor). Risk management and Public trust. Earthscan, London. Pp 143-159

Gaskell, G. (2004) Science policy and society: the British debate over GM agriculture, Current Opinion in Biotechnology Volume 15, Issue 3, June 2004, Pages 241-245

Gaskell, G., and Bauer, M. (Eds). (2001). Biotechnology 1996-2000: The years of controversy. London: Science Museum Press.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11, 255-274.

Greer, J. G. (1988) “What do open-ended questions measure?” *Public Opinion Quarterly 52*: pp. 365–371

Hayes, B and V. Tariq (2000) Gender Differences in Scientific Knowledge and Attitudes Towards Science', *Public Understanding of Science*, 9 (2000) : 433-447

Kleinman, D. L. (2000) “Democratizations of Science and Technology,” in D. L. Kleinman (ed.) *Science, Technology, and Democracy*, pp. 139–66. Albany, NY: SUNY Press.

Kotchetkova, Inna, Evans, Robert and Langer, Susanne (2007) Public opinion as questions not answers: how citizens evaluate the therapeutic potential of stem cell research in the context of T1 diabetes, QUALITI Working Paper, March 2007

Lazarsfeld, P. F., Bernard Berelson, and Hazel Gaudet (1944) The People's Choice:How the voter makes up his mind a presidential campaign, New York: Duell, Sloan and Pearce

Losh, S. C. (2010) Age, generational, and educational effects on American adult public understanding of science. In: Bauer MW, Shukla R and Allum N (eds) *The Culture of Science—How does the public relate to science across the globe?* New York: Routledge.

Moscovici, Serge (1963) Attitudes and Opinions, *Annual Review of Psychology* Vol. 14: 231-260

Newton, K. (2006) May the weak force be with you: The power of the mass media in modern politics. European Journal of Political Research, Vol. 45, No. 2, 209-234., (ISSN: 0304-4130)

Niglas, K. (2004). The combined use of qualitative and quantitative methods in educational research. Tallinn, Estonia: Tallinn Pedagogical University Dissertation on Social Sciences.

Nisbet, Matthew C. 2004. "THE POLLS--TRENDS: PUBLIC OPINION ABOUT STEM. CELL RESEARCH AND HUMAN CLONING." Public Opinion Quarterly 68. (1):131-54

Pattison J. UK Stem cell initiative: report and recommendations. London: DoH, 2005.

Reinert, M. (1998). Mondes lexicaux et topoi dans l'approche Alceste [Lexical worlds and topoi in the Alceste approach]. In S. Mellet & M. Vuillaume (Eds.) Mots chiffrés et déchiffrés. Mélanges offerts à Étienne Brunet, Travaux de linguistique quantitative, SLATKINE, Genève, pp. 289-303

Ruckelshaus, W. J. (1986) "Overview of the Problem. Communicating about. Risk ." In Davies, J. C., Vincent T. Covello, and Frederick W. Allen (eds.), *Risk Communication: Proceedings of the National Conference on Risk Communication*, Washington D.C., January 29-31, 1986 (Washington, D.C.: Conservation Foundation)

Salter, B. and Jones, M. (2006) Change in the policy community of human genetics: A pragmatic approach to open governance. Policy and Politics, 34(2), 347-366

Salter, B., and M. Jones. 2002. Human genetic technologies, European governance and the politics of bioethics. Nature Reviews Genetics 3 (10): pp. 808-14.

Sandelowski Margarete, Voils Corrine and Knafl George (2009) On Quantitizing, *Journal of Mixed Methods Research* July 2009 vol. 3 no. 3 208-222.

Schlesinger, P. (1993), Wishful Thinking: Cultural Politics, Media, and Collective Identities in Europe. Journal of Communication, 43: pp. 6–17

Schuman and Presser. (1981). Questions and Answers in Attitude Surveys: Experiments in Question Form, Wording, and Context. New York: Academic Press

Sturgis, P. and Allum, N. 2006 A literature review of research conducted on public interest, knowledge and attitudes to biomedical science. London: The Wellcome Trust.

Sturgis,P. and N. Allum (2004) Science in society: re-evaluating the deficit model of public attitudes. *Public Understanding of Science* 13, 55-74

The Phillips Report (2000). The BSE Inquiry, London, HMSO, available online at: www.bseinquiry.gov.uk/

Wilsdon J, Wynne B, Stilgoe J (2005) The Public Value of Science. London, UK: Demos

Wood, W., Rhodes, N., & Biek, M. (1995). Working knowledge and attitude strength: An information-processing analysis, pp. 282-313 In R. E. Petty &. J. A. Krosnick (Eds.), Attitude Strength: Antecedents and Consequences. Hillsdale, NJ: Lawrence Erlbaum Associates

APPENDIX A

Table 5: Characteristic statements: stem cells

1	1.	you get it from cord blood of babies use it help a sibling
	2.	they come from the umbilical cord of a baby they can save a child with leukaemia
	3.	blood taken from umbilical cord which can be used fight genetic diseases
	4.	its used help people with illness as cure them
	5.	something do with material taken from afterbirth of new born babies i e umbilical cord
	6.	taking cells from the umbilical cord as soon as the placenta has been delivered that is when they inject them i siblings or children who are suffering from a rare sort of cancer
	7.	take from the umbilical cord from a baby but what they actually do i don_t know but they can use them grow new cells
	8.	its do with haemophiliacs and can be a cure for leukaemia and cancers and blood disorders
	9.	taken from spinal cord and it may lead cures
	10.	the cell drawn from umbilical code of new born babies
2	1.	an undifferentiated cell which has yet specialise into a particular type of cell e g liver cell
	2.	basic cells the procuring cell prior differentiation into different cell types
	3.	a cell having potential turn into a particular type of cell
	4.	its the cells which develop into different types of cells a kind of blueprint
	5.	is an undifferentiated cell which has possibility of turning into a specific type of cell
	6.	non differentiated cells which have the ability turn into any type of cell
	7.	a non differentiated cell that can divide into different specific tissue types
	8.	it is the cell at a very early stage of development before it has started divide
	9.	its the very basic cell that isn_t differentiated into anything so you can get all the types of cells you need from it
	10.	basic cell which can develop into any kind of cell
3	1.	part of the body whereby your tissue can reproduce its linked the brain
	2.	you can grow identical parts of body which are damaged
	3.	they can reproduce other organs or parts of the body
	4.	they regenerate damaged parts of the body for example the nervous system
	5.	the growth for hormone in a persons body
	6.	it is part of the brain respondent can_t say more
	7.	recreate tissue that can regenerate itself

8. part of the dna grown specifically
9. parts of the body appertaining the brain
10. using a small part of a persons

Interest in medical research

"Medical research is about how the body works, the causes of illnesses and diseases and developing and testing new treatments. How interested, if at all, would you say you are in medical research?"

1 Not at all interested

2 Not very interested

3 Fairly interested

4 Very interested

5 Don't know

Interest in science

"Overall, how interesting did you find science lessons at school?"

1 Not at all interesting

2 Not very interesting

3 Fairly interesting

4 Very interesting

Has R ever studied science

"Have you ever studied for a qualification in biology or genetics at school, college or university"?

0 No

1 Yes

Scientific knowledge

"Now for a quick quiz about science. For each of the following statements, please say whether you think it is definitely true, probably true, probably false or definitely false. If you don't know, just say so and we'll go on to the next one."

The knowledge quiz included a range of items, focusing on a variety of aspects of science and ranging from things which it was anticipated would be commonly known to more obscure ones.

APPENDIX B

Optimistic for medical advances based on genetic research

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very optimistic,	301	25.5	25.5	25.5
	Somewhat optimistic,	691	58.6	58.6	84.1
	Not too optimistic, OR,	117	9.9	9.9	94.1
	Not at all optimistic?	44	3.7	3.7	97.8
	Don't Know	26	2.2	2.2	100.0
	Total	1179	100.0	100.0	

Understanding the ethical issues of genetic research

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly agree	59	5.0	5.0	5.0
	Agree	415	35.2	35.2	40.2
	Neither agree nor disagree	348	29.5	29.5	69.7
	Disagree	269	22.8	22.8	92.6
	Strongly disagree	77	6.5	6.5	99.1
	Don't Know	11	.9	.9	100.0
	Total	1179	100.0	100.0	

APPENDIX C

Alternate coding on dependent variable – combine 1&2 = 1

Table 9b: Binary logistic regression: class membership on optimism about medical research

	Stem cell classes			*DNA classes*		
	Exp(B)	Sig	S.E.	Exp(B)	Sig	S.E.
Constant	2.747	.026	.453	.202	.000	.379
Age 22-34	.444	.065	.440	.787	.482	.340
Age 35-44	.917	.841	.435	.759	.383	.316
Age 45-59	.614	.240	.415	.584	.085	.313
Age 60 +	.698	.390	.418	1.025	.939	.318
Sex (Female = 0)	1.252	.194	.173	1.540**	.003	.144
No qualification (GCSE/A level = 0)	.643*	.032	.206	.948	.798	.209
Degree and above (as above))	1.853	.056	.323	1.185	.359	.185
Interest in science (no = 0)	1.382	.165	.233	1.235	.192	.162
Knowledge score medium (low = 0)	1.990***	.001	.207	1.410	.155	.242
Knowledge score high (as above)	3.100***	.001	.332	2.694***	.000	.271
Stemcell non-response	.949	.823	.232			
Stemcell class one	1.888*	.051	.325			
Stemcell class two (knowledgeable def)	2.741	.062	.541			
Stemcell class three	3.204***	.000	.313			
DNA non-response				.380**	.009	.372
DNA class one				1.013	.947	.198
DNA class two				.660	.120	.267
DNA class three (knowledgeable def)				1.618*	.040	.234
DNA class four				1.089	.748	.264
DNA class five				.812	.482	.296
Cox & Snell R sq	.086			.082		
N	1179			1179		